Estimation of Net Ecosystem Carbon Exchange for the Conterminous United States by Combining MODIS and AmeriFlux Data

Jingfeng Xiao[1*], Qianlai Zhuang[2], Dennis D. Baldocchi[3], Paul V. Bolstad[4], Sean P. Burns[5], Jiquan Chen[6], David R. Cook[7], Peter S. Curtis[8], Bert G. Drake[9], David R. Foster[10], Lianhong Gu[11], Julian L. Hadley[12], David Y. Hollinger[13], Gabriel G. Katul[14], Beverly E. Law[15], Marcy Litvak[16], Siyan Ma[17], Timothy A. Martin[18], Roser Matamala[19], Steve McNulty[20], Tilden P. Meyers[21], Russell K. Monson[5], J. William Munger[22], Asko Noormets[23], Walter C. Oechel[24], Ram Oren[14], Andrew D. Richardson[25], Hans Peter Schmid[26], Russell L. Scott[27], Gregory Starr[18], Ge Sun[20], Andrew E. Suyker[28], Margaret S. Torn[29], Kyaw Paw U[30], Shashi B. Verma[28], Sonia Wharton[30], Steven C. Wofsy[31]

[1]Department of Earth & Atmospheric Sciences, Purdue Climate Change Research Center, Purdue University, West Lafayette, IN, USA

[2]Department of Earth & Atmospheric Sciences, Department of Agronomy, Purdue Climate Change Research Center, Purdue University, West Lafayette, IN, USA

[3]Atmospheric Science Center, University of California - Berkeley, Berkeley, Berkeley, CA, USA

[4]Department of Forest Resources, University of Minnesota, St. Paul, MN, USA

[5]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

[6]Earth, Ecological & Environmental Sciences, University of Toledo, Toledo, OH, USA

[7]Argonne National Laboratory, Environmental Science Division, Argonne, IL, USA

[8]Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, OH, USA

[9]Smithsonian Environmental Research Center, Edgewater, MD, USA

[10]Harvard Forest and Department of Organismic and Evolutionary Biology, Harvard University, Petersham, MA, USA

[11]Oak Ridge National Laboratory Environmental Sciences Division, Oak Ridge, TN, USA

[12]Harvard Forest, Harvard University, Petersham, MA, USA

[13]USDA Forest Service, Northeastern Research Station, Durham, NH, USA

[14]School of the Environment, Duke University, Durham, NC, USA

[15]College of Forestry, Oregon State University, Corvallis, OR, USA

[16]Department of Biology, University of New Mexico, Albuquerque, NM, USA

[17]Department of Environmental Science, Policy, & Management, University of California - Berkeley, Berkeley, CA, USA

[18]University of Florida, Gainesville, FL, USA

[19]Argonne National Laboratory, Biosciences Division, Argonne, IL, USA

[20]USDA-Forest Service, Southern Research Station, Raleigh, NC, USA

[21]NOAA/ARL, Atmospheric Turbulence and Diffusion Division, Oak Ridge, TN, USA

[22]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA

[23]Department of Forestry and Environmental Resources and Southern Global Change Program, North Carolina State University, Raleigh, NC, USA

[24]Department of Biology, San Diego State University, San Diego, CA, USA

[25]Complex Systems Research Center Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH, USA

[26]Department of Geography, Indiana University, Bloomington, IN, USA

[27]USDA-ARS Southwest Watershed Research Center, Tucson, AZ, USA

[28]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA

[29]Lawrence Berkeley National Laboratory, Earth Science Division, Berkeley, CA, USA

[30]Department of Land, Air and Water Resources, University of California - Davis, Davis, CA, USA

[31]Division of Engineering and Applied Science/Department of Earth and Planetary Science, Harvard University, Cambridge, MA, USA

*Corresponding author: jing@purdue.edu

Abstract

Eddy covariance flux towers provide continuous measurements of net ecosystem carbon exchange (NEE) at half-hourly or hourly time steps. The use of these flux measurements has improved our understanding of the net exchange of carbon dioxide between terrestrial ecosystems and the atmosphere for a wide range of climate and biome types. There is growing interest in scaling up these measurements to regional or continental scales. Here we used remote sensing data from the Moderate Resolution Imaging Spectrometer (MODIS) instrument on board the NASA's Terra satellite to extrapolate NEE measured at AmeriFlux sites to the continental scale. We first combined MODIS data and NEE measurements for representative U.S. ecosystems including forests, grasslands, shrublands, savannas, and croplands to develop a predictive NEE model using a regression tree approach. All explanatory variables of the model were derived from MODIS data. The regression tree model was trained and validated using AmeriFlux NEE measurements over the period 2000-2004 and 2005-2006, respectively. We then applied the model to the continental scale and estimated NEE for each 1 km × 1 km cell for the conterminous U.S. for each 8-day period in 2005 using spatially explicit MODIS data. We found that the model predicted NEE reasonably well at the continental scale, and generally captured the spatiotemporal patterns of NEE for the conterminous U.S. Our study demonstrated that our approach is powerful for scaling up eddy flux NEE measurements to the continental scale and producing spatially explicit NEE estimates. This approach could provide an independent dataset from simulations with biogeochemical models and inverse modeling approaches for examining the spatiotemporal patterns of NEE and constraining regional, continental, or global terrestrial carbon sink or source activities.

1. Introduction

Net ecosystem carbon exchange (NEE), the net effect of photosynthetic uptake and release of carbon dioxide (CO_2) by respiration from autotrophs (plants) and heterotrophs (e.g., microbial decomposition), represents the net exchange of carbon dioxide (CO_2) between terrestrial ecosystems and the atmosphere [Law et al., 2006]. The quantification of NEE for regions, continents, and the globe can improve our understanding of the feedbacks between the terrestrial biosphere and the atmosphere in the context of global change, and facilitate climate policy-making. Therefore, the estimation of NEE over large areas has important scientific and political implications.

To date, numerous techniques have been used to estimate NEE [Baldocchi et al., 2001]. For example, atmospheric inverse models have been used to provide aggregated information on carbon balances over large areas [e.g., Gurney et al., 2002]. The accuracy of those estimates is limited by the sparseness of the CO_2 observation network, their biased placement in the marine boundary layer, and the accuracy of the atmospheric transport models [Tans et al., 1990; Denning et al., 1996; Fan et al., 1998]. Moreover, this approach does not provide information about which ecosystems are contributing to the sinks/sources or the processes involved [Janssens et al., 2003]. During the last three decades, researchers have also been using global biogeochemical models such as BIOME-BGC [Running and Hunt, 1993], CASA [Potter et al., 1993; Field et al., 1995], CENTURY [Parton et al., 1993], and Terrestrial Ecosystem Model [TEM; Zhuang et al., 2003; Xiao et al., 2008] to quantify NEE over large areas. These models are dependent on site-level parameterizations, and the accuracy of model simulations at regional scales is therefore limited. In addition, besides atmospheric CO_2 and climate variability, other factors such as land use/land cover change [McGuire et al., 2001],

disturbances [Zhuang et al., 2002; Law et al., 2004], and management practices [Xiao and Moody, 2004a; Magnani et al., 2007] significantly affect NEE. It is a challenge for these biogeochemical models to consider all these factors due to model limitations and/or lack of data.

At the site level, eddy covariance flux towers provide continuous measurements of ecosystem-level exchanges of carbon, water, energy, and momentum at half-hourly or hourly time steps [Baldocchi et al., 2001]. At present, over 400 eddy covariance flux towers are operating on a long-term and continuous basis over the globe [FLUXNET, 2008]. This global network of flux towers, called FLUXNET, encompass a large range of climate and biome types [Baldocchi et al., 2001]. These flux towers provide the most extensive, reliable, and longest measurements of ecosystem carbon fluxes. However, these measurements only represent the fluxes at the scale of the tower footprint [Running et al., 1999], up to several square kilometers [Schmid, 1994]. The quantification of the terrestrial carbon sinks/sources for regions, continents, and the globe often require spatially explicit estimates of NEE. The way of using these flux data to quantify the terrestrial carbon fluxes at large scales is an intriguing question.

Some efforts have been made to scale up NEE measured at eddy covariance flux towers. For example, Mahadevan et al. [2008] developed the vegetation photosynthesis and respiration model (VPRM) for estimating NEE at hourly time steps. This model uses eddy covariance flux data to calibrate and validate the VPRM model. Similar to process-based biogeochemical models, this empirical model is also based on site-level parameterizations. Satellite remote sensing is a potentially valuable tool for scaling up NEE measurements to continental and global scales [Running et al., 1999]. Yamaji et al. [2007] linked satellite data to flux tower NEE data at the tower level. Wylie et al. [2007] estimated NEE for grasslands in the

northern Great Plains using satellite data and flux tower NEE measurements. Despite these efforts, to our knowledge, no study has scaled up flux tower NEE measurements to the continental scale.

Here we combined MODIS (Moderate Resolution Imaging Spectroradiometer) and eddy covariance flux data to scale up flux tower NEE measurements to the continental scale. First, we developed a predictive NEE model based on site-level MODIS and AmeriFlux data. Second, we validated the performance of the model with independent AmeriFlux data. Third, we applied the model to estimate NEE for each 1 km × 1 km cell across the conterminous U.S. for each 8-day period in 2005 using spatially explicit MODIS data. Finally, we examined the spatiotemporal patterns of NEE for representative U.S. ecosystems. This undertaking represents the first study of its kind and is the first step towards continental-scale extrapolations of flux tower NEE measurements across the conterminous U. S.

2. Methods

2.1. Regression tree

Regression tree algorithms predict class membership by recursively partitioning a dataset into more homogeneous subsets. The partitioning process splits each parent node into two child nodes, and each child node is treated as a potential parent node [Breiman et al., 1984]. The regression tree algorithm produces rule-based models containing one or more rules, each of which is a set of conditions associated with a linear submodel. Regression tree models can account for a nonlinear relationship between predictive and target variables and allow both continuous and discrete variables as input variables [Yang et al., 2003]. Regression tree methods are proven not only more effective than simple techniques including multivariate

linear regression, but also easier to understand than neural networks [Huang and Townshend, 2003].

We used the regression tree algorithm implemented in the commercial software called Cubist. Cubist is a powerful tool for generating rule-based predictive models. The predictive accuracy of a rule-based model can be improved by combining it with an instance-based/nearest-neighbor model that predicts the target value of a new case using the average predicted values of the *n* most similar cases. The use of the composite model can improve the predictive accuracy relative to the rule-based model alone. Cubist can also generate committee models made up of several rule-based models, and each member of the committee model predicts the target value for a case and the member's predictions are averaged to give a final prediction.

Cubist uses three statistical measures to measure the quality of the constructed regression tree model, including average error, relative error, and product-moment correlation coefficient. The average error is calculated as [Yang et al., 2003]:

$$AE = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{1}$$

where AE is the average error of a tree model, N is the number of samples used to establish the tree, y_i and $\hat{y}_i$ are the actual and predicted values of the response variable. The relative error is calculated as [Yang et al., 2003]:

$$RE = \frac{AE(T)}{AE(\mu)} \tag{2}$$

where RE is the relative error of a tree model, AE(T) is the average error of the tree model, and AE(μ) is the average error that would result from always predicting the mean value.

Cubist has been used to estimate percent land cover [Huang and Townshend, 2003], impervious area [Yang et al., 2003], forest biomass [Salajanu et al., 2005], and ecosystem carbon fluxes [Wylie et al., 2007]. Piecewise regression models were selected as the most appropriate approach for scaling the flux tower data to ecoregions [Wylie et al., 2007]. We chose Cubist to construct a predictive NEE model based on AmeriFlux NEE measurements and satellite data, and then applied the model to estimate NEE at the continental scale. All the three statistical measures provided by Cubist were used to evaluate the performance of the tree model.

2.2. Explanatory variable selection

NEE, the small difference between two large carbon fluxes of photosynthesis and respiration [Law et al., 1999], is influenced by a variety of meteorological, physiological, atmospheric, and edaphic variables. At the leaf level, photosynthesis or gross primary productivity (GPP) is influenced by several factors, including solar radiation, air temperature, soil moisture, nitrogen availability, and leaf area index (LAI). Ecosystem respiration (R_e) includes autotrophic (R_a) and heterotrophic respiration (R_h). R_a is mainly influenced by air temperature and vegetation carbon, whereas R_h is influenced by soil organic carbon pools as well as surface soil temperature and soil moisture [Tian et al., 1999]. Changes in atmospheric CO_2 concentrations, air temperature, and precipitation also influence R_h through effects on GPP and R_a that affect the pool size of soil organic matter through changes in litterfall input [Tian et al, 1999]. At the stand or regional level, NEE is also affected by fractional vegetation cover [DeFries et al., 2002].

Many of these factors influencing NEE can be measured or approximated by satellite remote sensing. Optical remote sensing systems measure the surface reflectance, the fraction of

solar energy that is reflected by the Earth's surface. For a given wavelength, different vegetation types and/or plant species may have different reflectance [Schmidt and Skidmore, 2003]. The reflectance of the same vegetation type also varies depending on wavelength region, biophysical properties (e.g., biomass, leaf area, and stand age), soil moisture conditions, and sun-object-sensor geometry [Ranson et al., 1985; Penuelas et al., 1993]. Therefore, reflectance values from multiple spectral bands can provide useful information for estimating NEE. Moreover, surface reflectance can be used to develop vegetation indices and biophysical parameters that may account for factors influencing NEE, such as the normalized difference vegetation index (NDVI), the enhanced vegetation index (EVI), the land surface temperature (LST), the normalized difference water index (NDWI), the fraction of photosynthetically active radiation absorbed by vegetation canopies (fPAR), and LAI.

The NDVI captures the contrast between the visible-red and near-infrared reflectance of vegetation canopies. It is defined as:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \tag{3}$$

where ρ_{red} and ρ_{nir} are the visible-red and near-infrared reflectance, respectively. NDVI is closely correlated to the fraction of photosynthetically active radiation (fPAR) absorbed by vegetation canopies [Asrar et al., 1984; Xiao & Moody, 2004b]. NDVI is also related to vegetation biomass [Myneni et al., 2001] and fractional vegetation cover [Xiao and Moody, 2005]. However, NDVI has several limitations, including saturation in a multilayer closed canopy and sensitivity to both atmospheric aerosols and soil background [Huete et al., 2002; Xiao and Moody, 2005]. To account for these limitations of NDVI, Huete et al. [1997] developed the improved vegetation index, EVI:

$$EVI = 2.5 \times \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + (6 \times \rho_{red} - 7.5 \times \rho_{blue}) + 1} \quad (4)$$

where ρ_{nir}, ρ_{red}, and ρ_{blue} are the spectral reflectance at the near-infrared, red, and blue wavelengths, respectively.

The LST derived from MODIS is a measure of the soil temperature at the surface. The MODIS LST agreed with in situ measured LST within 1 K in the range 263-322 K [Wan et al., 2002]. LST is likely a good indicator of R_e as both R_a and R_H are significantly affected by air/surface temperature. Rahman et al. [2005] demonstrated that satellite-derived LST was strongly correlated with R_e.

As the short infrared (SWIR) spectral band is sensitive to vegetation water content and soil moisture, a combination of NIR and SWIR bands have been used to derive water-sensitive vegetation indices [Ceccato et al., 2002]. Gao [1996] developed the NDWI from satellite data to measure vegetation liquid water:

$$NDWI = \frac{\rho_{nir} - \rho_{swir}}{\rho_{nir} + \rho_{swir}} \quad (5)$$

where ρ_{swir} is the reflectance at the shortwave infrared (SWIR) spectral band. The NDWI was shown to be strongly correlated with leaf water content (equivalent water thickness (EWT), g H_2O/m^2) [Jackson et al., 2004] and soil moisture [Fensholt and Sandholt, 2003] over time. It was incorporated into the vegetation photosynthesis model (VPM) as a water scalar for estimating GPP [Xiao et al., 2005].

Satellite data can also provide estimates for LAI and fPAR. These two variables characterize vegetation canopy functioning and energy absorption capacity [Myneni et al., 2002], and are key parameters in most ecosystem productivity and biogeochemical models [Sellers et al., 1997]. These two variables are closely associated with GPP.

We therefore selected surface reflectance, EVI, LST, NDWI, fPAR, and LAI as explanatory variables. All these variables were derived from MODIS data.

2.3. Data

We obtained the following three types of data, including NEE data from eddy covariance flux towers, explanatory variables derived from MODIS data, and a land cover map derived from MODIS data.

2.3.1 AmeriFlux data

As a part of the FLUXNET, the AmeriFlux network coordinates regional analysis of observations from eddy covariance flux towers across North America, Central America, and South America [AmeriFlux, 2007]. We obtained the Level 4 NEE product for 42 AmeriFlux sites for the period 2000-2006 from the AmeriFlux website (http://public.ornl.gov/ameriflux/) (Table 1). These sites involve a variety of vegetation types, including forests, shrublands, savannas, grasslands, and croplands (Table 1), and are distributed across the conterminous U.S. (Figure 1). We therefore believed that these sites are representative of the U.S. terrestrial ecosystems.

The Level 4 product consists of two types of NEE data, including standardized (NEE_st) and original (NEE_or) NEE [AmeriFlux, 2007]. NEE_st was calculated using the storage obtained from the discrete approach (single point on the top of the tower) with the same approach for all the sites, whereas NEE_or was calculated using the storage sent by the PI that can be obtained with the discrete approach or using the profile system. The average data coverage during a year is only 65% due to system failures or data rejection, and therefore robust and consistent gap filling methods are required to provide complete data sets [Falge et al., 2001]. Both NEE_st and NEE_or were filled using the Marginal Distribution Sampling

(MDS) method [Reichstein et al., 2005] and the Artificial Neural Network (ANN) method [Papale and Valentini, 2003]. The ANN method was generally, if only slightly, superior to the MDS method [Moffat et al., 2007]. Therefore, we used the NEE data filled using the ANN method. For each site, if the percentage of the remaining missing values for NEE_st was lower than that for NEE_or, we selected NEE_or; otherwise, we used NEE_st.

The Level 4 product consists of NEE data with four different time steps, including half-hourly, daily, weekly (8-day), and monthly. We used 8-day NEE data (g C m^{-2} day^{-1}) to mach the compositing intervals of MODIS data.

2.3.2. MODIS data

MODIS is a key instrument on board the NASA's Terra and Aqua satellites. The Terra MODIS and Aqua MODIS are viewing the entire Earth's surface every one to two days, acquiring data in 36 spectral bands and with the spatial resolution of 250m, 500m, and 1km. A variety of MODIS data products are currently available at the USGS-NASA Distributed Active Archive Center. We used the following four MODIS data products, including surface reflectance [MOD09A1; Vermote and Vermeulen, 1999], daytime and nighttime LST [MOD11A2; Wan et al., 2002], EVI [MOD13A1; Huete et al., 2002], and LAI/fPAR [MOD15A2; Myneni et al., 2002]. Surface reflectance data consist of reflectance values of seven spectral bands: blue (459-479 nm), green (545-565 nm), red (620-670 nm), near infrared (841-875 nm, 1230-1250 nm), shortwave infrared (1628-1652 nm, 2105-2155 nm). Surface reflectance and EVI are at spatial resolution of 500m, while LAI, fPAR, and LAI are at spatial resolution of 1km. Surface reflectance, fPAR, and LAI are at temporal resolution of 8 days, while EVI is at temporal resolution of 16 days.

For each AmeriFlux site, we obtained the MODIS ASCII subsets (Collection 4) consisting of 7 km × 7 km regions centered on the flux tower from the Oak Ridge National Laboratory's Distributed Active Archive Center [ORNL DAAC, 2006]. The objective of the MODIS ASCII Subset project is to provide data in support of validation of models or remote sensing products for selected FLUXNET or field sites. We extracted average values for the central 3 × 3 km area within the 7 × 7 km cutouts to better represent the flux tower footprint [Schmid, 2002; Rahman et al., 2005]. For each variable, we determined the quality of the value of each pixel within the area using the quality assurance (QA) flags included in the product. At each time step, we averaged the values of each variable using the pixels with good quality within the area to represent the values at the flux site. If none of the values within the 3 × 3 km area was of good quality, we treated the period as missing. Each 16-day EVI value was split into two 8-day values to correspond with the compositing interval of other MODIS data products.

For the continental-scale estimation of NEE, we obtained continental-scale MODIS data including surface reflectance, EVI, daytime and nighttime LST, and fPAR and LAI from Earth Observing System (EOS) Data Gateway. For each variable and for each 8- or 16-day period, a total of 22 tiles were needed to cover the conterminous U. S., and these tiles were mosaiced to generate a continental-scale image. For each variable, we determined the quality of the value of each pixel using the QA flags, and replaced the values with bad quality using a linear interpolation approach [Zhao et al., 2005]. The NDWI was calculated from band 2 (near-infrared, 841-876nm) and band 6 (shortwave infrared, 1628-1652) of the surface reflectance product (MOD09A1) according to equation (5). Each 16-day EVI composite was split into two 8-day composites to correspond with the compositing interval of other MODIS data products.

2.3.3. Land cover

To construct the predictive NEE model, we obtained the land cover type for each AmeriFlux site based on the site descriptions (Table 1). Here we used the International Geosphere-Biosphere Programme (IGBP) land-cover classification system (Belward and Loveland, 1996). Although the 42 AmeriFlux sites that we used cover a variety of vegetation types of this classification system, some vegetation types of the classification system have only one or two sites or even none. We therefore reclassified the vegetation types of this system to seven broader classes (Table 2). Specifically, evergreen needleleaf forests and evergreen broadleaf forests were merged to evergreen forests, deciduous needleleaf forests and deciduous broadleaf forests to deciduous forests, closed shrublands and open shrublands to shrublands, and woody savannas and savannas to savannas.

To estimate NEE for each 1 km × 1 km cell at the continental scale, we obtained the land cover type for each cell from the MODIS land cover map with the University of Maryland (UMD) classification system [Friedl et al., 2002]. The UMD system is similar to the IGBP classification system, but the UMD system does not include the following two vegetation classes including permanent wetlands and cropland/natural vegetation mosaics. Similarly, we reclassified the vegetation types of the MODIS land cover map to the seven broader classes (Table 2). The resulting reclassified land-cover map is shown in Figure 1.

2.4. Model development

We developed the predictive NEE model using Cubist based on the site-level MODIS and AmeriFlux NEE data. Our explanatory variables included surface reflectance (7 bands), daytime and nighttime LST, EVI, fPAR, and LAI, and our target variable was NEE. We temporally split the site-level data set of AmeriFlux and MODIS data into a training set (2000-

2004) and a test set (2005-present). If a site only had NEE observations for the period 2000-2004, the site was only included in the training set; if a site only had NEE observations for the period 2005-2006, the site was only included in the test set; otherwise, the site was included in both training and test sets. For example, the Fort Peck site (MT) had NEE data from the entire study period, and we used data from 2000 to 2004 for training, and data from 2005 to 2006 for validation; the Harvard Forest EMS Tower site had data available for 2000-2004, and we used all the data for training; the Fermi National Accelerator Laboratory Agricultural site had data for 2005 and 2006 only, and we used all the data for validation (Figure 2). The training and test sets included 40 and 34 AmeriFlux sites, respectively. We had a total of 4596 cases for the training set, and 2257 cases for the test set.

We used Cubist to construct a rule-based model by combining it with an instance-based model and committee models. We trained the model with the training set (2000-2004), and tested the model with the test set (2005-2006). In addition to the full model including all the explanatory variables, we also developed a series of models by dropping one or more variables at a time using Cubist. We then evaluated the performance of each model using average error, relative error, correlation coefficient, scatterplots of predicted versus observed NEE, and seasonal variations between the predicted and observed NEE. Finally, we chose the best model to estimate NEE at the continental scale.

2.5. Continental-scale estimation of NEE

The AmeriFlux network is representative of the conterminous U.S. ecoregions [Hargrove et al., 2003], while the 42 sites used in this study included most of the active flux sites in the network (Figure 1). Moreover, these sites cover a variety of vegetation types and are spatially distributed across the conterminous U.S. We therefore believed that the predictive

NEE model constructed from the 42 sites can be extrapolated to the conterminous U.S. We applied the predictive NEE model to estimate NEE for each 1 km × 1 km cell for the conterminous U.S. for each 8-day period in 2005 using spatially explicit MODIS data. We then examined the spatiotemporal patterns of our NEE estimates for representative U.S. ecosystems.

3. Results and discussion

3.1. Model development

We found that the best model contained the following explanatory variables, including surface reflectance bands 1-6, EVI, daytime and nighttime LST, and NDWI (relative error = 0.64, average error = 0.986, r = 0.73). This model achieved slightly higher performance than the full model (relative error = 0.66, average error = 1.01, r = 0.72). The best model estimated NEE reasonably well (Figure 3) considering that we used multiple years of data from a number of sites involving a variety of vegetation types across the conterminous U.S. The model slightly underestimated positive NEE values, and overestimated negative NEE values. In absolute magnitudes, the model slightly underestimated both carbon release and sequestration rates.

The analysis of residuals (Figure 4) indicated that the residuals were not randomly distributed. In absolute magnitudes, low NEE values were generally associated with low prediction errors, whereas high NEE values were associated with high prediction errors. This indicated that the explanatory variables included in the model could not completely explain the variance of NEE. It is likely because that the independent variables we used may not sufficiently account for the sizes of vegetation and soil organic carbon pools, and therefore lowered the prediction performance of the model.

We calculated the average error and relative error across all AmeriFlux sites for each 8-day period, and then plotted these two types of error against time (Figure 5). The average error showed a strong seasonality. In absolute magnitudes, winter had low average errors (about 0.6 g C m^{-2} day^{-1}), whereas warm season errors often exceeded 1 g C m^{-2} day^{-1}. This also suggests the relatively large uncertainties associated with NEE estimates, further indicating that NEE is difficult to estimate. The relative error, however, exhibited less seasonality, indicating that the relative error was less variable throughout the year than the average error.

We also compared our NEE estimates against observed NEE for each AmeriFlux site (Figure 6). The NEE estimates captured most features of observed NEE seasonality and interannual variability for the period 2005-2006. For some specific sites, episodes of under- or over-prediction occurred. The model could not well capture exceptionally high and low NEE values that represented large carbon release and sequestration rates, respectively for some sites, such as the Audubon Research Ranch site (AZ), Fermi National Accelerator Laboratory Agricultural site (IL), Goodwin Creek site (MS), and Fort Peck (MT). In absolute magnitudes, the model substantially underestimated those exceptional values. For example, the model estimates were far below the observed NEE values higher than 2 g C m^{-2} day^{-1} at the Greek Creek site (MS), and were far above the observed NEE values below -3 g C m^{-2} day^{-1} at the Audubon Research Ranch site (AZ). Overall, the model performed better for deciduous forests, savannas, grasslands and croplands than for evergreen forests and shrublands.

The disagreement between estimated and observed NEE values is due to the following three reasons. First, the MODIS and tower footprints did not always matched with each other. As mentioned earlier, for each explanatory variable derived from MODIS data, we used the values averaged within the 3 km × 3 km area (MODIS footprint) surrounding each flux tower

to represent the values at the tower. The footprints of MODIS and AmeriFlux matched with each other for most sites because the vegetation structure within the 3 km × 3 km area surrounding the flux tower is similar to that at the tower. However, some ecosystems are fairly complex in structure and topography, even over the relatively small area represented by a MODIS cell [Running et al., 1999], and the vegetation structure at the flux tower is different from that within the MODIS footprint. For example, the Tonzi Ranch site (CA) is dominated by deciduous blue oaks (*Quercus douglasii*), and the understory and open grassland are dominated by cool-season C_3 annual species [Ma et al., 2007]. The MODIS footprint consists of a larger fraction of grassland. The phenologies of blue oaks and grassland are distinct from each other [Ma et al., 2007], and therefore these two plant species had differential contributions to the NEE integrated over the MODIS footprint. In the spring, wet conditions along with warm temperatures facilitated the fast growth of grass, leading to large carbon sequestration rates within the MODIS footprint. Therefore, in absolute magnitudes, our NEE estimates were higher than the observed values at the tower. Grasses senesced by the end of the spring as the rainy season ended [Ma et al., 2007]. The senescence of grasses provided carbon sources in the summer, and thus lowered the carbon sequestration rates within the MODIS footprint. Therefore, in absolute magnitudes, our NEE values were much lower than the observed values at the tower in the summer.

Second, we estimated NEE at 8-day time steps, and therefore may not be able to capture the variability of NEE within each 8-day period. The MODIS LST and EVI data products were averaged from the corresponding daily products over a period of 8 and 16 days, respectively [Huete et al., 2002; Wan et al., 2002]. For each period, only data with good quality were retained for compositing, and thus the number of days actually used for compositing is

often lower than the total number of days over the period. The compositing technique for the MODIS surface reflectance product is based on the minimum-blue criterion that selects the clearest conditions over the 8-day period [Vermote and Vermeulen, 1999]. Therefore, the 8- or 16-day values did not always represent the average conditions over the 8- or 16-day period. The exclusion of days with high and low values could lead to underestimation and overestimation of NEE values, respectively. For example, each 16-day EVI composite was an average of daily EVI over a period of 16 days. The number of acceptable pixels over a 16-day compositing period is typically less than 10 and often less than 5 due to cloud contaminations and extreme off-nadir sensor view angles [Huete et al., 2002]. The compositing process may exclude high EVI values that represented high fPAR or fractional vegetation cover, therefore leading to lower estimates of NEE. On the other hand, the compositing process may also exclude low EVI values that represented low fPAR or fractional vegetation cover, thereby leading to higher NEE estimates.

Third, the exceptionally high or low NEE values that our model could not effectively capture might be affected by spikes remaining in the eddy covariance measurements. Among these exceptional NEE values, positive values were often associated with R_e values with much smaller magnitudes, leading to negative GPP estimates that were ecophysiologically impossible. Eddy covariance measurements are often affected by spikes, due to different reasons both biophysical (changes in the footprint or fast changes in turbulence conditions) and instrumental (e.g., water drops on sonic anemometer or on open path IRGA) [Papale et al., 2006].

We averaged our estimated and observed 8-day NEE for each AmeriFlux site and examined the relationship between estimated and observed mean 8-day NEE across the sites

(Figure 7). The model estimated NEE reasonably well at the site level ($r^2 = 0.72$, $p < 0.00001$). Overall, in absolute magnitudes, the model underestimated NEE. The performance of the model also varied with site. On average, some sites provided carbon sources, whereas other sites provided carbon sinks. High overestimation of carbon sequestration occurred at the Toledo Oak Openings site (OH), whereas high underestimation of carbon sequestration occurred at the Mature Red Pine site (WI), Duke Forest Pine site (NC), Duke Forest Hardwoods (NC), and North Carolina Pine (NC). High overestimation of carbon release occurred at Audubon Research Ranch (AZ), ARM Oklahoma (OK), and Freeman Ranch Mesquite (TX), whereas high underestimation of carbon release occurred at Mead Irrigated (NE), Goodwin Creek (MS), and Austin Cary (FL).

We also averaged our estimated and observed 8-day NEE over all AmeriFlux sites for each vegetation type, and examined the relationship between estimated and observed NEE across the vegetation types (Figure 8). The model predicted NEE at the biome level very well ($r^2 = 0.95$, $p < 0.00001$). Again, in absolute magnitudes, the model underestimated NEE. The performance of the model also varied with vegetation type. In absolute magnitudes, high overestimation occurred for evergreen forests and shrublands.

Our study demonstrated that MODIS data have great potential for scaling up flux tower NEE data to continental scales across a variety of vegetation types. NEE is much more difficult to estimate or simulate than GPP because the carbon pools and associated heterotrophic respiration are difficult to estimate or simulate [Running et al., 2004]. The performance of our model for estimating NEE is remarkable given the wide variety of ecosystem types, age structures, fire and insect disturbances, and management practices. In future research, additional explanatory variables should be selected in order to better account for vegetation

and soil carbon pools. For example, spatially explicit estimates of forest biomass [e.g., Zhang et al., 2006] and soil organic carbon [e.g., Global Soil Data Task Group, 2000] may provide useful information for vegetation and soil carbon pools and potentially improve the estimation of NEE. In addition, the use of higher temporal resolution (e.g., daily) may also improve the estimates of NEE, spikes in particular, by capturing the variability of NEE within each 8-day period.

3.2. Continental-scale estimation of NEE

We estimated NEE for each 1 km × 1 km cell for the conterminous U.S. for each 8-day interval over the period 1/1/2005-2/28/2006. Figure 9 shows examples of four 8-day NEE images in 2005, including March 14-21, July 12-19, September 14-21, and December 11-18. The regression tree model trained at the AmeriFlux sites generally captured the expected spatiotemporal patterns of NEE. The majority of the conterminous U.S. were carbon neutral or released carbon in March or December because at this time of the year most ecosystems were dormant; in July, the ecosystems in the east sequestered carbon from the atmosphere, whereas many areas in the west released carbon due to high R_e; in September, the ecosystems in the east sequestered less carbon as vegetation began to senesce. Some ecosystems in the Pacific Northwest and California sequestered carbon from the atmosphere throughout the year.

We aggregated 8-day NEE estimates for each season in 2005 (Figure 10). The carbon balances of terrestrial ecosystems at temperate and high latitudes exhibited strong seasonal fluctuations [Falge et al., 2002]. The NEE estimates also varied substantially over space. In the spring, many areas in the eastern half of the conterminous U.S. including the Southeast and the Gulf Coast exhibited negative NEE values, indicating that these ecosystems sequestered carbon from the atmosphere. The growing season of these ecosystems started in the mid- to late

spring, and GPP quickly exceeded R_e, leading to net carbon sequestration in the season. By contrast, the Upper Great Lakes region, the northern Great Plains, and the New England region exhibited positive NEE values and provided carbon sources. The Upper Great Lakes region and the northern Great Plains are dominated by croplands. Most crops were planted between April-June, with corn planted between April and mid-May, soybeans between mid-May and mid-June, and sorghum between late May and late June [Shroyer et al., 1996]. Crops were sparse in the beginning of the growing season and R_e was far above GPP, thereby leading to carbon releases. The New England region and the northern portion of the Upper Great Lakes region are dominated by temperate-boreal transitional forests, and their relatively late greenup due to low air temperatures led to carbon releases in the spring. Many regions in the western half of the conterminous U. S. released carbon into the atmosphere in the spring because of the sparse vegetation and the dominance of R_e over GPP. The Pacific Coast slightly sequestered carbon even in the spring because the dominant evergreen forests in the region sequestered carbon due to mild temperatures and moist conditions [Anthoni et al., 2002]. The Mediterranean region sin California also sequestered carbon in the spring. The Mediterranean climate is characterized by mild winter temperatures concomitant with the rainy season as opposed to severe summer droughts and heat [Barbero et al., 1992]. These ecosystems sequestered carbon due to precipitation surplus and relatively warm temperatures in the spring [Xu and Baldocchi, 2004].

In the summer, the eastern half of the conterminous U.S. provided carbon sinks as GPP far exceeded R_e due to optimal temperature and soil moisture conditions. By contrast, a vast majority of the land across the western counterpart released carbon into the atmosphere, including the Great Basin, the Colorado Plateau, and the western Great Plains. The 2005

summer drought affecting these regions [National Climatic Data Center, 2008] reduced GPP, whereas the high temperatures increased R_e, leading to net carbon releases. Some other regions in the west sequestered carbon from the atmosphere, including the northern Rocky Mountains and the Pacific Coast. Some Mediterranean ecosystems in California also released carbon into the atmosphere because summer is a part of the dry season.

In the fall, the Southeast and the Gulf Coast still provided carbon sinks, but the absolute values of NEE or the rates of carbon sequestration substantially decreased relative to those in the summer. This is because vegetation began to senesce in these regions in the fall. The Upper Great Lakes region and the Great Plains were largely carbon sources due to the harvesting of crops. The majority of the land across the west including the Great Plains, the Great Basin, and the Colorado Plateau released carbon into the atmosphere. The northern Pacific Coast, however, still provided carbon sinks. The Mediterranean ecosystems in California released carbon as the dry season spanned into the fall.

In the winter, the vast majority of the conterminous U.S. exhibited positive NEE values, indicating that the U.S. terrestrial ecosystems provided carbon sources in this season as most ecosystems were dormant at this season of the year. Some regions in the Pacific Coast sequestered carbon even in the winter because of the dominance of evergreen forests and mild temperatures in the regions. This agreed with the finding of Anthoni et al. [2002] that old-growth ponderosa pine in Oregon slightly sequestered carbon in the winter season. For the Mediterranean ecosystems in California, a smaller part of the region released carbon into the atmosphere relative to the fall as the wet season started in the winter.

Figure 11a shows the trajectory of the mean 8-day NEE (g C m^{-2} day^{-1}) for each vegetation type averaged over the entire conterminous U.S. throughout 2005. Overall,

deciduous forests, croplands, savannas, and mixed forests showed large intra-annual variability in NEE, whereas evergreen forests, grasslands, and shrublands exhibited much less interannual variability. The season patterns of NEE were determined by the seasonal differences in LAI, physiological capacity, meteorological conditions, the length of the growing season, soil temperature, and moisture status [Falge et al., 2002]. In the late fall, winter, and early spring, on average, the U.S. terrestrial ecosystems exhibited positive NEE values and therefore released carbon into the atmosphere. All vegetation types except evergreen forests and grasslands released carbon. Among vegetation types exhibited positive NEE values, deciduous forests exhibited highest values, followed by mixed forests; croplands exhibited intermediate values; shrublands and savannas exhibited lowest values. Evergreen forests still sequestered carbon from the atmosphere. During the growing season, on average, the U.S. terrestrial ecosystems exhibited negative NEE values and sequestered carbon from the atmosphere. All vegetation types except shrublands sequestered carbon. In absolute magnitudes, the highest NEE values occurred for deciduous forests, followed by croplands, savannas, and mixed forests; intermediate values occurred for evergreen forests; the lowest values occurred for grasslands. Shrublands released carbon because of high temperatures and low soil moisture conditions. Baldocchi et al. [2001] showed that the net CO_2 exchange of temperate deciduous forests increases by about 5.7 g C m^2 day^{-1} for each additional day that the growing season, defined as the period over which mean daily CO_2 exchange is negative due to net uptake by ecosystems, is extended. We found that on average, the CO_2 exchange of deciduous forests across the conterminous U.S. increased 3.3 g C m^{-2} day^{-1} for each additional day that the growing season is extended, Our continental-scale estimate was 42% lower than the estimate by Baldocchi et al. [2001] based on three temperate deciduous forests.

Figure 11b shows the trajectory of the total 8-day NEE (Tg C day^{-1}) aggregated from the NEE estimates over the conterminous U.S. for each vegetation type. The differences in the trajectories of total 8-day NEE among vegetation types were different from those of mean 8-day NEE because of the differences in the areas among vegetation types (Figure 12). In the late fall, winter, and early spring, the U.S. terrestrial ecosystems provided a carbon source (1-2 Tg C day^{-1}). Croplands, deciduous forests, and mixed forests provided carbon sources, whereas evergreen forests provided a carbon sink. Shrublands, savannas, and grasslands, however, were nearly carbon neutral. During the growing season, overall, the U.S. terrestrial ecosystems provided a carbon sink, with peak total NEE values of -17 Tg C day^{-1}. All vegetation types except shrublands functioned as carbon sinks. In absolute magnitudes, the highest total NEE values (about 10 Tg C day^{-1}) occurred for croplands; the intermediate values occurred for deciduous forests, savannas, and mixed forests; the lowest values occurred for evergreen forests and grasslands. By contrast, shrublands provided a carbon source. Total 8-day NEE exhibited largest intra-annual variability for croplands, intermediate variability for deciduous forests, savannas, and mixed forests, and lowest variability for evergreen forests, grasslands, and shrublands.

Our spatially explicit estimates of NEE across the conterminous U.S. provided an independent dataset from simulations by biogeochemical modeling and inverse modeling for examining the spatiotemporal patterns of NEE and constraining U.S. terrestrial carbon sinks/sources. Our estimates have advantages over these models simulations in that we took advantage of NEE measurements from a number of AmeriFlux sites involving representative U.S. ecosystems. Moreover, compared to biogeochemical modeling, our scaling-up approach implicitly considered the effects of climate variability, land use/land cover change,

disturbances, extreme climate events, and management practices. Compared to inverse modeling techniques, our approach provided estimates at high spatial (1 km × 1 km) and temporal resolutions (8 day). NEE is notoriously difficult to quantify over large areas [Running et al., 2004], and the accuracy of simulated NEE for regions and continents by biogeochemical models is poorly known due to lack of spatially explicit, independent validation datasets. Our estimates may provide an independent validation dataset for these model simulations.

4. Summary and conclusions

We combined MODIS and NEE data from 42 AmeriFlux sites involving a variety of vegetation types to develop a predictive NEE model using a regression tree approach. The model estimated NEE reasonably well. We then applied the model to estimate NEE for each 1 km × 1 km cell for the conterminous U.S. for each 8-day period in 2005. The model generally captured the spatiotemporal patterns of NEE. Deciduous forests, croplands, savannas, and mixed forests showed large intra-annual variability in NEE, whereas evergreen forests, grasslands, and shrublands exhibited much less interannual variability. Total 8-day NEE exhibited largest intra-annual variability for croplands, intermediate variability for deciduous forests, savannas, and mixed forests, and lowest variability for evergreen forests, grasslands, and shrublands.

Our study demonstrated that MODIS has great potential for scaling up AmeriFlux NEE measurements to the continental scale. Our estimates may provide an independent dataset from simulations with biogeochemical models and inverse modeling approaches for examining the spatiotemporal patterns of NEE and constraining terrestrial carbon sinks/sources for regions and continents. More importantly, our scaling-up approach implicitly considered the effects of climate variability, land use/land cover change, disturbances, extreme climate events,

and management practices that could not be easily considered altogether in other modeling approaches. Our spatially explicit NEE estimates at high spatial and temporal resolutions may therefore provide an independent dataset from simulations with biogeochemical models and inverse modeling approaches for continental-scale terrestrial carbon cycling studies.

Acknowledgements:

This study was funded by the National Science Foundation (NSF) Carbon and Water Program (EAR-0630319) and the Department of Energy (DOE) (Contract No. DE-AC02-05CH11231). We thank the PIs of the MODIS data products used, including Drs. Huete, Wan, Myeni, Vermote and other contributors as well as the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DACCC) and the Earth Observing System (EOS) Data Gateway for making these products available. We also thank T. A. Boden at the Carbon Dioxide Information Analysis Center, ORNL, S. K. S. Vannan at the ORNL DACC, Dr. M. Zhao at the University of Montana, and Dr. Z. Wan at the University of California, Santa Barbara for helpful discussion about AmeriFlux data, MODIS ASCII subsets, MODIS Quality Assurance (QA) flags, and MODIS LST, respectively. Computing support was provided by the Rosen Center for Advanced Computing, Purdue University.

References:

AmeriFlux (2007), AmeriFlux Network. http://public.ornl.gov/ameriflux/. Visited on 10/04/2007.

Anthoni, P. M., M. H. Unsworth, B. E. Law, J. Irvine, D. D. Baldocchi, S. V. Tuyl, and D. Moore (2002), Seasonal differences in carbon and water vapor exchange in young and old-growth ponderosa pine ecosystems, *Agri. For. Meteor., 111*, 203-222.

Asrar, G., M. Fuchs, E. T. Kanemasu, and J. L. Hatfield (1984), Estimating of absorbed photosynthesis radiation and leaf area index from spectral reflectance in wheat, *Agron. J., 76*, 300-306.

Baldocchi, D., E. Falge, L. Gu, R. Olson, D. Hollinger et al. (2001), FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, *Bull. Amer. Meteor. Soc., 82*, 2415-2434.

Barbero, M., P. Loisel, P. Quezel (1992), Biogeography, ecology and history of Mediterranean *Quercus ilex* ecosystems, *Vegetatio*, *99/100*, 19-34.

Belward, A. S., J. E. Estes, and K. D. Kline (1999), The IGBP-DIS Global 1-km Land-Cover Data Set DISCover: A Project Overview, *Photogram. Eng. Remote Sens., 65*, 1013-1020.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. 1984. *Classification and regression trees*. Chapman and Hall, New York, 358pp.

Ceccato, P., N. Gobron, S. Flasse, B. Pinty, and S. Tarantola (2002), Designing a spectral index to estimate vegetation water content from remote sensing data: Part 1 - Theoretical approach, *Remote Sen. Environ., 82*, 188-197.

DeFries, R. S., R. A. Houghton, M. C. Hansen, C. B. Field, D. Skole, and J. Townshend (2002), Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s, *PNAS, 99*, 14256-14261.

Denning, A. S., J. G. Gollatz, C. Zhang, D. A. Randall, J. A. Berry et al. (1996), Simulations of terrestrial carbon metabolism and atmospheric CO_2 in a general circulation model. Part 1: Surface carbon fluxes, *Tellus, 48B*, 521-542.

Dore, S., G. J. Hymus, D. P. Johnson, C. R. Hinkle, R. Valentini, and B. G. Drake (2003), Cross validation of open-top chamber and eddy covariance measurements of ecosystem CO2 exchange in a Florida scrub-oak ecosystem, *Global Change Biol., 9*, 84-95.

Falge, E., D. Baldocchi, R. Olson, P. Anthoni, M. Aubinet et al. (2001), Gap filling strategies for defensible annual sums of net ecosystem exchange, *Agri. For. Meteor., 107*, 43-69.

Falge, E., D. Baldocchi, J. Tenhunen, M. Aubinet, P. Bakwin et al. (2002), Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements, *Agri. For. Meteor., 113*, 53-74.

Fan, S. M., M. Gloor, J. Mahlman, S. Pacala, J. Sarmiento et al. (1998), A large terrestrial carbon sink in North America implied by atmospheric and oceanic carbon dioxide data and models, *Science, 282*, 442-446.

Fensholt, R., and I. Sandholt (2003), Derivation of a shortwave infrared water stress index from MODIS near- and shortwave infrared data in a semiarid environment, *Remote Sens. Environ., 87*, 111-121.

Field, C. B., J. T. Randerson, and C. M. Malmstrom (1995), Global net primary production: combining ecology and remote sensing, *Remote Sens. Environ., 51*, 74-88.

FLUXNET (2008), FLUXNET Project. http://daac.ornl.gov/FLUXNET. Visited on 01/02/2008.

Friedl, M. A., D. K. McIver, J. C. F. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf (2002), Global land cover mapping from MODIS: algorithms and early results, *Remote Sens. Environ., 83*, 287-302.

Gao, B. C. (1996), NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sens. Environ., 58*, 257-266.

Global Soil Data Task Group. 2000. Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS). [Global Gridded Surfaces of Selected Soil Characteristics (International Geosphere-Biosphere Programme - Data and Information System)]. Data set. Available on-line [http://www.daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

Gough, C. M., C. S. Vogel, H. P. Schmid, H-B. Su, and P. S. Curtis (2007), Multi-year convergence of biometric and meteorological estimates of forest carbon storage, Agri. For. Meteor., in press.

Gu, L., T. Meyers, S. G. Pallardy, P. J. Hanson, B. Yang, M. Heuer, K., P. Hosman, Q. Liu, J.S. Riggs, D. Sluss, S. D. Wullschleger (2007), Influences of biomass heat and biochemical energy storages on the land surface fluxes and radiative temperature, *J. Geophys. Res.*, D02107, doi:10.1029/2006JD007425.

Gu, L., T. Meyers, S. G. Pallardy, P. J. Hanson, B. Yang, M. Heuer, K., P. Hosman, J.S. Riggs, D. Sluss, S. D. Wullschleger (2006), Direct and indirect effects of atmospheric conditions

and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site, *J. Geophys. Res.*, D16102, doi:10.1029/2006JD007161.

Gurney, K.R., R.M. Law, A.S Denning, P.J. Rayner, D. Baker et al. (2002), Towards robust regional estimates of CO2 sources and sinks using atmospheric transport models, *Nature, 415*, 626-630.

Hargrove, W. W., F. M. Hoffman, and B. E. Law (2003), New analysis reveals representativeness of the AmeriFlux network, *EOS Trans. AGU*, 84, 529-544.

Hollinger, D. Y., S. M. Goltz, E. A. Davidson, J. T. Lee, K. Tu, and H. T. Valentine (1999), Seasonal patterns and environmental control of carbon dioxide and water vapor exchange in an ecotonal boreal forest, *Global Change Biol., 5*, 891-902

Hollinger, D.Y., J. Aber, B. Dail, E. A. Davidson, S. M. Goltz, H. Hughes, M. Y. Leclerc, J. T. Lee, A. D. Richardson, C. Rodrigues, N. A. Scott, D. Achuatavarier, and J. Walsh (2004), Spatial and temporal variability in forest-atmosphere CO2 exchange, *Global Change Biol., 10*, 1689-1706.

Huang, C., and J. R. G. Townshend (2003), A stepwise regression tree for nonlinear approximation: applications to estimating subpixel land cover, *Int. J. Remote Sens., 24*, 75-90.

Huete, A. R., H. Q. Liu, K. Batchily, and W. vanLeeuwen (1997), A comparison of vegetation indices global set of TM images for EOS-MODIS, *Remote Sens. Environ., 59*, 440-451.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira (2002), Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sens. Environ., 83*, 195-213.

Irvine, J., B. E. Law, and K. A. Hibbard (2007), Postfire carbon pools and fluxes in semiarid ponderosa pine in Central Oregon, *Global Change Biol., 13*, 1748-1760.

Jackson, T. J., D. Chen, M. Cosh, F. Li, M. Anderson, C. Walthall, P. Doriaswamy, and E. R. Hunt (2004), Vegetation water content mapping using Landsat data derived normalized difference water index fro corn and soybeans, *Remote Sens. Environ., 92*, 475-482.

Janssens, I. A., A. Freibauer, P. Ciais, P. Smith, G-J. Jabuurs et al. (2003), Europe's terrestrial biosphere absorbs 7 to 12% of European anthropogenic CO_2 emissions, *Science, 300*, 1538-1542.

Law, B. E., M. G. Ryan, and P. M. Anthoni (1999), Seasonal and annual respiration of a ponderosa pine ecosystem, *Global Change Biol., 5*, 169-182.

Law, B. E., O. J. Sun, J. Campbell, S. V. Tuyl, and P. E. Thornton (2003), Changes in carbon storage and fluxes in a chronosequence of ponderosa pine, *Global Change Biol., 9*, 510-524.

Law, B.E., D. Turner, J. Campbell, O. J. Sun, S. Van Tuyl, W. D. Ritts, and W. B. Cohen (2004), Disturbance and climate effects on carbon stocks and fluxes across Western Oregon USA, *Global Change Biol., 10*, 1429-1444.

Law, B.E., D. Turner, M. Lefsky, J. Campbell, M. Guzy, O. Sun, S. Van Tuyl, W. Cohen. 2006. Carbon fluxes across regions: Observational constraints at multiple scales. In J. Wu, B. Jones, H. Li, O. Loucks, eds. *Scaling and Uncertainty Analysis in Ecology: Methods and Applications*. Springer, USA. Pages 167-190.

Lipson, D. A., R. F. Wilson, and W. C. Oechel (2005), Effects of Elevated Atmospheric CO2 on Soil Microbial Biomass, Activity, and Diversity in a Chaparral Ecosystem, *App. Environ. Microbiol., 71*, 8573-8580.

Ma, S., D. D. Baldocchi, L. Xu, and T. Hehn, (2007), Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California, *Agri. For. Meteor., 147,* 157-171.

Magnani, F., M. Mencuccini, M. Borghetti, P. Berbigier, F. Berninger et al. (2007), The human footprint in the carbon cycle of temperate and boreal forests, *Nature, 447*, 849-851.

Mahadevan, P., S. C. Wofsy, D. M. Matross, X. Xiao, A. L. Dunn, J. C. Lin, C. Gerbig, J. W. Munger, V. Y. Chow, and E. W. Gottlieb (2008), A satellite-based biosphere parameterization for net ecosystem CO exchange: vegetation photosynthesis and respiration model (VPRM), *Global Biogeochem. Cyc.*, doi:10.1029/2006GB002735, in press.

McGuire, A. D., et al. (2001), Carbon balance of the terrestrial biosphere in the twentieth century: analyses of CO_2, climate and land-use effects with four process-based ecosystem models, *Global Biogeochem. Cyc., 15*, 183-206.

Monson, R. K., A. A. Turnipseed, J. P. Sparks, P. C. Harley, L. E. Scott-Denton, K. Sparks, T. E. Huxman (2002), Carbon sequestration in a high-elevation, subalpine forest, *Global Change Biol.*, *8,* 459-478.

Moffat, A. M., D. Papale, M. Reichstein, D. Y. Hollinger, A. D. Richardson, A. G. Barr, C. Beckstein, B. H. Braswell, G. Churkina, A. R. Desai, E. Falge, J. H. Gove, M. Heimann, D. Hui, A. J. Jarvis, J. Kattge, A. Noormets, and V. J. Stauch (2007), Comprehensive comparison of gap-filling techniques for eddy covariance net carbon fluxes, *Agri. For. Meteor., 147,* 209-232.

Myneni, R. B., J. Dong, C. J. Tucker, R. K. Kaufmann, P. E. Kauppi et al. (2001), A large carbon sink in the woody biomass of northern forests, *PNAS, 98*, 14784-14789.

Myneni, R. B., S. Hoffman, Y. Knyazikhin, J. L. Privette, Y. Tian et al. (2002), Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, *Remote Sens. Environ., 83*, 214-231.

National Climatic Data Cneter (2008), http://lwf.ncdc.noaa.gov/oa/climate/research/2005/perspectives.html, visited on 01/28/2008.

Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC). 2006. MODIS subsetted land products, Collection 4. Available on-line [http://www.daac.ornl.gov/MODIS/modis.html] from ORNL DAAC, Oak Ridge, Tennessee, U.S.A. Accessed Month 06, 2007.

Oren R., C.-I. Hsieh, P. Stoy, J. Albertson, H. R. McCarthy, P. Harrell, G. G. Katul (2006), Estimating the uncertainty in annual net ecosystem carbon exchange: spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements, *Global Change Biol.,* 12, 883-896.

Oren R., B. E. Ewers, P. Todd, N. Phillips, G. Katul (1998) ,Water balance delineates the soil layer in which moisture affects canopy conductance, *Ecol. App., 8*, 990-1002.

Papale, D., and A. Valentini (2003), A new assessment of European forests carbon exchange by eddy fluxes and artificial neural network spatialization, *Global Change Biol., 9*, 525-535.

Parton, W. J., J. M. O. Scurlock, D. S. Ojima et al. (1993), Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide, *Global Biogeochem. Cyc., 7*, 785-809.

Pataki D. E., R. Oren (2003), Species difference in stomatal control of water loss at the canopy scale in a bottomland deciduous forest, *Adv. Water Res., 26*, 1267-1278.

Penuelas, J., J. A. Gamon, K. L. Griffin, and C. B. Field (1993), Assessing community type, plant biomass, pigment composition, and photosynthetic efficiency of aquatic vegetation from spectral reflectance, *Remote Sens. Environ., 46*, 110-118.

Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klosster (1993), Terrestrial ecosystem production – a process model based on global satellite and surface data, *Global Biogeochem. Cyc., 7*, 811-841.

Powell, T. L., G. Starr, K. L. Clark, T. A. Martin, and H. L. Gholz (2005), Ecosystem and understory water and energy exchange for a mature, naturally regenerated pine flatwoods forest in north Florida, *Can. J. For. Res., 35*, 1568-1580.

Rahman, A. F., Sims, D. A., Cordova, V. D., and El-Masri, B. Z. 2005. Potential of MODIS EVI and surface temperature for directly estimating per-pixel ecosystem C fluxes. *Geophy. Res. Lett., 32,* L19404, doi:10.1029/2005GL024127.

Ranson, K. J., C. S. T. Daughtry, L. L. Biehl, and M. E. Bauer (1985), Sun-view angle effects on reflectance factors of corn canopies, *Remote Sens. Environ., 18*, 147-161.

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

Running, S. W., and E. R. Hunt (1993), Generalization of a forest ecosystem process model for other biomes, Biome-BGC, and an application for global-scale models. Scaling processes

between leaf and landscape levels. In: *Scaling Physiological Processes: Leaf to Globe* (Ehleringer J. R., Field C. B., eds.), pp. 141-158. Academic Press, San Diego.

Running, S. W., D. D. Baldocchi, D. P. Turner, S. T. Gower, P. S. Bakwin, and K. A. Hibbard (1999), A global terrestrial monitoring network integrating tower fluxes, flask sampling, ecosystem modeling and EOS satellite data, *Remote Sens. Environ., 70*, 108-127.

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H. (2004), A continuous satellite-derived measure of global terrestrial primary production, *BioScience, 54*, 547-560.

Salajanu, D., and D. M. Jacobs. 2005. Assessing biomass and forest area classifications from MODIS satellite data while incrementing the number of FIA data panels. In: Pecora 16 "Global Priorities in Land Remote Sensing", October 23-27, 2005, Sioux Falls, South Dakota. 10pp.

Schmid, H. P. (1994), Source areas for scalars and scalar fluxes, *Bound. –Layer Meteor., 67*, 293-318.

Schmid, H. P. (2002), Footprint modeling for vegetation atmospheric exchange studies.: A review and perspective, *Agri. For. Meteor., 113*, 159-183.

Schmid, H. P., C. S. B. Grimmond, F. Cropley, B. Offerle, and H. B. Su (2000), Measurements of CO2 and energy fluxes over a mixed hardwood forest in the mid-western United States, *Agri. For. Meteor., 103*, 357-374.

Schmidt, K. S., and A. K. Skidmore (2003), Spectral discrimination of vegetation types in a coastal wetland, *Remote Sens. Environ., 85*, 92-108.

Sellers, P. J., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., Collatz, G. J., Denning, A. S., Mooney, H. A., Nobre, C. A., Sato, N., Field, C. B., & Henderson-sellers, A. (1997),

Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, *Science, 275*, 502– 509.

Shroyer, J. P., C. Thompson, R. Brown, P. D. Ohlenbach, D. L. Fjell, S. Staggenborg et al. (1996), Kansas crop planting guide. Publication, Vol. L-818 (pp. 2). Manhattan, KS: Kansas State University.

Tans, P. P., I. Y. Fung, and T. Takahashi (1990), Observational constraints on the global atmospheric CO_2 budget, *Science, 247*, 1431-1438.

Tian, H., J. M. Melillo, D. W. Kicklighter, A. D. McGuire, and J. Helfrich (1999), The sensitivity of terrestrial carbon storage to historical climate variability and atmospheric CO_2 in the Untied States, *Tellus, 51B*, 414-452.

Urbanski, S., C. Barford, S. Wofsy, C. Kucharik, E. Pyle, J. Budney, K. McKain, D. Fitzjarrald, M. Czikowsky, J. W. Munger (2007), Factors controlling CO2 exchange on timescales from hourly to decadal at Harvard Forest. *J. Geophys. Res, 112*, Art. No. G02020 MAY 9 2007.

Verma, S.B., A. Dobermann, K.G. Cassman, D.T. Walters, J.M. Knops, T.J. Arkebauer, A.E. Suyker, G.G. Burba, B. Amos, H. Yang, D. Ginting, K.G. Hubbard, A.A. Gitelson, and E.A. Walter-Shea (2005), Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, *Agri. For. Meteor., 131*, 77-96.

Vermote, E. F., and A. Vermeulen (1999), MODIS Algorithm Technical Background Document – Atmospheric Correction Algorithm: Spectral Reflectances (MOD09), Version 4.0. http://modis.gsfc.nasa.gov/data/atbd/atbd_mod08.pdf.

Wan, Z., Y. Zhang, Q. Zhang, and Z.-L. Li (2002), Validation of the land-surface temperature products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data, *Remote Sens. Environ., 83*, 163-180.

Watts, C.J., R. L. Scott, J. Garatuza-Payan, J. C. Rodriguez, J. Prueger, W. Kustas, and M. Douglas (2007), Changes in vegetation condition and surface fluxes during NAME 2004, *J. Clim., 20,* 1810-1820, doi: 10.1175/JCLI4088.1

Wylie, B. K., E. A. Fosnight, T. G. Gilmanov, A. B. Frank, J. A. Morgan et al. 2007. Adaptive data-driven models for estimating carbon fluxes in the Northern Great Plains, *Remote Sensing of Environment, 106*, 399-413.

Xiao, J., and A. Moody (2004a), Trends in vegetation activity and their climatic correlates: China 1982 to 1998, *Int. J. Remote Sens.*, *25*, 5669-5689.

Xiao, J., and A. Moody (2004b), Photosynthetic activity of US biomes: responses to the spatial variability and seasonality of precipitation and temperature, *Global Change Biol., 10*, 437-451.

Xiao, J., A. Moody (2005), A comparison of methods for estimating fractional green vegetation cover within a dsert-to-upland transition zone in central New Mexico, USA, *Remote Sens. Environ., 98*, 237-250.

Xiao, J., Q. Zhuang, E. Liang, A. D. McGuire, D. W. Kicklighter, A. Moody, and J. M. Melillo (2008), Droughts and their impacts on carbon cycling of terrestrial ecosystems during the 20th century, *J. Geophys. Res.*, *revised.*

Xiao, X., Q. Zhang, S. Saleska, L. Hutyra, P. D. Camargo et al. (2005), Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest, *Remote Sens. Environ., 94*, 105-122.

Xu L., D. D. Baldocchi (2004), Seasonal variation in carbon dioxide exchange over a Mediterranean annual grassland in California, *Agri. For. Meteor., 123*, 79-96.

Yamaji, T., T. Sakai, T. Endo, P. J. Baruah, T. Akiyama et al. (2007), Scaling-up technique for net ecosystem productivity of deciduous broadleaved forests in Japan using MODIS data, *Ecol. Res.*, doi:10.1007/s11284-007-0438-0.

Yang, L., C. Huang, C. Homer, B. K. Wylie, and M. J. Coan (2003), An approach for mapping large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial resolution imagery, *Can. J. Remote Sens., 29*, 230-240.

Zhang, X., and S. Kondragunta (2006), Estimating forest biomass in the USA using generalized allometric models and MODIS land products, *Geophys. Res. Lett., 33*, L09402, doi:10.1029/2006GL025879.

Zhao, M., Heinsch, F. A., Nemani, R. R., and Running, S. W. (2005), Improvements of the MODIS terrestrial gross and net primary production global data set, *Remote Sens. Environ., 95*, 164-175.

Zhuang, Q., A. D. McGuire, K. P. O'Neill, J. W. Harden, V. E. Romanovsky, and J. Yarie. (2002), Modeling the soil thermal and carbon dynamics of a fire chronosequence in Interior Alaska, *J. Geophy. Res.*, *107*, 8147, doi:10.1029/2001JD001244, 2002. [Printed 108(D1), 2003]

Zhuang, Q., A. D. McGuire, J. M. Melillo, J. S. Clein, R. J. Dargaville, D. W. Kicklighter, R. B. Myneni, J. Dong, V. E. Romanovsky, J. Harden, and J. E. Hobbie (2003), Carbon cycling in extratropical terrestrial ecosystems of the Northern Hemisphere during the 20th Century: A modeling analysis of the influences of soil thermal dynamics, *Tellus*, *55B*, 751-776.

Figure captions:

Figure 1. Location and spatial distribution of the AmeriFlux sites used in this study. The base map is the reclassified MODIS land-cover map that was used to for the continental-scale estimation of NEE.

Figure 2. Examples of the splitting of NEE data into a training set (2000-2004) and a validation set (2005-2006): (a) Harvard Forest EMS Tower (MA); (b) FNAL Agricultural Site (IL); (c) Fort Peck (MT).

Figure 3. Scatterplot of observed 8-day NEE versus predicted 8-day NEE. The solid line is the 1:1 line.

Figure 4. Scatterplot of predicted 8-day NEE versus residuals (observed - predicted) over the period 2005-2006.

Figure 5. The average error and relative error across all AmeriFlux sites for each 8-day period.

Figure 6. Observed (red line) and predicted (green line) 8-day NEE (g C m^{-2} day^{-1}) for each AmeriFlux site over the period 2005-2006.

Figure 7. Scatterplot of observed mean NEE versus predicted mean NEE across the AmeriFlux sites. The abbreviations of these sites are given in Table 1.

Figure 8. Scatterplot of observed mean NEE versus predicted mean NEE across vegetation types: EF - evergreen forests; DF - deciduous forests; MF - mixed forests; Sh - shrublands; Sa - savannas; Gr - grasslands; Cr – Croplands.

Figure 9. Predicted NEE for four 8-day periods in 2005: (a) March 14 - 21; (b) July 12 - 19; (c) September 14 - 21; (d) December 11 - 18. Positive values indicate carbon release, whereas negative values indicate carbon sequestration.

Figure 10. Predicted NEE for each season in 2005: (a) spring (March-May); (b) summer (June-August); (c) fall (September-November); (d) winter (December-February). Positive values indicate carbon release, whereas negative values indicate carbon sequestration.

Figure 11. Estimated 8-day NEE for each vegetation type in 2005. (a) Mean 8-day NEE (g C m^{-2} day^{-1}); (b) Total 8-day NEP (Tg C day^{-1}).

Figure 12. The area of each vegetation type across the conterminous U. S.

Table 1. Site descriptions including name, latitude, longitude, vegetation structure, years of data available, and references for each flux site in this study.

Site	State	Lat	Lon	Vegetation structure	Vegetation type	Year	References
Audubon Research Ranch (ARR)	AZ	31.59	-110.51	Desert grasslands	Grasslands	2002-2006	
Santa Rita Mesquite (SRM)	AZ	31.82	-110.87	Mesquite-dominated savanna	Savannas	2004-2006	Watts et al., 2007
Walnut Gulch Kendall Grasslands (WGK)	AZ	31.74	-109.94	Warm season C4 grassland	Grasslands	2004-2006	
Sky Oaks Old Stand (SOO)	CA	33.37	-116.62	Chaparral (Mediterranean-type ecosystems)	Shrublands	2004-2006	Lipson et al., 2005
Sky Oaks Young stand (SOY)	CA	33.38	-116.62	Chaparral (Mediterranean-type ecosystems)	Shrublands	2001-2006	Lipson et al., 2005
Tonzi Ranch (TR)	CA	38.43	-120.97	Oak savanna, grazed grassland dominated by blue oak and grasses	Savannas	2001-2006	Ma et al., 2007
Vaira Ranch (VR)	CA	38.41	-120.95	Grazed C3 grassland opening in a region of oak/grass savanna	Grasslands	2001-2006	Xu et al., 2004
Niwot Ridge Forest (NRF)	CO	40.03	-105.55	Subalpine coniferous forest dominated by subalpine, Engelmann spruce, and lodgepole pine	Evergreen forests	2000-2003	Monsoon et al., 2002
Kennedy Space Center - Scrub Oak (KSC)	FL	28.61	-80.67	Scrub-oak palmetto dominated by schlerophyllous evergeen oaks and the Saw Palmetto Serenoa repens	Shrublands	2000-2006	Dore et al., 2003
Austin Cary - Slash Pine (AC)	FL	29.74	-82.22	Slashpine dominated by Pinus palustris/Pinus ellottii	Evergreen forests	2001-2005	Powell et al., 2005
Bondville (Bon)	IL	40.01	-88.29	Annual rotation between corn (C4) and soybeans (C3)	Croplands	2001-2006	
FNAL agricultural site (FAg)	IL	41.86	-88.22	Soybean/corn	Croplands	2005-2006	
FNAL Prairie site (FPr)	IL	41.84	-88.24	Tall grass prairie	Grasslands	2004-2006	
Morgan Monroe State Forest (MMS)	IN	39.32	-86.41	Mixed hardwood deciduous forest dominated by sugar maple, tulip poplar, sassafras, white oak, and black oak	Deciduous forests	2000-2005	Schmid et al., 2000
Harvard Forest EMS Tower (HFE)	MA	42.54	-72.17	Temperate deciduous forest dominated by red oak, red maple, black birch, white pine, and hemlock	Deciduous forests	2000-2004	Urbanski et al. 2007
Harvard Forest Hemlock Site (HFH)	MA	42.54	-72.18	Temperate coniferous forest dominated by hemlock	Evergreen forests	2004	
Little Prospect Hill (LPH)	MA	42.54	-72.18	Temperate deciduous forest dominated by red oak,	Deciduous forests	2002-2005	

				red maple, black birch, white pine, and hemlock			
Howland forest (HF)	ME	45.20	-68.74	Boreal--northern hardwood transitional forest consisting of hemlock-spruce-fir, aspen-birch, and hemlock-hardwood mixtures	Evergreen forests	2000-2004	Hollinger et al., 1999, 2004
Howland forest (west tower) (HFW)	ME	45.21	-68.75	Deciduous needle forest, Boreal/northern hardwood ecoton, old coniferous	Deciduous forests	2000-2004	Hollinger et al., 1999, 2004
Sylvania Wilderness Area (SWA)	MI	46.24	-89.35	Old-growth eastern hemlock/sugar maple/basswood/yellow birch	Mixed forests	2002-2006	
Univ. of Mich. Biological Station (UMB)	MI	45.56	-84.71	Mid-aged conifer and deciduous, northern hardwood, pine understay,aspen, mostly deciduous, old growth hemlock	Mixed forests	2000-2003	Gough et al., 2007
Missouri Ozark (MO)	MO	38.74	-92.20	Oak hickory forest	Deciduous forests	2004-2006	Gu et al. 2006, 2007
Goodwin Creek (GC)	MS	34.25	-89.97	Temperate grassland	Grasslands	2002-2006	
Fort Peck (FPe)	MT	48.31	-105.10	Grassland	Grasslands	2000-2006	
Duke Forest loblolly pine (DFP)	NC	35.98	-79.09	Even-aged loblolly pine forest	Evergreen forests	2001-2005	Oren et al. 1998, 2006
Duke Forest hardwoods (DFH)	NC	35.97	-79.10	An uneven-aged closed-canopy stand in an oak-hickory type forest composed of mixed hardwood species with pine (P. taeda) as a minor component	Deciduous forests	2003-2005	Pataki and Oren, 2003
North Carolina loblolly pine (NCP)	NC	35.80	-76.67	Loblolly pine plantation	Evergreen forests	2005-2006	
Mead -irrigated continuous maize site (MIC)	NE	41.17	-96.48	Continuous maize	Croplands	2001-2005	Verma et al. 2005
Mead irrigated rotation (MIR)	NE	41.16	-96.47	Maize-soybean rotation	Croplands	2001-2005	Verma et al. 2005
Mead rainfed (MR)	NE	41.18	-96.44	Maize-soybean rotation	Croplands	2001-2005	Verma et al. 2005
Bartlett Experimental Forest (BEF)	NH	44.06	-71.29	Temperate northern hardwood forest dominated by American beech, red maple, paper birch, and hemlock	Deciduous forests	2004-2005	
Toledo Oak Openings (TOP)	OH	41.55	-83.84	Oak Savannah dominated by quercus rebrua, quercus alba, and acer rubrum	Savannas	2004-2005	
ARM Oklahoma (ARM)	OK	36.61	-97.49	Winter wheat, some pasture and summer crops	Croplands	2003-2006	
Metolius intermediate aged ponderosa pine (MI)	OR	44.45	-121.56	Temperate coniferous forest dominated by pinus ponderosa, purshia tridentate, arctostaphylos patula	Evergreen forests	2003-2005	Law et al. 2003; Irvine et al. 2007
Metolius new young pine (MN)	OR	44.32	-121.61	Temperate coniferous forest dominated by pinus ponderosa and purshia tridentata	Evergreen forests	2004-2005	Law et al. 2003; Irvine et

						al. 2007
Brookings (Bro)	SD	44.35	-96.84	Temperate grassland	Grasslands	2004-2006
Freeman Ranch Mesquite Juniper (FRM)	TX	29.95	-98.00	Grassland in transition to an Ashe juniper-dominated woodland	Savannas	2004-2006
Wind River Crane Site (WRC)	WA	45.82	-121.95	Temperate coniferous forest dominated by douglas-fir and western hemlock	Evergreen forests	2000-2004
Lost Creek (LC)	WI	46.08	-89.98	Alder-willow deciduous wetland	Deciduous forests	2000-2005
Willow Creek (WC)	WI	45.81	-90.08	Temperate/Boreal forest dominated by white ash, sugar maple, basswood, green ask, and red oak	Deciduous forests	2000-2006
Wisconsin intermediate hardwood (WIH)	WI	46.73	-91.23		Deciduous forests	2003
Wisconsin mature red pine (MRP)	WI	46.74	-91.17		Evergreen forests	2002-2005

Descriptions on vegetation structures are from the site information available at http://public.ornl.gov/ameriflux/ for all sites except Duke Forest - hardwoods. The description on the vegetation for Duke Forest - hardwoods is from http://www.env.duke.edu/other/AMERIFLUX/hwsite.html.

Table 2. The seven broader vegetation types used in the study.

Vegetation types	IGBP class	Definition [Belward and Loveland, 1996]
Evergreen forests	Evergreen needleleaf forests (1), evergreen broadleaf forests (2)	Tree canopy cover > 60% and tree height > 2m. Most of the canopy remains green all year
Deciduous forests	Deciduous needleleaf forests (3), deciduous broadleaf forests (4)	Tree canopy cover > 60% and tree height > 2m. Most of the canopy is deciduous
Mixed forests	Mixed forests (5)	Tree canopy cover > 60% and tree height > 2m. Mixed evergreen and deciduous canopy
Shrublands	Closed shrublands (6), open shrublands (7)	Shrub canopy cover > 10% (10-60% for open shrublands, >60% for closed shrublands) and height < 2m
Savannas	Woody savannas (8), savannas (9)	Forest canopy cover between 10-60% (30-60% for woody savannas, 10-30% for savannas) and height > 2m
Grasslands	Grasslands (10)	Herbaceous cover. Woody cover < 10%
Croplands	Croplands (12)	Temporary crops followed by harvest and a bare soil period

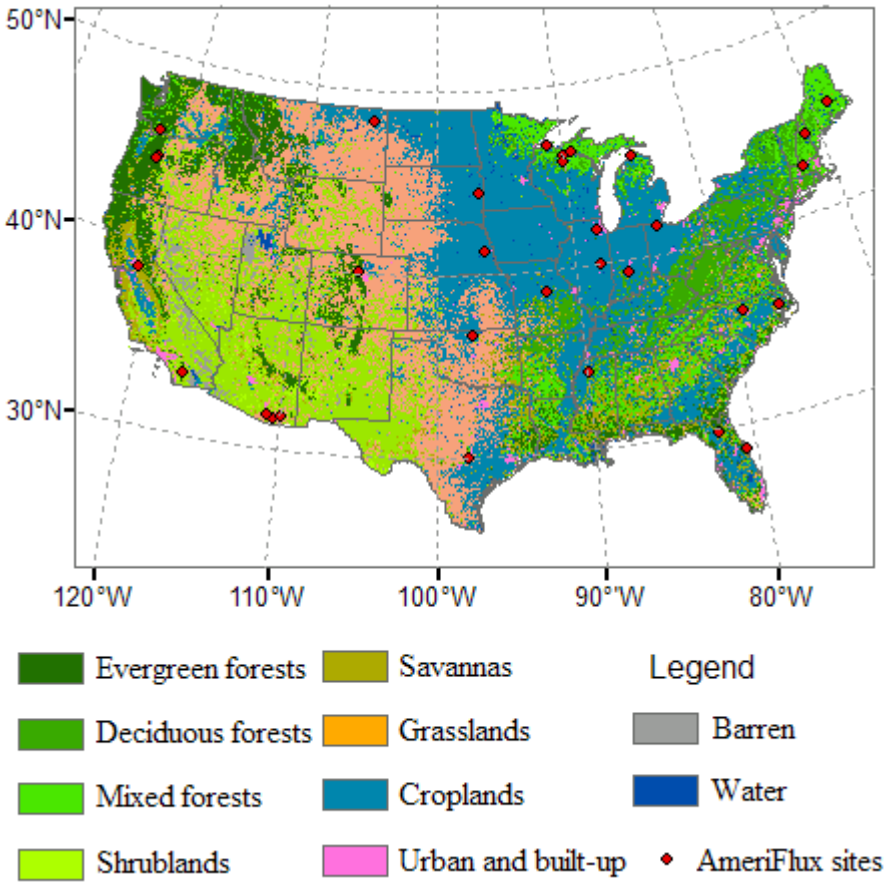


Figure 1. Location and spatial distribution of the AmeriFlux sites used in this study. The base map is the reclassified MODIS land-cover map that was used to for the continental-scale estimation of NEE.

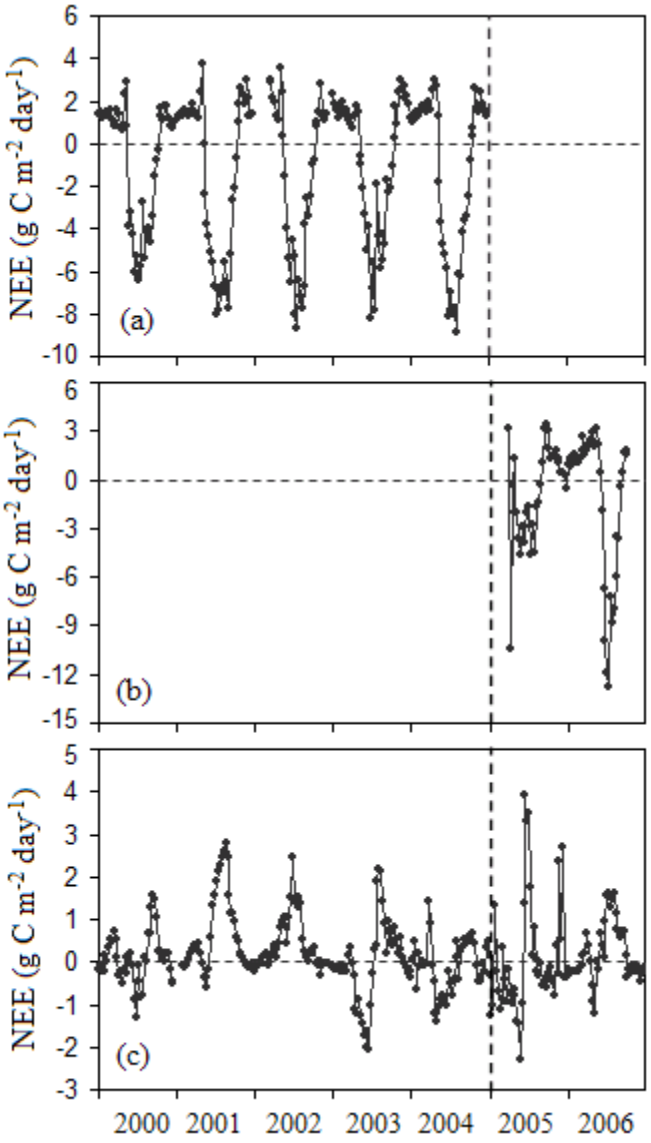


Figure 2. Examples of the splitting of NEE data into a training set (2000-2004) and a validation set (2005-2006): (a) Harvard Forest EMS Tower (MA); (b) FNAL Agricultural Site (IL); (c) Fort Peck (MT).

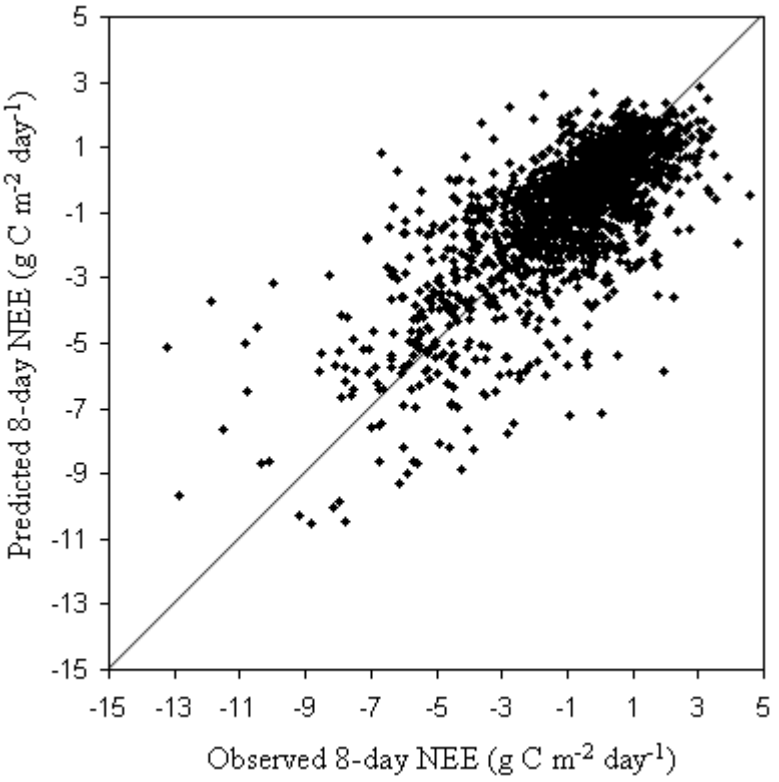


Figure 3. Scatterplot of observed 8-day NEE versus predicted 8-day NEE. The solid line is the 1:1 line.

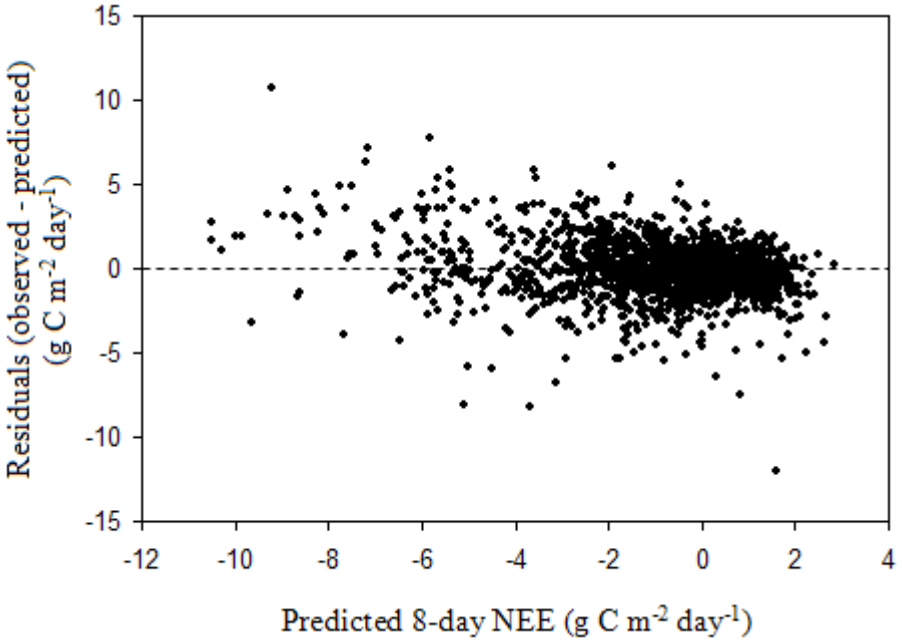


Figure 4. Scatterplot of predicted 8-day NEE versus residuals (observed - predicted) over the period 2005-2006.

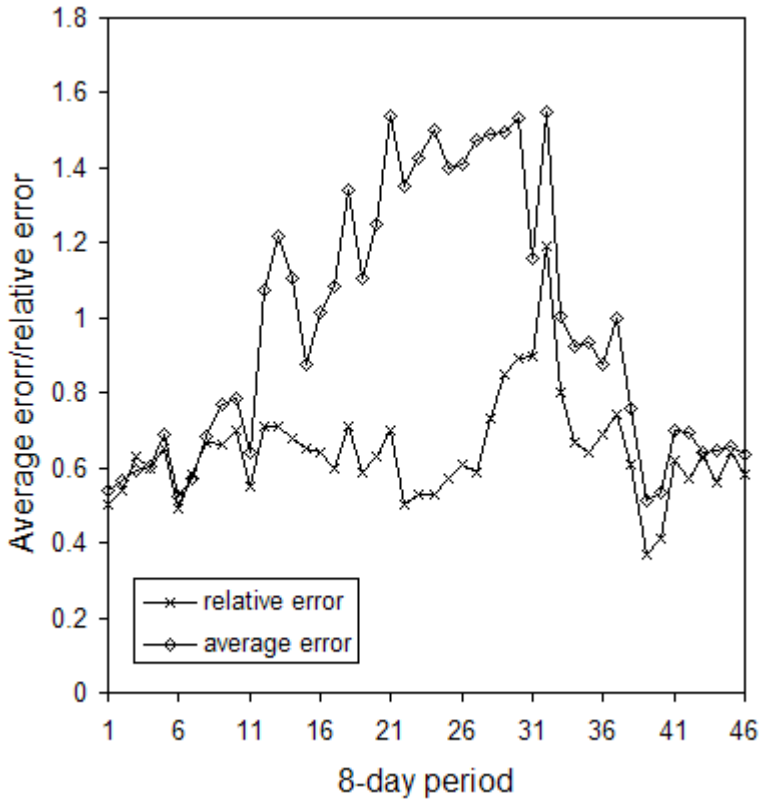
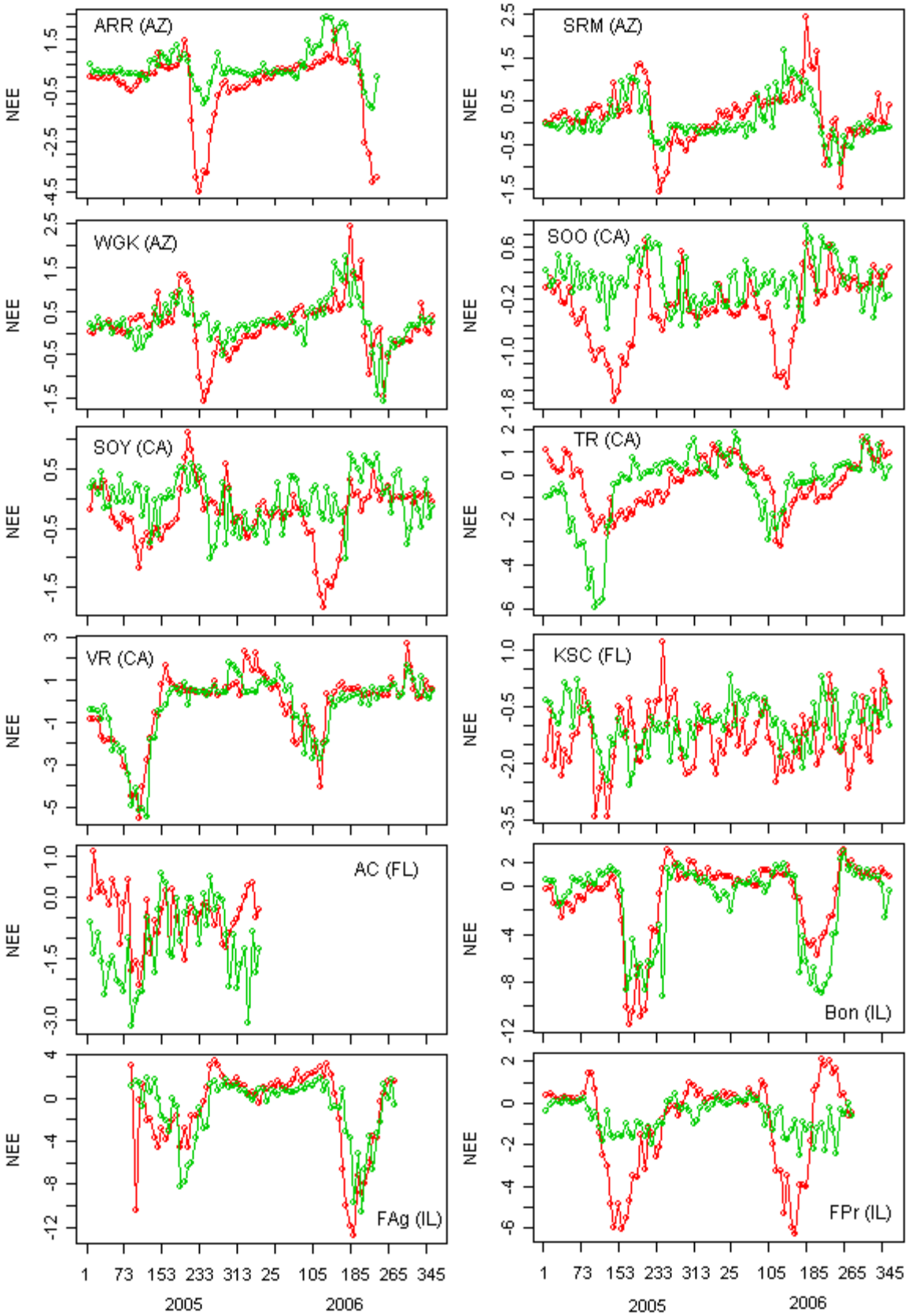
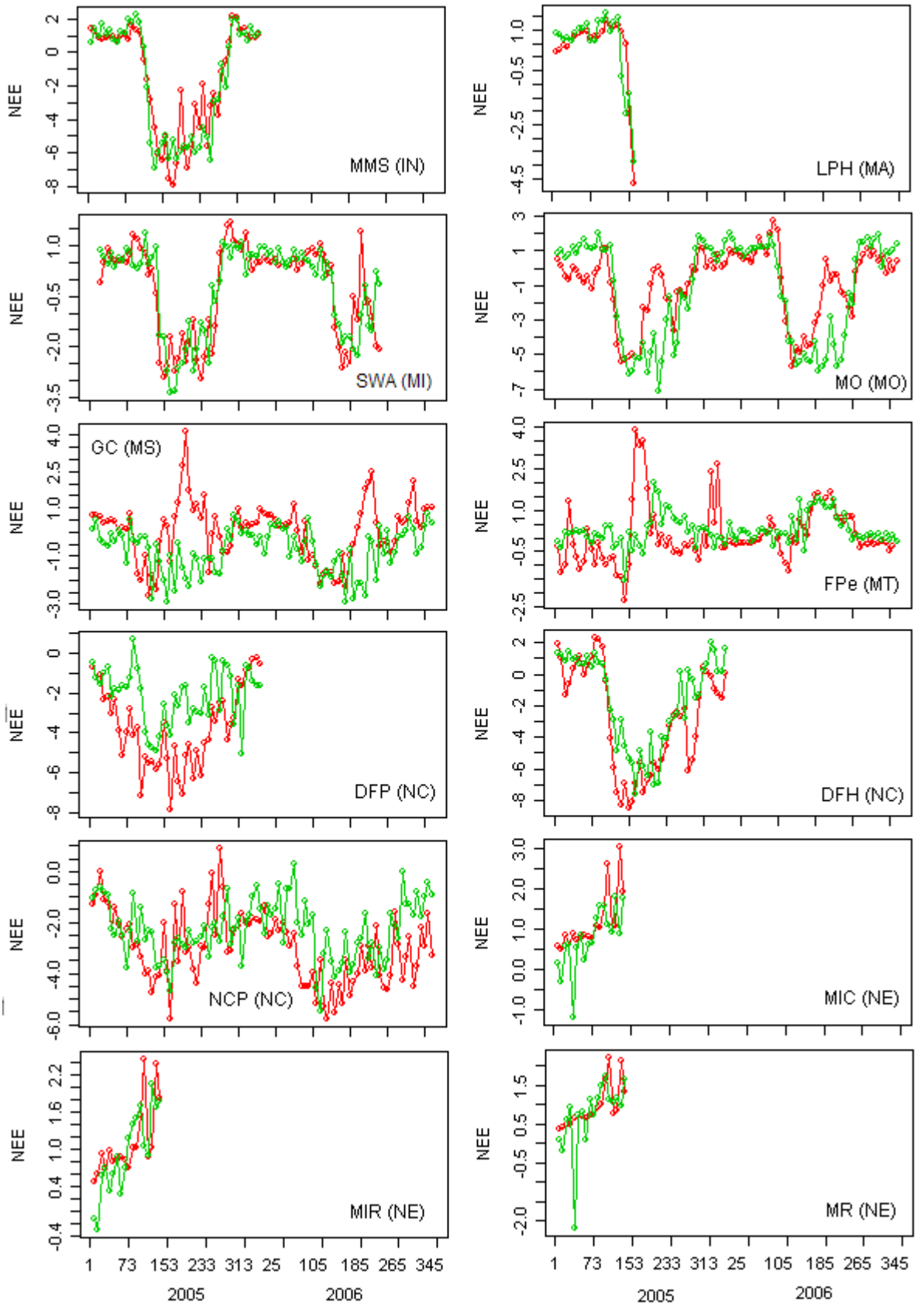


Figure 5. The average error and relative error across all AmeriFlux sites for each 8-day period.


ARR (AZ)
SRM (AZ)
WGK (AZ)
SOO (CA)
SOY (CA)
TR (CA)
VR (CA)
KSC (FL)
AC (FL)
Bon (IL)
FAg (IL)
FPr (IL)
NEE
1 73 153 233 313 25 105 185 265 345
2005
2006


MMS (IN)
LPH (MA)
SWA (MI)
MO (MO)
GC (MS)
FPe (MT)
DFP (NC)
DFH (NC)
NCP (NC)
MIC (NE)
MIR (NE)
MR (NE)
NEE
1 73 153 233 313 25 105 185 265 345
2005
2006

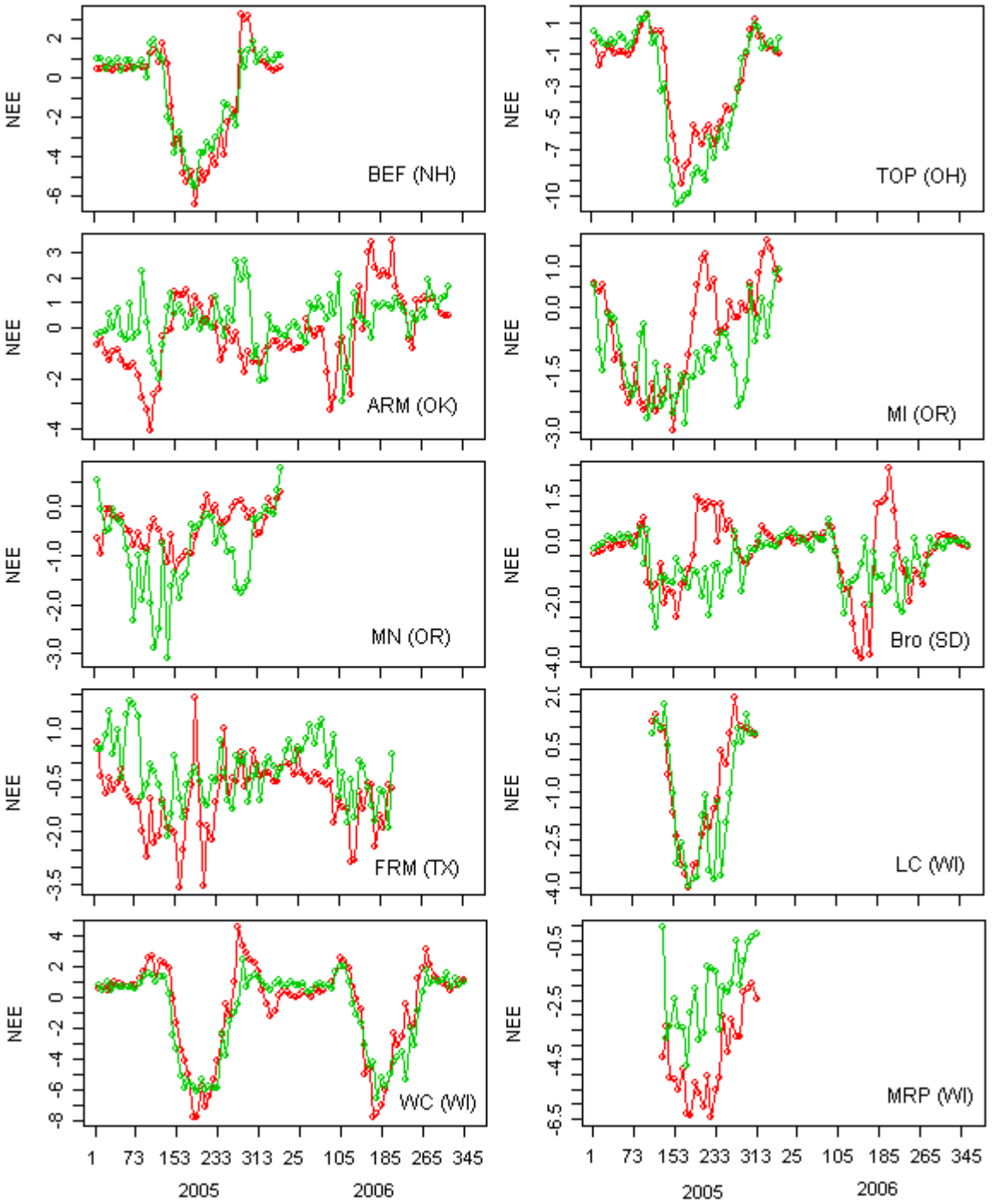


Figure 6. Observed (red line) and predicted (green line) 8-day NEE (g C m^{-2} day^{-1}) for each AmeriFlux site over the period 2005-2006.

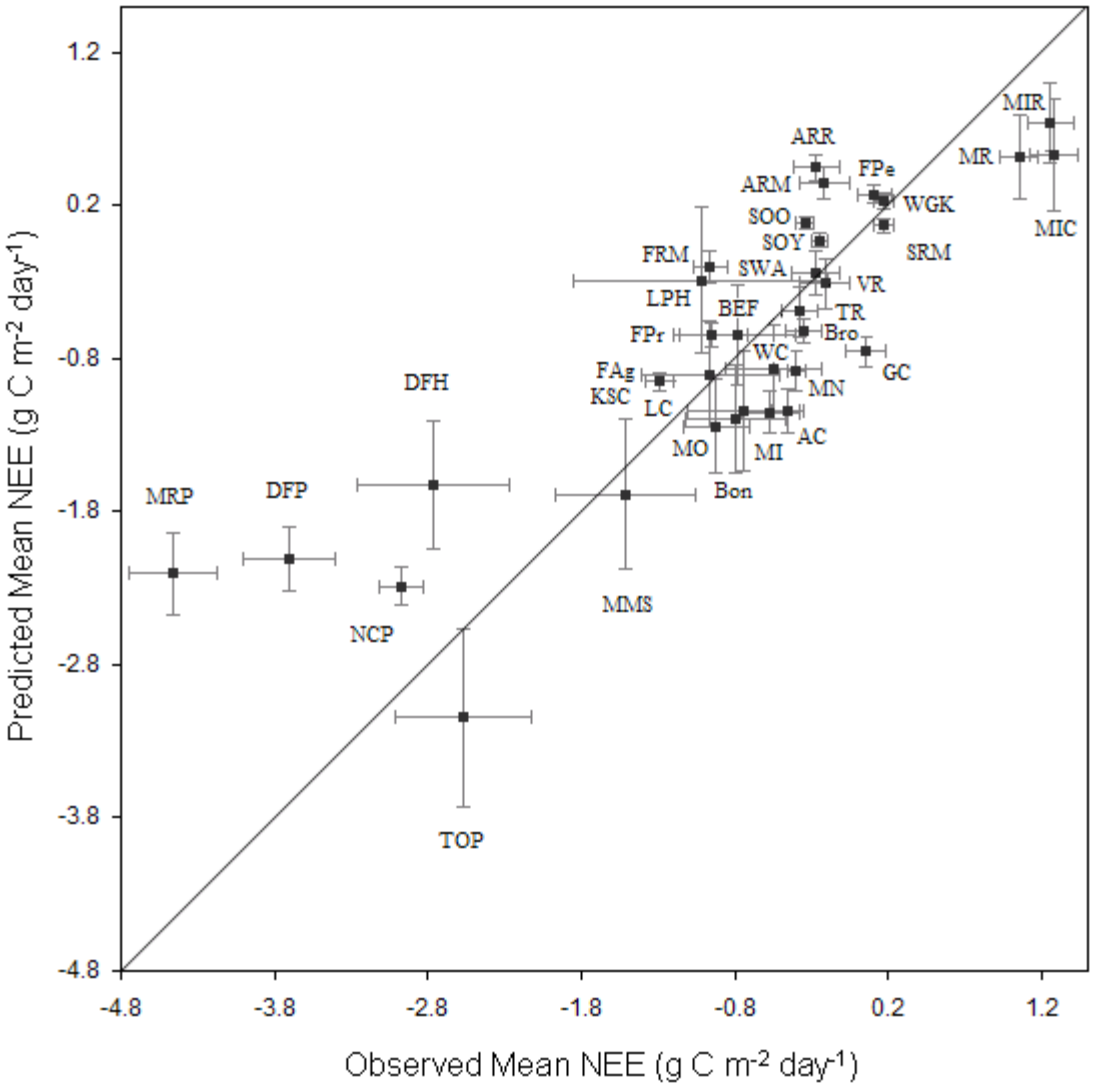


Figure 7. Scatterplot of observed mean NEE versus predicted mean NEE across the AmeriFlux sites. The abbreviations of these sites are given in Table 1.

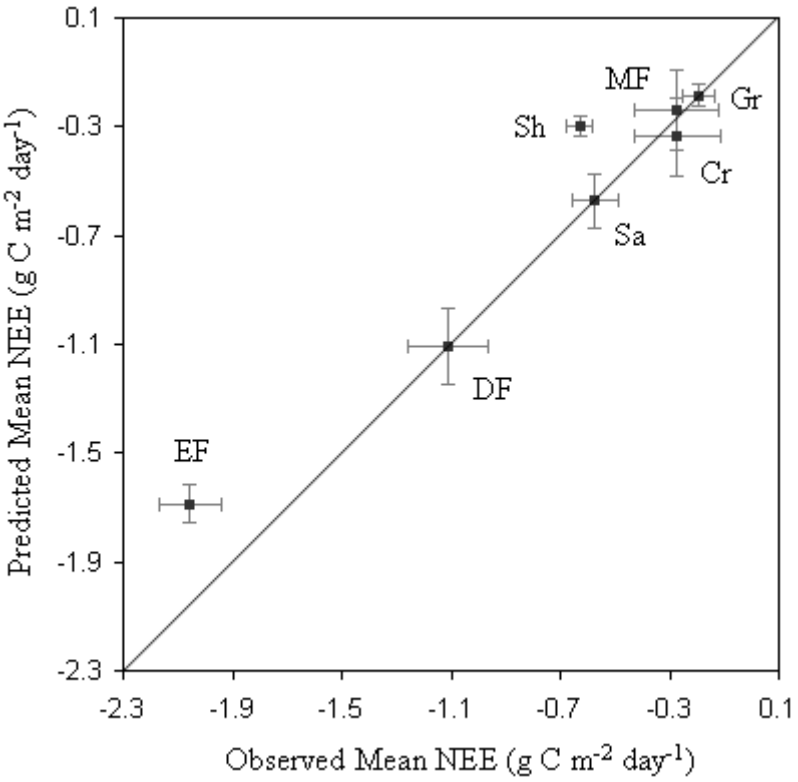


Figure 8. Scatterplot of observed mean NEE versus predicted mean NEE across vegetation types: EF - evergreen forests; DF - deciduous forests; MF - mixed forests; Sh - shrublands; Sa - savannas; Gr - grasslands; Cr – Croplands.

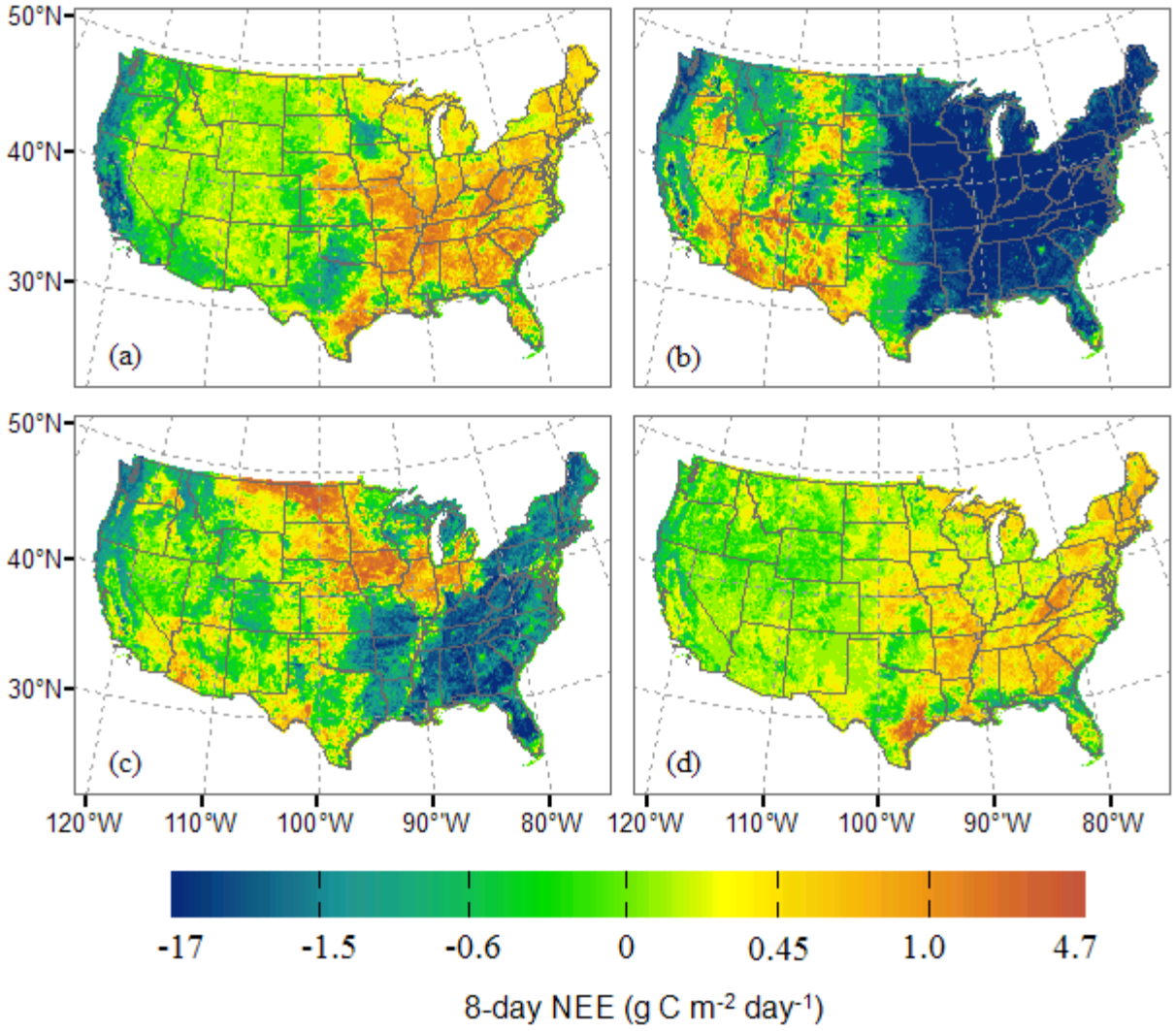


Figure 9. Predicted NEE for four 8-day periods in 2005: (a) March 14 - 21; (b) July 12 - 19; (c) September 14 - 21; (d) December 11 - 18. Positive values indicate carbon release, whereas negative values indicate carbon sequestration.

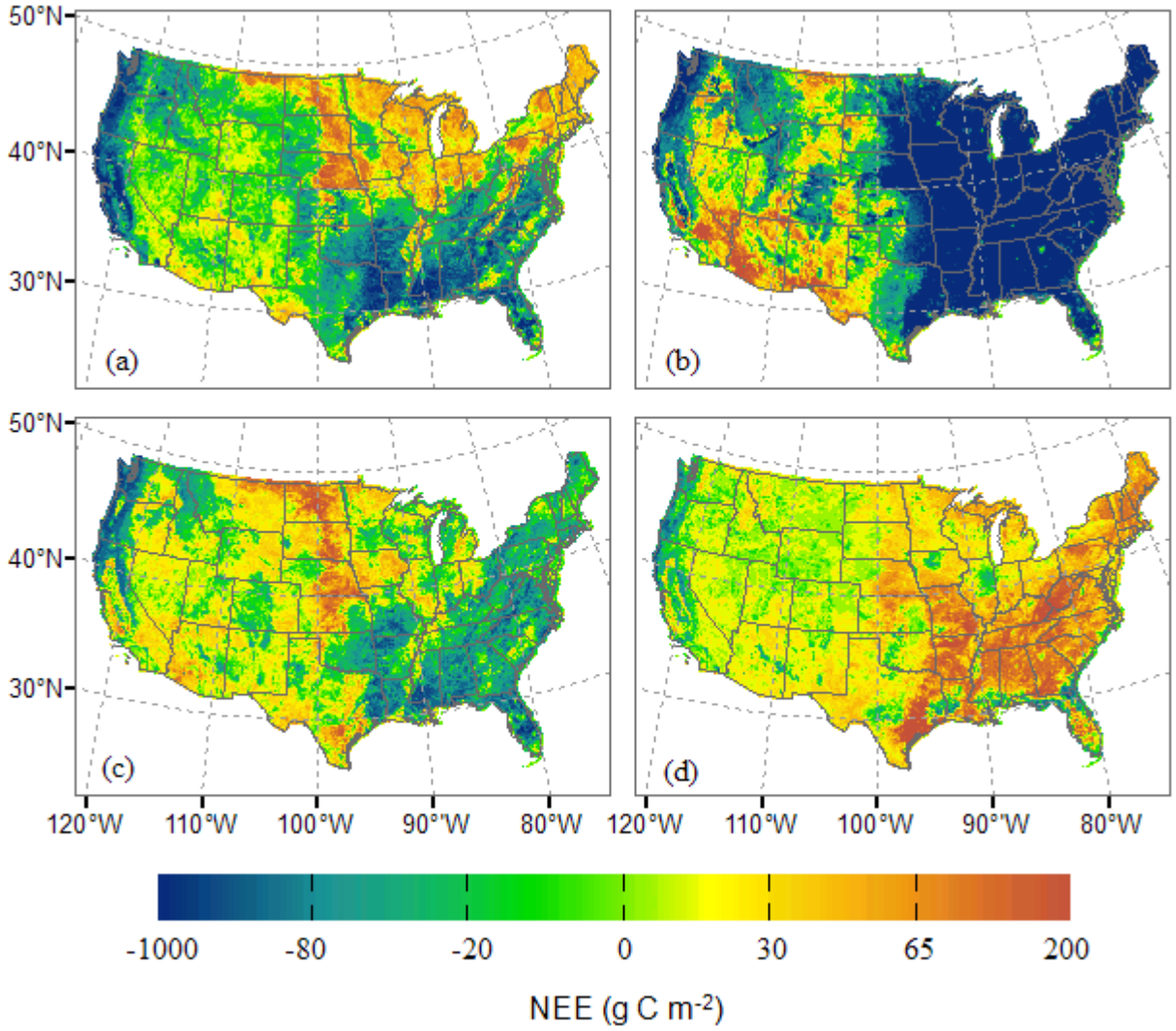


Figure 10. Predicted NEE for each season in 2005: (a) spring (March-May); (b) summer (June-August); (c) fall (September-November); (d) winter (December-February). Positive values indicate carbon release, whereas negative values indicate carbon sequestration.

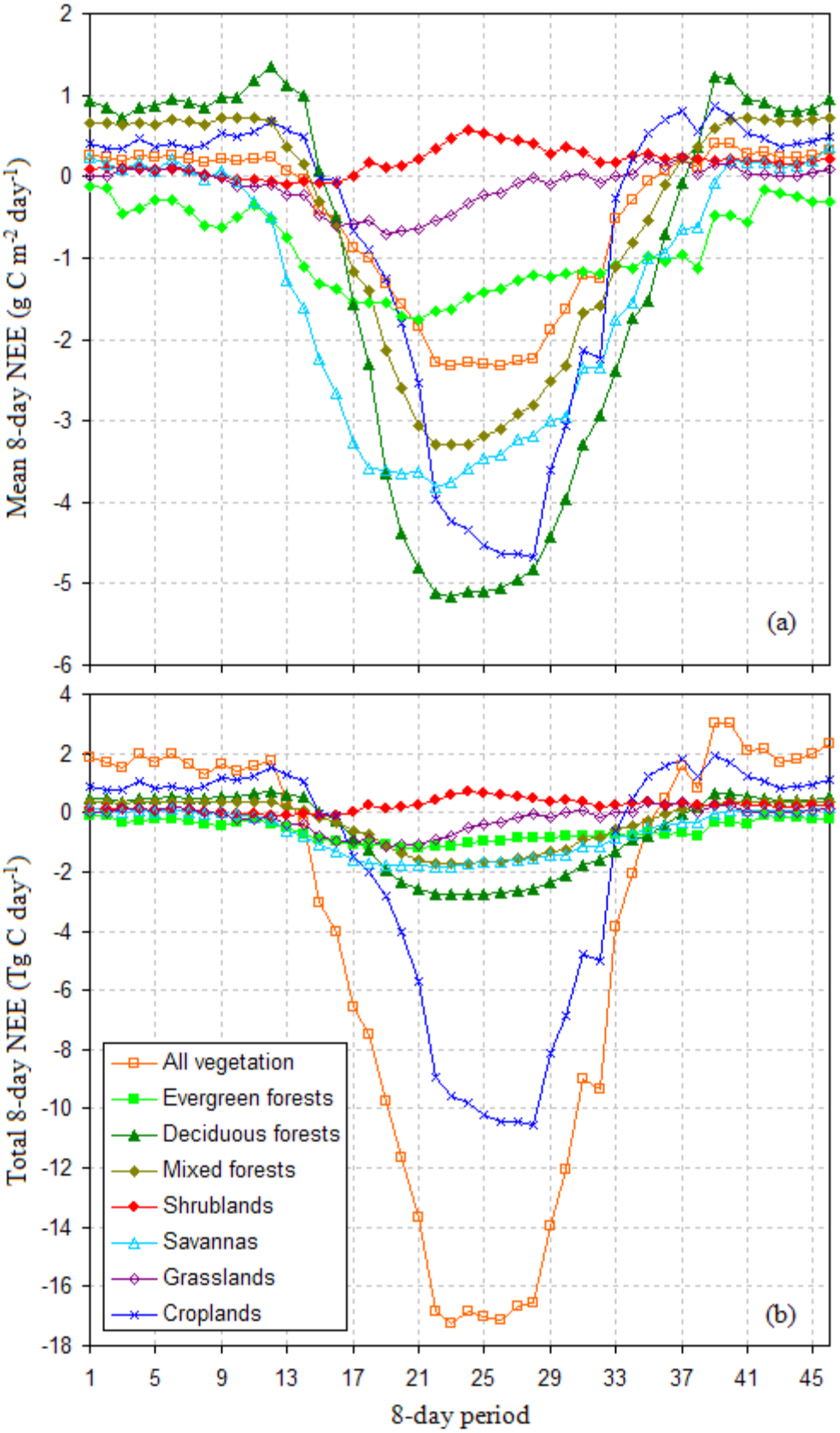

Mean 8-day NEE (g C m-2 day-1)
(a)
Total 8-day NEE (Tg C day-1)
(b)
All vegetation
Evergreen forests
Deciduous forests
Mixed forests
Shrublands
Savannas
Grasslands
Croplands
8-day period

Figure 11. Estimated mean and total 8-day NEE for each vegetation type in the conterminous U.S. in 2005. (a) Mean 8-day NEE (g C m^{-2} day^{-1}); (b) Total 8-day NEE (Tg C day^{-1}).

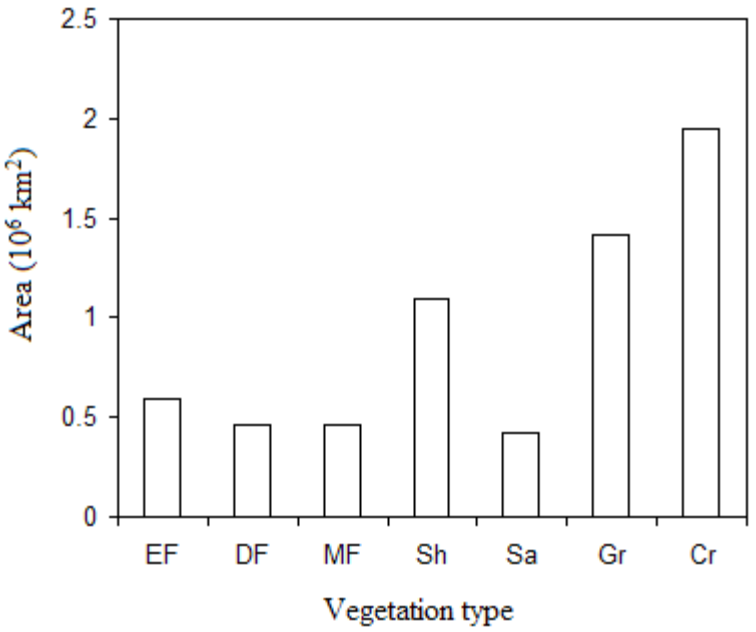


Figure 12. The area of each vegetation type across the conterminous U. S.