A Continuous Measure of Gross Primary Production for the Conterminous U.S. Derived from MODIS and AmeriFlux Data

Jingfeng Xiao, Qianlai Zhuang, Beverly E. Law, Jiquan Chen, Dennis D. Baldocchi, David R. Cook, Ram Oren, Andrew D. Richardson, Sonia Wharton, Siyan Ma, Timothy A. Martin, Shashi B. Verma, Andrew E. Suyker, Russell L. Scott, Russell K. Monson, Marcy Litvak, David Y. Hollinger, Ge Sun, Kenneth J. Davis, Paul V. Bolstad, Sean P. Burns, Peter S. Curtis, Bert G. Drake, Matthias Falk, Marc L. Fischer, David R. Foster, Lianhong Gu, Julian L. Hadley, Gabriel G. Katul, Roser Matamala, Steve McNulty, Tilden P. Meyers, J. William Munger, Asko Noormets, Walter C. Oechel, Kyaw Tha Paw U, Hans Peter Schmid, Gregory Starr, Margaret S. Torn, Steven C. Wofsy

This work was supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

A Continuous Measure of Gross Primary Production for the Conterminous U.S. Derived from MODIS and AmeriFlux Data

Jingfeng Xiao[1*], Qianlai Zhuang[2], Beverly E. Law[3], Jiquan Chen[4], Dennis D. Baldocchi[5], David R. Cook[6], Ram Oren[7], Andrew D. Richardson[8], Sonia Wharton[9], Siyan Ma[5], Timothy A. Martin[10], Shashi B. Verma[11], Andrew E. Suyker[11], Russell L. Scott[12], Russell K. Monson[13], Marcy Litvak[14], David Y. Hollinger[15], Ge Sun[16], Kenneth J. Davis[17], Paul V. Bolstad[18], Sean P. Burns[13], Peter S. Curtis[19], Bert G. Drake[20], Matthias Falk[9], Marc L. Fischer[21], David R. Foster[22], Lianhong Gu[23], Julian L. Hadley[24], Gabriel G. Katul[7], Roser Matamala[25], Steve McNulty[16], Tilden P. Meyers[26], J. William Munger[27], Asko Noormets[28], Walter C. Oechel[29], Kyaw Tha Paw U[9], Hans Peter Schmid[30], Gregory Starr[31], Margaret S. Torn[32], Steven C. Wofsy[33]

[1]Department of Earth & Atmospheric Sciences, Purdue Climate Change Research Center, Purdue University, West Lafayette, IN 47907, USA

[2]Department of Earth & Atmospheric Sciences, Department of Agronomy, Purdue Climate Change Research Center, Purdue University, West Lafayette, IN 47907, USA

[3]College of Forestry, Oregon State University, Corvallis, OR 97331, USA

[4]Department of Environmental Sciences, University of Toledo, Toledo, OH 43606, USA

[5]Ecosystem Science Division, Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA 94720, USA

[6]Argonne National Laboratory, Environmental Science Division, Argonne, IL 60439, USA

[7]Nicholas School of the Environment, Duke University, Durham, NC 27708, USA

[8]Complex Systems Research Center Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, NH 03824, USA

[9]Department of Land, Air and Water Resources, University of California, Davis, Davis, CA 95616, USA

[10]School of Forest Resources & Conservation, University of Florida, Gainesville, FL 32611, USA

[11]School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE 68583, USA

[12]USDA-ARS Southwest Watershed Research Center, Tucson, AZ 85719, USA

[13]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA

[14]Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA

[15]USDA Forest Service, Northeastern Research Station, Durham, NH 03824, USA

[16]USDA Forest Service, Southern Research Station, Raleigh, NC 27606, USA

[17]Department of Meteorology, Pennsylvania State University, University Park, PA 16802, USA

[18]Department of Forest Resources, University of Minnesota, St. Paul, MN 55108, USA

[19]Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, OH 43210, USA

[20]Smithsonian Environmental Research Center, Edgewater, MD 21037, USA

[21]Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division, Atmospheric Science Department, Berkeley, CA 94720, USA

[22]Harvard Forest and Department of Organismic and Evolutionary Biology, Harvard University, Petersham, MA 01366, USA

[23]Oak Ridge National Laboratory Environmental Sciences Division, Oak Ridge, TN 37831, USA

[24]Harvard Forest, Harvard University, Petersham, MA 01366, USA

[25]Argonne National Laboratory, Biosciences Division, Argonne, IL 60439, USA

[26]NOAA/ARL, Atmospheric Turbulence and Diffusion Division, Oak Ridge, TN 37831, USA

[27]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA 02138, USA

[28]Department of Forestry and Environmental Resources and Southern Global Change Program, North Carolina State University, Raleigh, NC 27695, USA

[29]Department of Biology, San Diego State University, San Diego, CA 92182, USA

[30]Department of Geography, Indiana University, Bloomington, IN 47405, USA; Atmospheric Environmental Research, Institute of Meteorology and Climate Research, Research Center Karlsruhe (FZK/IMK-IFU), Kreuzeckbahnstr, 19, 82467 Garmisch-Partenkirchen, Germany

[31]Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA

[32]Lawrence Berkeley National Laboratory, Earth Science Division, Berkeley, CA 94720, USA

[33]Division of Engineering and Applied Science/Department of Earth and Planetary Science, Harvard University, Cambridge, MA 02138, USA

*Corresponding author: jing@psu.edu. Now at Department of Meteorology, Pennsylvania State University, University Park, PA 16802

Keywords: Gross Primary Productivity; MODIS; AmeriFlux; Eddy Covariance; Regression Tree; US; Carbon Fluxes; Interannual Variability; Satellite Data; Biomes

Submitted to *Remote Sensing of Environment* on 01/28/2009

Revised: 10/08/2009

Abstract The quantification of carbon fluxes between the terrestrial biosphere and the atmosphere is of scientific importance and also relevant to climate-policy making. Eddy covariance flux towers provide continuous measurements of ecosystem-level exchange of carbon dioxide spanning diurnal, synoptic, seasonal, and interannual time scales. However, these measurements only represent the fluxes at the scale of the tower footprint. Here we used remotely-sensed data from the Moderate Resolution Imaging Spectroradiometer (MODIS) to upscale gross primary productivity (GPP) data from eddy covariance flux towers to the continental scale. We first combined GPP and MODIS data for 42 AmeriFlux towers encompassing a wide range of ecosystem and climate types to develop a predictive GPP model using a regression tree approach. The predictive model was trained using observed GPP over the period 2000-2004, and was validated using observed GPP over the period 2005-2006 and leave-one-out cross-validation. Our model predicted GPP fairly well at the site level. We then used the model to estimate GPP for each 1 km × 1 km cell across the U.S. for each 8-day interval over the period from February 2000 to December 2006 using MODIS data. Our GPP estimates provide a spatially and temporally continuous measure of gross primary production for the U.S. that is a highly constrained by eddy covariance flux data. Our study demonstrated that our empirical approach is effective for upscaling eddy flux GPP data to the continental scale and producing continuous GPP estimates across multiple biomes. With these estimates, we then examined the patterns, magnitude, and interannual variability of GPP. We estimated a gross carbon uptake between 6.91 and 7.33 Pg C yr^{-1} for the conterminous U.S. Drought, fires, and hurricanes reduced annual GPP at regional scales and could have a significant impact on the U.S. net ecosystem carbon exchange. The sources of the interannual variability of U.S. GPP were dominated by these extreme climate events and disturbances.

92 **1. Introduction**

93 The quantification of ecosystem carbon fluxes for regions, continents, or the globe can
94 improve our understanding of the feedbacks between the terrestrial biosphere and the
95 atmosphere in the context of global change and facilitate climate policy decisions (Law et al.
96 2006). Gross primary productivity (GPP) is the amount of carbon fixed by vegetation through
97 photosynthesis and a key component of ecosystem carbon fluxes and the carbon balance
98 between the biosphere and the atmosphere (Mäkelä et al. 2008). The accurate estimation of
99 GPP is essential for the quantification of net ecosystem carbon exchange (NEE) as the latter is
100 often a small difference of two large carbon fluxes – GPP and ecosystem respiration (R_e). The
101 estimation of GPP for regions, continents, or the globe, however, can only be made by using
102 ecosystem models (e.g., Prince and Goward 1995) and/or remotely sensed data (e.g., Running
103 et al. 2004).

104 Eddy covariance flux towers have been providing continuous measurements of
105 ecosystem-level exchange of carbon, water, and energy spanning diurnal, synoptic, seasonal,
106 and interannual time scales since the early 1990s (Wofsy et al. 1993; Baldocchi et al. 2001). At
107 present, over 500 eddy covariance flux towers are operating on a long-term and continuous
108 basis around the world (FLUXNET, http://daac.ornl.gov/FLUXNET). This global network
109 encompasses a large range of climate and biome types (Baldocchi et al. 2001), and provides
110 probably the best estimates of ecosystem-level carbon fluxes. The flux towers directly measure
111 NEE that can be separated into two major components: GPP and R_e (Reichstein et al. 2005;
112 Desai et al. 2008). However, these estimates only represent fluxes at the scale of the tower
113 footprint with longitudinal dimensions ranging between a hundred meters and several
114 kilometers depending on homogeneous vegetation and fetch (Schmid, 1994; Göckede et al.,

115 2008). To quantify the exchange of CO_2 between the terrestrial biosphere and the atmosphere,
116 significant efforts are needed to scale up flux tower measurements from the stand scale to
117 landscape, regional, continental, or global scales.

118 Satellite remote sensing is a potentially valuable tool for scaling-up efforts (Running et
119 al. 1999; Turner et al. 2003; Xiao et al. 2008). Several studies have integrated flux data with
120 remote sensing data to quantify GPP over large areas. Zhang et al. (2007) estimated GPP for
121 the Northern Great Plains grasslands using satellite and flux tower data. Yang et al. (2007)
122 linked satellite observations to flux tower GPP data for the estimation of GPP for two broad
123 vegetation types in the U.S. using a machine learning approach. Despite these efforts, to our
124 knowledge, no study has upscaled AmeriFlux GPP data to the continental scale to produce
125 spatially-explicit estimates of GPP across multiple biomes and to examine the patterns,
126 magnitude, and interannual variability of GPP over the conterminous U.S.

127 Here we used a regression tree approach and remotely-sensed data from the Moderate
128 Resolution Imaging Spectroradiometer (MODIS) to upscale flux tower GPP to the continental
129 scale and produced wall-to-wall GPP estimates for multiple biomes across the conterminous
130 U.S. First, we developed a predictive GPP model based on site-specific MODIS and flux tower
131 GPP data, and validated the model using eddy flux data in both temporal and spatial domains.
132 Second, we applied the model to estimate GPP for each 1 km × 1 km cell across the
133 conterminous U.S. for each 8-day interval over the period 2000-2006 using wall-to-wall
134 MODIS data. Third, we examined the patterns, magnitude, and interannual variability of GPP
135 across the conterminous U.S.

136 **2. Data and Methods**

137 *2.1. Regression tree approach*

We used a modified regression tree approach implemented in the commercial software, Cubist, to upscale flux tower GPP to the continental scale. Regression tree algorithms typically 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. Regression tree models can account for a nonlinear relationship between predictive and target variables and allow both continuous and discrete variables. Previous studies showed that regression tree methods are not only more effective than simple techniques including multivariate linear regression, but also easier to understand than neural networks (e.g., Huang and Townshend 2003).

Cubist constructs an unconventional type of regression tree, in which the terminal nodes or leaves are linear regression models instead of discrete values (Minasny and McBratney 2008). Cubist produces rule-based models containing one or more rules, each of which is a set of conditions associated with a multivariate linear submodel. Cubist is a powerful tool for generating rule-based predictive models. A Cubist model resembles a piecewise linear model, except that the rules can overlap with one another (RuleQuest 2008). Details on regression tree approaches and Cubist were described in Yang et al. (2003), Wylie et al. (2007), and Xiao et al. (2008). In our previous study, we used Cubist to develop a predictive NEE model and upscaled NEE estimates to the continental scale for the conterminous U.S. (Xiao et al. 2008). In this study, we used Cubist to construct a predictive GPP model based on MODIS and AmeriFlux GPP data. Cubist uses three statistical measures to evaluate the quality of the constructed predictive model, including mean absolute error (MAE), relative error (RE), and product-moment correlation coefficient (Yang et al. 2003; Xiao et al. 2008). MAE is calculated as:

161 $$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (1)$$

162 where N is the number of samples used to establish the predictive model, and y_i and $\hat{y}_i$ are
163 the actual and predicted values of the response variable, respectively. RE is calculated as:

164 $$RE = \frac{MAE_T}{MAE_\mu} \quad (2)$$

165 where MAE_T is the MAE of the constructed model, and MAE_μ is the MAE that would result
166 from always predicting the mean value. All three statistical measures were used to evaluate the
167 performance of the constructed model.

168 *2.2. Explanatory variables*

169 GPP is influenced by a variety of physical, physiological, atmospheric, hydrological,
170 and edaphic variables. At the leaf level, GPP is influenced by several factors, including
171 incoming solar radiation, air temperature, vapor pressure deficit, soil moisture, and nitrogen
172 availability (Ruimy et al. 1995; Clark et al. 1999, 2004). At the canopy or ecosystem level,
173 GPP is also influenced by leaf area index (LAI) (Ruimy et al. 1995) and canopy phenology
174 (Richardson et al. 2008a). At the stand or regional level, GPP is significantly affected by
175 disturbances such as fire and harvest (Law et al. 2004). Many of these factors can be
176 effectively assessed by satellite remote sensing. Surface reflectance depends on vegetation
177 type, biophysical properties (e.g., biomass, leaf area, and stand age), soil background, soil
178 moisture conditions, and sun-object-sensor geometry (Ranson et al. 1985; Penuelas et al. 1993;
179 Schmidt and Skidmore 2003). Vegetation indices including normalized difference vegetating
180 index (NDVI) and enhanced vegetation index (EVI) are closely correlated to the fraction of
181 photosynthetically active radiation (fPAR; Asrar et al. 1984), and are also related to vegetation
182 biomass and fractional vegetation cover (e.g., Tucker et al. 1985; Persson et al. 1993; Myneni

183 et al. 2001; Chen et al. 2004). Compared to NDVI, EVI is more responsive to canopy structural
184 variations, such as LAI, canopy type, plant physiognomy, and canopy architecture (Gao et al.
185 2000). The normalized difference water index (NDWI; Gao 1996) was shown to be strongly
186 correlated with leaf water content (Jackson et al. 2004) and soil moisture (Fensholt & Sandholt
187 2003) over time. LAI and fPAR characterize vegetation canopy functioning and energy
188 absorption capacity (Myneni et al. 2002) and are key parameters in most ecosystem
189 productivity and biogeochemical models (Sellers et al. 1997). We therefore selected surface
190 reflectance, EVI, LST, LAI, fPAR, and NDWI as explanatory variables. All of these variables
191 were derived from MODIS data, which also avoided the complications and difficulties to
192 merge disparate data sources (Xiao et al. 2008).

193 *2.3. AmeriFlux data*

194 We obtained the following three types of data: GPP from eddy covariance flux towers,
195 explanatory variables derived from MODIS, and a land cover map. The AmeriFlux network
196 coordinates regional analysis of observations from eddy covariance flux towers across North
197 America, Central America, and South America (Law 2006). We obtained the Level 4 data
198 product for 42 AmeriFlux sites over the period 2000-2006 from the AmeriFlux website
199 (http://public.ornl.gov/ameriflux) (Table 1). This product includes NEE data from most of the
200 active flux sites in the network. These sites are distributed across the conterminous U.S. and
201 cover a range of vegetation types: forests, shrublands, savannas, grasslands, and croplands.
202 Moreover, the distribution of these sites in the mean annual climate space indicates that the
203 sites we selected are fairly representative of typical U.S. climate types (Xiao et al. 2008). In
204 addition, some of the forested sites (e.g., Austin Cary, FL; Metolius new young pine, OR;
205 Metolius intermediate aged ponderosa pine, OR; Wisconsin intermediate hardwood, WI) are at

206 different stages since stand replacing disturbance, which are located in disturbance clusters of
207 sites. In addition, some of the sites have received treatment, including the Howland Forest
208 West Tower (ME; nitrogen fertilizer) and the Mead cropland sites (NE; irrigation versus
209 rainfed, continuous maize versus maize/soybean rotation). We therefore believe that these sites
210 are fairly representative of typical U.S. ecosystem and climate types.

211 The Level 4 product consists of two types of GPP data, including standardized
212 (GPP_st) and original (GPP_or) GPP. GPP was calculated from NEE and ecosystem
213 respiration (R_e):

214 $$GPP\_st = R_e - NEE\_st \quad (3)$$

215 and

216 $$GPP\_or = R_e - NEE\_or \quad (4)$$

217 where NEE_st and NEE_or are standardized and original NEE, respectively. NEE_st was
218 calculated using the storage obtained from the discrete approach (single point on the top of the
219 tower) with the same approach for all the sites, whereas NEE_or was calculated using the
220 storage sent by the principal investigators that can be obtained with the discrete approach or
221 using a vertical CO_2 profile system. Both NEE_st and NEE_or were gap-filled using the
222 Marginal Distribution Sampling (MDS) method (Reichstein et al. 2005) and the Artificial
223 Neural Network (ANN) method (Papale and Valentini 2003). The ANN method was generally,
224 if only slightly, superior to the MDS method (Moffat et al. 2007). A number of methods are
225 available for estimating GPP. Although Stoy et al. (2006) showed that the non-rectangular
226 hyperbolic method (Gilmanov et al. 2003) produce estimates more consistent with independent
227 data, we chose to use a method that relies on gap-filled nighttime data because it is more
228 frequently used and less computationally demanding. We used GPP calculated from NEE data

229 that was gap-filled using the ANN method. For each site, if the percentage of the remaining
230 missing values for GPP_st was lower than that for GPP_or, we selected GPP_or; otherwise, we
231 used GPP_st. GPP_st was the first choice so that the processing procedure for GPP was the
232 same for as many sites as possible. We used 8-day average GPP data (g C m^{-2} day^{-1}) to match
233 the compositing intervals of MODIS data.

234 *2.4. MODIS data*

235 We used the following four MODIS data products (Collection 4), including surface
236 reflectance (MOD09A1; Vermote and Vermeulen 1999), daytime and nighttime LST
237 (MOD11A2; Wan et al. 2002), EVI (MOD13A1; Huete et al. 2002), and LAI/fPAR
238 (MOD15A2; Myneni et al., 2002). Surface reflectance and EVI are at a spatial resolution of
239 500m, while LST, LAI, and fPAR are at spatial resolution of 1 km. Surface reflectance, LST,
240 LAI, and fPAR are at a temporal resolution of 8 days, while EVI is at a temporal resolution of
241 16 days. Sims et al. (2005) showed that the midday values of gross CO_2 exchange during
242 satellite overpasses can be used to estimate 8-day mean gross CO_2 exchange, bridging the
243 connection between continuous measurements of flux tower data and 8-day MODIS data. We
244 used the 16-day EVI product instead of EVI calculated from 8-day surface reflectance despite
245 the lower temporal resolution of the 16-day EVI product. Each 16-day EVI composite was
246 composited from 16 daily observations (Huete et al. 2002). The VI algorithm applies a filter to
247 the data based on quality, cloud, and viewing geometry, and only the higher quality, cloud-
248 free, filtered data are retained for compositing; the maximum value composite (MVC) method
249 employed selects the observation with the highest VI value to represent the composting period
250 (16 days) (Huete et al. 2002). MVC minimizes the contamination of clouds and aerosols and
251 the effects of sensor view angles on VI (Hoblen 1986). For the 8-day surface reflectance

252 product, each pixel contains the best possible daily observation during an 8-day period as
253 selected on the basis of high observation coverage, low view angle, the absence of clouds or
254 cloud show, and aerosol loading (Vermote and Kotchenova, 2008), and the EVI calculated
255 from the 8-day surface reflectance is less representative of the composting period than the
256 MODIS EVI product.

257 For each AmeriFlux site, we obtained MODIS ASCII (American Standard Code for
258 Information Interchange) subsets (Collection 4) consisting of 7 km × 7 km regions centered on
259 the flux tower, including surface reflectance, daytime and nighttime LST, EVI, LAI, and fPAR
260 over the period 2000-2006 from the Oak Ridge National Laboratory's Distributed Active
261 Archive Center (ORNL DAAC 2006). We extracted average values for the central 3 km × 3
262 km area within the 7 km × 7 km cutouts to better represent the flux tower footprint (Schmid
263 2002; Rahman et al. 2005; Xiao et al. 2008). For each variable, we determined the quality of
264 the value of each pixel within the area using the quality assurance (QA) flags included in the
265 product. At each time step, we averaged the values of each variable using the pixels with good
266 quality within the area to represent the values at the flux site. If none of the values within the 3
267 × 3 km area were of good quality, we treated the period as missing. Each 16-day EVI value
268 was used for the two 8-day intervals corresponding with the compositing interval of other
269 MODIS data products. NDWI was calculated from band 2 and band 6 of the surface
270 reflectance product.

271 To estimate GPP at the continental scale, we obtained wall-to-wall MODIS data
272 including surface reflectance, daytime and nighttime LST, LAI, and EVI over the period from
273 February 2000 to December 2006 from the Earth Observing System (EOS) Data Gateway. For
274 each variable, we determined the quality of the value of each pixel using the QA flags and

275 replaced the bad-quality value using a linear interpolation approach (Zhao et al. 2005). The
276 NDWI was calculated from band 2 (near-infrared, 841-876 nm) and band 6 (shortwave
277 infrared, 1628-1652 nm) of the surface reflectance product (MOD09A1). Similarly, each 16-
278 day EVI composite was used for two 8-day intervals corresponding to the compositing interval
279 of other MODIS products. NDWI was calculated from band 2 and band 6 of the surface
280 reflectance product for each 8-day interval.

281 We also used the MODIS 8-day GPP product (MOD17A2; Running et al. 2004) to
282 evaluate the performance of our model at both the site level. The MODIS GPP product is at a
283 spatial resolution of 1 km, and a temporal resolution of 8 days. We obtained MODIS ASCII
284 subsets (Collection 4) for the MODIS 8-day GPP product over the period 2005-2006 from the
285 Oak Ridge National Laboratory's Distributed Active Archive Center (ORNL DAAC 2006).
286 We also obtained the MODIS annual GPP product (MOD17A3; Running et al. 2004) for 2005
287 from the Numerical Terradynamic Simulation Group, University of Montana
288 (http://www.ntsg.umt.edu) to evaluate the performance of our model at the continental scale.

289 *2.5. Land cover*

290 To construct a predictive GPP model, we obtained the land cover type for each
291 AmeriFlux site based on the site descriptions (Table 1) and categorized each site into a class of
292 the UMD (University of Maryland) land-cover classification system. Although the 42
293 AmeriFlux sites used in this study cover a variety of vegetation classes of this classification
294 system, some classes (e.g., deciduous needleleaf forests, open shrublands) were not covered by
295 any site. We therefore reclassified all vegetation classes of the UMD classification system to
296 seven broader classes, following Xiao et al. (2008). Specifically, evergreen needleleaf forests
297 and evergreen broadleaf forests were merged to evergreen forests, deciduous needleleaf forests

298 and deciduous broadleaf forests to deciduous forests, closed shrublands and open shrublands to
299 shrublands, and woody savannas and savannas to savannas.

300 To estimate GPP for each 1 km × 1 km pixel across the conterminous U.S., we
301 obtained the land-cover type for each pixel from the MODIS land-over map with the UMD
302 classification system (Friedl et al. 2002). Similarly, we reclassified the vegetation classes of the
303 MODIS land-cover map to the seven broader classes. We then used the reclassified land cover
304 map to specify the land cover of each 1 km × 1 km cell across the conterminous U.S.

305 *2.6. Model development*

306 We developed a predictive GPP model using Cubist based on the site-specific MODIS
307 and AmeriFlux GPP data. Our explanatory variables included land cover, surface reflectance
308 (bands 1-7), daytime and nighttime LST, EVI, NDWI, fPAR, and LAI, and our response
309 variable was GPP (g C m^{-2} day^{-1}). Land cover was included as a categorical variable in the
310 model. We split the site-level data set of AmeriFlux and MODIS data into a training set (2000-
311 2004) and a test set (2005-2006). If a site only had GPP observations for the period 2000-2004,
312 the site was only included in the training set; if a site only had GPP observations for the period
313 2005-2006, the site was only included in the test set; otherwise, the site was included in both
314 training and test sets. The training and test sets included 40 and 34 AmeriFlux sites,
315 respectively. We had a total of 4529 and 2240 data samples for the training and test sets,
316 respectively. In addition to the full model that includes all of the 14 independent variables, we
317 also developed a series of models by dropping one or more variables at a time using Cubist. To
318 select the best model, we evaluated the performance of each model based on MAE, RE, and
319 correlation coefficient. We chose the model with the minimal MAE and RE and maximum
320 correlation coefficient as the best model. We also evaluated the model performance using the

321 Root Mean Squared Error (RMSE), scatterplots of predicted GPP versus observed GPP, and
322 seasonal variations between the predicted and observed GPP.

323 We also evaluated the performance of our model in the spatial domain using leave-one-
324 out cross-validation. In this approach, the data from a single site was used for validation, and
325 the data from the remaining sites were used for training. The training and validation data were
326 from different sites and were therefore independent from each other as these sites are generally
327 hundreds of kilometers away from each other and the spatial autocorrelation between these
328 sites was negligible. The leave-one-out cross-validation was conducted for each site,
329 separately.

330 *2.7. Continental-scale estimation of GPP*

331 As mentioned earlier, the AmeriFlux sites used in this study are fairly representative of
332 typical U.S. ecosystem and climate types. We believe that the predictive GPP model
333 constructed from the 42 sites can be extrapolated to the conterminous U.S. We used the model
334 to estimate GPP for each 1 km × 1 km cell across the conterminous U.S. for each 8-day
335 interval over the period 2000-2006 using wall-to-wall MODIS data. GPP was not estimated for
336 non-vegetated cells (e.g., urban, barren), and water bodies. We compared our estimate with the
337 MODIS GPP product (MOD17A3; Running et al. 2004). With our 8-day GPP estimates, we
338 examined the patterns, magnitude, and interannual variability of GPP.

339 **3. Results and Discussion**

340 *3.1. Model development*

341 *3.1.1 Predictive GPP model*

342 We chose the model containing five explanatory variables – land cover, EVI, daytime
343 LST, LAI, and NDWI as the best model to predict GPP at the continental scale (RE = 0.38,

MAE = 1.22 g C m^{-2} day^{-1}, R^2 = 0.74). The performance of the model was comparable to that of the full model (RE = 0.37, MAE = 1.19 g C m^{-2} day^{-1}, R^2 = 0.74). Having only five explanatory variables could substantially reduce the computational complexity for continental-scale predictions compared to the full model. The model consisted of five committee models, each of which was made of a number of rule-based submodels. For instance, the first committee model was made of the following 23 rule-based submodels:

Rule 1: if land cover in {Deciduous forests, Mixed forests}, EVI <= 0.37, and LAI <= 2.64, then

$$GPP = 0.28 + 0.012\ LST_{day}$$

Rule 2: if LST_{day} <= 3.27, EVI > 0.22, then

$$GPP = -0.64 + 4.8EVI + 0.6NDWI + 0.06LAI$$

Rule 3: if land cover in {Evergreen forests, Shrublands, Savannas, Grasslands, Croplands}, EVI <= 0.22, LAI <= 2.64, then

$$GPP = 0.15 + 1.4EVI + 0.01\ LST_{day}$$

Rule 4: if land cover in {Deciduous forests, Mixed forests, Savannas, Croplands}, LST_{day} > 3.26, EVI <= 0.37, LAI <= 1.78, then

$$GPP = -1.10 + 0.75LAI + 4.6EVI$$

...

Rule 22: if NDWI <= -0.28, LST_{day} > 17.65, EVI > 0.58, then

$$GPP = 10.49 + 1.9EVI + 0.02LST_{day} - 0.7NDWI$$

Rule 23: if land cover = Croplands, LST_{day} > 29.92, EVI > 0.37, then

$$GPP = -24.85 + 5.47LAI + 58.9EVI + 0.124LST_{day} - 0.5NDWI$$

366 where LST_{day} is the daytime LST. As mentioned earlier, the rules of the model could overlap
367 with one another. For instance, rule 1 overlapped with rule 4 as land cover could be deciduous
368 forests in both cases; rules 22 and 23 overlapped with each other as EVI could be greater than
369 0.58 in both cases.

370 *3.1.2. Model evaluation*

371 The analysis of model residuals indicated that the residuals were not randomly
372 distributed. Low GPP values were generally associated with low prediction errors, whereas
373 high GPP values were associated with high prediction errors. The uncertainties of carbon flux
374 measurements are directly proportional to the magnitudes of the fluxes (Richardson et al.
375 2008b). The residuals also exhibited a systematic component. For example, large GPP tended
376 to have consistently negative residuals. The residuals also had a random component that arose
377 partially from errors/uncertainties in the measured fluxes as well as MODIS data. Random
378 errors in AmeriFlux GPP data are significant (Hagen et al. 2006) and these errors may
379 ultimately limit the agreement between observed and predicted GPP values. In addition, the
380 explanatory variables included in the model could not completely explain the variance of GPP.
381 For example, the independent variables used in the model could not account for nitrogen
382 availability, and may affect the accuracy of the model.

383 We compared our GPP estimates with observed GPP for each AmeriFlux site over the
384 period 2005-2006 (Fig. 2). Our estimates captured most features of observed GPP including
385 seasonality and year-to-year variations over the period 2005-2006. GPP was under- and over-
386 predicted for some sites. The model could not capture exceptionally high GPP values for some
387 sites, such as Audubon Research Ranch (AZ), Santa Rita Mesquite (AZ), and Fort Peck (MT).
388 We averaged observed and predicted 8-day GPP for each site, and plotted mean predicted GPP

389 against observed GPP (Fig. 3). The model estimated GPP fairly well at the site level ($y = 0.95x$
390 $+ 0.21$, $R^2 = 0.84$, $p < 0.0001$; RMSE = 0.77 g C m^{-2} day^{-1}). Overall, the model slightly under-
391 and overestimated GPP for values greater or lower than ~4 g C m^{-2} day^{-1}, respectively. The
392 model performance also varied with site. Large underestimation occurred at some sites such as
393 the North Carolina Loblolly Pine (NCP, NC), Freeman Ranch Mesquite Juniper (FRM, TX),
394 and Walnut Gulch Kendall Grasslands (WGK, AZ), whereas large overestimation occurred at
395 some sites such as Lost Creek (LC, WI) and Mead Rainfed (MR, NE). The model predicted
396 GPP remarkably well at the biome level ($y = 0.99x - 0.13$, $R^2 = 0.91$, $p < 0.00001$; RMSE =
397 0.42 g C m^{-2} day^{-1}). The model slightly overestimated GPP for deciduous forests and croplands
398 and slightly underestimated GPP for all other biomes.

399 The disagreement between predicted and observed GPP values is likely due to the
400 following reasons. First, the MODIS and tower footprints do not match with each other and the
401 vegetation structure at the flux tower could be significantly different from that within the
402 MODIS footprint (Xiao et al. 2008). For example, the Tonzi Ranch site (CA) is dominated by
403 deciduous blue oaks (*Quercus douglasii*) and the understory and open grassland are mainly
404 cool-season C_3 annual species (Ma et al. 2007). The MODIS footprint consists of a larger
405 fraction of grassland than the tower footprint. Blue oaks and grasses have distinct phenologies
406 (Ma et al. 2007) and therefore had differential contributions to the carbon fluxes integrated
407 over the MODIS footprint over time (Xiao et al. 2008). Second, MODIS data is less sensitive
408 to changes in understory vegetation and damage to canopies that do not increase canopy gaps,
409 leading to overestimation of carbon assimilation rates. Third, the independent variables
410 included in the model could not account for other factors such as nitrogen availability (Clark et
411 al. 1999, 2004) and stand age (Ryan et al. 2004), all of which may influence GPP. Finally, we

412 estimated GPP for each 8-day interval, and therefore our estimates may not capture the
413 variability of GPP within that period. The 8- or 16-day LST and EVI values do not always
414 represent average environmental conditions and average fluxes over the 8- or 16-day period
415 (Xiao et al. 2008), and the exclusion of days with high and low values could lead to
416 underestimation and overestimation of GPP values, respectively. In addition, during drought or
417 days with high vapor pressure deficits, the midday GPP may not be representative because of
418 the skewed diurnal variation in GPP (Anthoni et al. 1999).

419 *3.1.3. Model validation*

420 We validated our model in both temporal and spatial domains. We first validated the
421 model in the temporal domain using the test set over the period 2005-2006 (Fig. 1a). Our
422 model estimated GPP fairly well ($R^2 = 0.74$, $p < 0.0001$; RMSE = 1.99 g C m^{-2} day^{-1}), although
423 it slightly under- and overestimated GPP values greater and less than 3 g C m^{-2} day^{-1},
424 respectively. By contrast, the MODIS GPP product estimated GPP for the AmeriFlux sites
425 with a RMSE of 2.43 g C m^{-2} day^{-1} ($y = 0.50x + 1.01$, $R^2 = 0.58$, $p < 0.0001$; Fig. 1b). The R^2
426 and RMSE of our model were 28% higher and 18% lower than those of the MODIS GPP
427 product, respectively.

428 We then validated the model in the spatial domain using leave-one-out cross validation.
429 The cross-validation also showed that our model estimated GPP fairly well (Fig. 4). The model
430 performance varied with site and biome type. Our model had a higher performance for forest
431 ecosystems and croplands than shrublands, savannas, and grasslands. We averaged observed
432 and predicted 8-day GPP for each site, and then plotted mean predicted GPP against observed
433 GPP (Fig. 5). The cross-validation showed that our model estimated GPP fairly well at the site
434 level ($y = 0.73x + 1.07$, $R^2 = 0.69$, $p < 0.0001$; RMSE = 1.19 g C m^{-2} day^{-1}) although the slope

435 and R^2 were lower and the RMSE was higher than those of the validation in the temporal
436 domain, respectively. The cross-validation also showed that our model estimated GPP
437 remarkably well at the biome level ($y = 0.86x + 0.48$, $R^2 = 0.91$, $p < 0.0001$; RMSE = 0.33 g C
438 m^{-2} day^{-1}).

439 The validation of the model in both temporal and spatial domains showed that the
440 performance of our model is encouraging, given the diversity in ecosystem types, age
441 structures, fire and insect disturbances, and management practices. Our approach extensively
442 used eddy covariance flux tower data involving typical U.S. ecosystem and climate types. Our
443 study demonstrated that our empirical approach has great potential for upscaling flux tower
444 GPP data to continental scales across a variety of biomes.

445 *3.2. Gross Primary Production*

446 *3.2.1. Seasonal patterns*

447 Our 8-day GPP estimates were highly constrained by eddy flux data, and provided a
448 spatially and temporally continuous measure of GPP with high spatial and temporal resolution
449 for the conterminous U.S., which made it possible to examine the patterns, magnitude, and
450 interannual variability of GPP across the U.S. Our estimates showed that GPP exhibited large
451 spatial variability and strong seasonal fluctuations (Fig. 6). The seasonal patterns of GPP and
452 its spatial variability reflected the controlling effects of climate conditions. In the spring
453 months, the Southeast and the Gulf Coast significantly assimilated carbon with GPP values
454 reaching ~100-250 g C m^{-2} mo^{-1} as the growing season started in early to mid-spring in these
455 regions. The Pacific Coast is dominated by evergreen forests, and these ecosystems also
456 assimilated carbon due to mild temperatures and moist conditions during the spring (Anthoni et
457 al. 2002). The Mediterranean regions in California also assimilated carbon in the spring

458 because of a surplus of precipitation and relatively warm temperatures (Xu & Baldocchi 2004;
459 Ma et al. 2007). By contrast, the Upper Great Lakes region and the northern Great Plains are
460 dominated by croplands with most crops planted between April and June (Shroyer et al. 1996),
461 whereas the New England region and the northern portion of the Upper Great Lakes region are
462 dominated by temperate-boreal transitional forests. The relatively late greenup in these regions
463 led to low GPP.

464 In the summer months, the eastern U.S., the Coastal Pacific Northwest, and some
465 regions in California exhibited high GPP values (~250–450 g C m^{-2} mo^{-1}) owing to favorable
466 temperature and soil moisture conditions, while the vast majority of western landscapes,
467 including the Great Basin, the Colorado Plateau, and the western Great Plains exhibited much
468 lower GPP values due to spare vegetation and precipitation deficits.

469 In the fall months (September-November), the GPP values of the Southeast and the
470 Gulf Coast substantially decreased relative to those in the summer because vegetation began to
471 senesce and days became shorter in these regions. The spatial patterns and magnitude of GPP
472 were similar to those of the spring. The Upper Great Lakes region and the Great Plains had
473 very low GPP values due to the harvesting of crops.

474 In the winter months (December-February), the majority of the U.S. had little or no
475 photosynthesis as the canopies of most ecosystems were dormant. Some regions in the Pacific
476 Coast, California, the Gulf Coast, and the Southeast slightly assimilated carbon. because of the
477 dominance of evergreen forests and mild temperatures (Waring & Franklin 1979; Clark et al.
478 1999; Anthoni et al. 2002; McGarvey et al. 2004). For example, Douglas-fir, a major species in
479 the Pacific Northwest and California, is known to be highly plastic and able to photosynthesize
480 in winter when temperatures are above the freezing point (Xiao et al. 2008).

481 Fig. 7 showed the trajectories of the spatially averaged and integrated 8-day GPP for
482 each biome from February 2000 to December 2006. Deciduous forests and croplands had the
483 largest intra-annual variability in spatial averaged GPP, followed by mixed forests; evergreen
484 forests and savannas had intermediate intra-annual variability; grasslands and shrublands had
485 the least variability (Fig. 7a). The temporal variability of spatially integrated GPP (or spatial
486 total) also showed clear dependence on biome (Fig. 7). Collectively, the terrestrial ecosystems
487 substantially assimilated carbon and had a peak spatial total of 50-55 Tg C day^{-1}. Taken
488 separately, croplands had the highest intra-annual variability in spatially integrated GPP, with a
489 peak spatial total of 25-30 Tg C day^{-1}, followed by deciduous forests; evergreen forests, mixed
490 forests, savannas, and grasslands had intermediate intra-annual variability in spatially
491 integrated GPP; shrublands had the least variability. Both spatially averaged and integrated
492 GPP showed interannual variability for each biome.

493 *3.2.2. Annual fluxes*

494 We calculated annual GPP for each year over the period 2001-2006 from our 8-day
495 GPP estimates, and then calculated the average annual GPP over the 6-year period (Fig. 8).
496 Annual GPP varied considerably over space, and exhibited a large spatial gradient from the
497 east to the west. The Gulf Coast, the Southeast, the coastal Pacific Northwest, and a part of the
498 Pacific Southwest had high annual GPP (~1500-2000 g C m^{-2} yr^{-1}); the Midwest and the
499 Northeast had intermediate values (~1200-1500 g C m^{-2} yr^{-1}), and the majority of the western
500 half of the country including the Southwest, the western Great Plains, and the Rocky Mountain
501 region had GPP values generally lower than 500 g C m^{-2} yr^{-1}.

502 We estimated a total gross carbon uptake of 7.06 Pg C yr^{-1} for the conterminous U.S.
503 over the period 2001-2006. A quantitative breakdown of the 6-yr average annual GPP map by

504 regions (Table 2) showed that the North Central and South Central regions had the highest
505 GPP, followed by the Southeast, the Rocky Mountain region, and the Northeast; the Pacific
506 Northwest and the Pacific Southwest had the lowest GPP. The spatially averaged annual GPP
507 of the U.S. was ~1100 g C m^{-2} yr^{-1} (Table 2). Regionally, the Southeast had the highest
508 spatially averaged annual GPP, followed by the Northeast and the South Central regions; the
509 North Central region, the Pacific Northwest, and the Pacific Southwest had intermediate
510 values; the Rocky Mountain region had the lowest spatially averaged annual GPP.

511 A quantitative breakdown of the 6-yr average annual GPP by biomes (Table 3) showed
512 spatially integrated annual GPP varied with biome. Croplands had the highest spatially
513 integrated annual GPP; shrublands had lowest annual GPP; other biomes including evergreen
514 forests, deciduous forests, mixed forests, savannas, and grasslands had intermediate annual
515 GPP. Spatially averaged annual GPP also varied with biome (Table 3). Deciduous forests had
516 the highest spatially averaged annual GPP, followed by evergreen forests, mixed forests,
517 croplands, and savannas; grasslands had intermediate values; shrublands had the lowest values.

518 We compared our annual GPP estimate with annual GPP estimate from the MODIS
519 GPP product (MOD17A3; Running et al. 2004) for 2005 (Fig. 9). Both estimates showed a
520 large spatial gradient from the east to the west: the Southeast had the highest annual GPP; the
521 Midwest had intermediate annual GPP, while the Rocky Mountain region had the lowest GPP.
522 Both estimates showed that annual GPP is ~300 g C m^{-2} yr^{-1} in the Rocky Mountain region.
523 Annual GPP is also very similar in the New England region and the Upper Peninsula of
524 Michigan. However, large discrepancies were observed between our estimate and the MODIS
525 GPP product. Our estimate exhibited larger spatial variability than the estimate from the
526 MODIS GPP product. Moreover, compared to our estimate, the MODIS GPP product

527 substantially underestimated GPP in some regions, particularly the Midwest and the coastal
528 Pacific Northwest where ecosystems are highly productive. For example, our annual GPP for
529 the croplands in the Midwest is ~1200-1500 g C m^{-2} yr^{-1}, while the MODIS annual GPP is only
530 ~700 g C m^{-2} yr^{-1}. In many areas in the Southeast, our annual GPP estimate is ~1500-2000 g C
531 m^{-2} yr^{-1}, while the MODIS annual GPP is only ~1000-1500 g C m^{-2} yr^{-1}.

532 The large discrepancies in annual GPP between our estimate and the MODIS GPP
533 product can be attributed to the following reasons. First, the MODIS GPP product was
534 developed using an algorithm optimized for global applications and meteorological fields with
535 coarse resolution (1 by 1.25 degree) and large uncertainties (Zhao et al. 2006), and thereby
536 likely contributed to the smaller spatial variability of MODIS GPP and affected the accuracy of
537 the MODIS GPP estimates. Second, the maximum light use efficiency (ε_{max}) is an essential
538 parameter of the LUE model used to develop the MODIS GPP product. The ε_{max} is only 0.68 g
539 C MJ^{-1} for croplands in the MOD17 algorithm (Heinsch et al. 2003), which is likely too low
540 for croplands and results in substantial underestimation of cropland GPP. Zhang et al. (2008)
541 showed that MODIS annual GPP for an irrigated cropland in China was only about 20-30% of
542 annual GPP derived from eddy covariance flux measurements, and attributed the substantial
543 underestimation of cropland GPP to the underestimation of ε_{max} in the MOD17 algorithm.
544 Third, our estimate was highly constrained eddy flux data, while some geographical regions
545 and biomes are underrepresented by the AmeriFlux network, which could affect the accuracy
546 of our estimates. For example, we merged savannas (tree cover 10-30%) and woody savannas
547 (tree cover 30-60%) together in the development of the model because no sites representing
548 typical savannas with tree cover below 30% were available in the model, which could lead to
549 overestimation of GPP for areas that were classified as savannas (tree cover 10-30%).

550 Global annual GPP has been estimated to be 120 Pg C yr^{-1} using ^{18}O measurements of
551 atmospheric CO_2 (Ciais et al. 1997) and 110 Pg C yr^{-1} from 2001 to 2003 using the MODIS
552 GPP product (Zhao et al. 2005). Our estimates suggested that the terrestrial ecosystems in the
553 conterminous U.S. accounted for 5.9-6.5% of the global annual GPP, while its land area
554 accounts for ~5.4% of the global land area. Our estimate of U.S. annual GPP was higher than
555 other estimates. For example, the average annual GPP over the period 2001-2006 derived from
556 MODIS GPP product (Running et al. 2004) was ~6.2 Pg C yr^{-1}. Our estimate was about 14%
557 higher than the MODIS GPP estimate. Potter et al. (2007) estimated annual NPP between 2.67
558 and 2.79 Pg C yr^{-1} over the period 2001-2004 using MODIS data and the NASA-CASA model.
559 Our estimate was also higher than that estimated by Potter et al. (2007) assuming that NPP is
560 about half of GPP (Lloyd & Farquhar 1996; Waring et al. 1998).

561 Our predictive model has advantages over empirical or process-based ecosystem
562 models. Most ecosystem models are dependent on site-level parameterizations that are used as
563 default parameters for a much broader spectrum of vegetation types and conditions, which may
564 limit the accuracy of model simulations over large areas (e.g., Prince and Goward 1995;
565 Running et al. 2004; Xiao et al. 2009). By contrast, our model was highly constrained by eddy
566 flux data from a number of towers encompassing a range of ecosystem and climate types, and
567 may lead to model parameters that are more representative of the full spectrum of vegetation
568 and climate types and thereby more accurate estimates of carbon fluxes at regional scales.
569 Moreover, our model consisted of rule-based, multivariate linear regression models, and is
570 easier to understand and implement. Our model could substantially reduce computational
571 complexity compared to many ecosystem models. On the other hand, however, our model also
572 has disadvantages over empirical or process-based ecosystem models. Our model is an

573 empirical approach, and does not involve ecosystem processes such as photosynthesis and
574 nitrogen cycling. Moreover, our model did not explicitly consider some factors influencing
575 GPP such as nitrogen availability, stand age, and disturbance history that may be explicitly
576 simulated in process-based ecosystem models (e.g., Aber et al. 1997). All these model
577 differences contributed to the discrepancies in annual GPP estimates between our empirical
578 approach and ecosystem models.

579 *3.2.3. Interannual variability*

580 The U.S. annual GPP varied between 6.91 and 7.33 Pg C yr^{-1} over the period 2001-
581 2006. The years 2002, 2004, and 2006 had lower GPP values relative to 2001, 2003, and 2005.
582 Annual GPP exhibited positive and negative anomalies for each year, and the spatial patterns
583 of these anomalies also varied from year to year (Fig. 10). The annual GPP anomalies were
584 likely due to climate variability, disturbances, and management practices. Most notably, large
585 negative GPP anomalies occurred in the Great Plains in 2002 and 2006. Moderate to severe
586 drought affected over 50% of the country in both years, including the Southwest, the Great
587 Plains, the Gulf Coast, and the coastal Southeast, particularly Texas and Oklahoma (U.S.
588 Drought Monitor, http://www.drought.unl.edu). The annual precipitation of these two years
589 was 467 and 458 mm for the U.S., respectively - lower than the 30-year mean annual
590 precipitation (480 mm) taken over the PRISM climate database. Our GPP anomaly maps
591 showed large negative GPP anomalies in many of the drought-affected regions. Notably, large
592 negative GPP anomalies occurred in the Great Plains in 2002 and 2006. Our results further
593 demonstrated severe drought could substantially affect ecosystem carbon fluxes (Xiao et al.
594 2009).

595 At landscape to regional scales, annual GPP also exhibited large anomalies. For
596 example, our results showed large negative GPP anomalies in the region experienced the
597 Biscuit fire in Oregon (Fig. 11). The 2002 Biscuit Fire was among the largest forest fires in
598 modern U.S. history, encompassing > 2000 km^2 primarily within the Rogue-Siskiyou National
599 Forest (RSNF) in southwest Oregon (Campbell et al., 2007; Thompson et al. 2007). Large fires
600 such as the Biscuit Fire damaged both overstory and/or understory vegetation, leading to a
601 reduction in GPP and large negative GPP anomalies in the region in 2003, which may result in
602 net carbon release into the atmosphere. Numerous wildfires occurred over the western half of
603 the country due to dry weather and high winds, burning 1.5-4.0 × 10^4 km^2 of forests from 2000
604 to 2006 (U.S. Fire Administration, http://www.usfa.dhs.gov). The drought along with wildfires
605 likely led to the negative GPP anomalies in the western half of the U.S.
606 Our results also showed hurricanes could reduce GPP and lead to large negative GPP
607 anomalies (Fig. 12). For example, Hurricane Katrina occurred in late August 2005 affected
608 over 2 × 10^4 km^2 of forest across Mississippi, Louisiana and Alabama, with damage ranging
609 from downed trees, snapped trunks and broken limbs to stripped leaves (USDA Forest Service,
610 http://www.srs.fs.usda.gov). Forest inventories indicated that the potential timber losses from
611 Hurricane Katrina amounted to roughly 1.2 × 10^8 m^3 (USDA Forest Service,
612 http://www.srs.fs.usda.gov). Our results showed large negative GPP anomalies in 2006 in the
613 Gulf Coast region severely affected by Hurricane Katrina (Fig. 12). The reduction in GPP and
614 increased R_e resulting from increased litter could lead to substantial carbon release into the
615 atmosphere (Chambers et al. 2007).

616 **4. Conclusions**

617 We used a regression tree approach and remotely-sensed data from MODIS to upscale
618 AmeriFlux GPP data to the continental scale and to produce a GPP dataset with 8-day temporal
619 resolution and 1 km spatial resolution for the conterminous U.S. over the period 2000-2006.
620 Our results demonstrated that our empirical approach has great potential for upscale eddy flux
621 GPP data to large areas across multiple biomes. Our GPP estimates provided a spatially and
622 temporally continuous measure of gross primary production for the conterminous U.S. Our
623 estimates also provided an alternative, independent dataset from the MODIS GPP product and
624 simulations with biogeochemical models. Our GPP estimates were highly constrained by flux
625 tower data from towers encompassing a large range of ecosystem and climate types as well as
626 disturbance history. Our approach can be applied to the entire North America, other geographic
627 regions including Europe, Southeast Asia, and South America, or to the global scale, and to
628 produce continuous GPP estimates over continents or the globe. This approach can also be
629 used to upscale other fluxes including evapotranspiration to large areas.

630 Our GPP estimates exhibited large spatial variability and strong seasonal variations,
631 which reflected the controlling effects of climate conditions and vegetation distributions. We
632 estimated a total gross carbon uptake of 7.06 Pg C yr^{-1} for the conterminous U.S. over the
633 period 2001-2006. Annual GPP varied substantially with geographical region and biome type.
634 Our annual GPP estimate exhibited large spatial variability than the MODIS GPP product
635 (MOD17A3; Running et al. 2004). Our results also showed that the U.S. annual GPP varied
636 between 6.91 and 7.33 Pg C yr^{-1} over the 6-year period. Extreme climate events (e.g., drought)
637 and disturbances (e.g., fires and hurricanes) reduced annual GPP at regional scales and could
638 have a significant impact on the U.S. net ecosystem carbon exchange. The interannual

variability of GPP was mainly caused by these extreme climate events (e.g., drought) and disturbances (e.g., fire, hurricane).

Acknowledgements:

This research was partly supported by the National Science Foundation (NSF) Carbon and Water Program (*EAR*-0630319). We thank the principal investigators of the MODIS data products including Drs. A.R. Huete, R.B. Myneni, S.W. Running, E.F. Vermote, and Z. Wan. We also thank D.A. Sims, T.A. Boden, S.K.S. Vannan, M. Zhao, Z. Wan, and L. Delp for helpful discussion. The MODIS data products were obtained from the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center (DAAC) and the Earth Observing System (EOS) Data Gateway for making these products available. The PRISM climate database was provided by the PRISM Group, Oregon State University (http://www.prismclimate.org).

References:

Aber, J.D., Ollinger, S.V., Driscoll, C.T. (1997). Modeling nitrogen saturation in forest ecosystems in response to land use and atmospheric deposition. *Ecological Modelling, 101*, 61-78.

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

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

662 Baldocchi, D., Falge, E., Gu, L. et al. (2001). FLUXNET: A new tool to study the temporal
663 and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux
664 densities. *Bulletin of the American Meteorological Society, 82,* 2415-2434.
665 Campbell, J., Donato, D., Azuma, D., & Law, B. (2007). Pyrogenic carbon emission from a
666 large wildfire in Oregon, United States. *J. Geophys. Res., 112,* G04014,
667 doi:10.1029/2007JG000451.
668 Chambers, J.Q., Fisher, J.I., Zeng, H., Chapman, E.L., Baker, D.B., Hurtt, G.C. (2007).
669 Hurricane Katrina's carbon footprint on U.S. Gulf Coast forests. *Science, 318,* 1107.
670 Chen, X., Vierling, L., Powell, E., DeFelice T. (2004). Using lidar and effective LAI data to
671 evaluate IKONOS and Landsat 7 ETM+ vegetation cover estimates in ponderosa pine
672 forest. Remote Sensing of Environment, 91, 14-26.
673 Ciais, P. et al. (1997). A thee-dimensional synthesis study of y18O in atmospheric CO2. 1.
674 surface Fluxes. *Journal of Geophysical Research, D102,* 5857–5872.
675 Clark, K.L., Gholz, H.L., Moncrieff, J.B., Cropley, F., Loescher, H.W. (1999). Environmental
676 controls over net carbon dioxide from contrasting Florida ecosystems. *Ecological*
677 *Applications, 9,* 936-948.
678 Clark, K.L., Gholz, H.L., Castro, M.S. (2004). Carbon dynamics along a chronosequence of
679 slash pine plantations in N. Florida. *Ecological Applications, 4,* 1154-1171.
680 Cook, B.D., Davis, K.J., Wang, W. et al. (2004). Carbon exchange and venting anomalies in an
681 upland deciduous forest in northern Wisconsin, USA. *Agricultural Forest Meteorology,*
682 *126,* 271-295.
683 DeFries, R.S., Houghton, R.A., Hansen, M.C., Field, C.B., Skole, D., & Townshend, J. (2002).
684 Carbon emissions from tropical deforestation and regrowth based on satellite observations

685 for the 1980s and 1990s. *Proceedings of the National Academy of Sciences of the United*
686 *States of America*, *99*, 14256-14261.
687 Desai, A.R., Bolstad, P.V., Cook, B.D., Davis, J., & Carey, E.V. (2005). Comparing net
688 ecosystem exchange of carbon dioxide between an old-growth and mature forest in the
689 upper Midwest, USA. *Agricultural Forest Meteorology*, *128*, 33-55.
690 Desai, A.R., Richardson, A.D., Moffat, A.M. et al. (2008). Cross site evaluation of eddy
691 covariance GPP and RE decomposition techniques. Agricultural and Forest Meteorology,
692 in press.
693 Dore, S., Hymus, G.J., Johnson, D.P., Hinkle, C.R., Valentini, R., & Drake, B.G. (2003). Cross
694 validation of open-top chamber and eddy covariance measurements of ecosystem CO2
695 exchange in a Florida scrub-oak ecosystem. *Global Change Biology*, *9*, 84-95.
696 Falk, M., Wharton, S., Schroeder, M., Ustin, S., Paw U, K.T. (2008). Flux partitioning in an
697 old-growth forest: seasonal and interannual dynamics. *Tree Physiology*, *28*, 509-520.
698 Fensholt, R., & Sandholt, I. (2003). Derivation of a shortwave infrared water stress index from
699 MODIS near- and shortwave infrared data in a semiarid environment. *Remote Sensing*
700 *Environment*, *87*, 111-121.
701 Friedl, M.A., McIver, D.K., Hodges, J.C.F. et al. (2002). Global land cover mapping from
702 MODIS: algorithms and early results. *Remote Sensing Environment*, *83*, 287-302.
703 Gao, B.C. (1996). NDWI - A normalized difference water index for remote sensing of
704 vegetation liquid water from space. *Remote Sensing Environment*, *58*, 257-266.
705 Gao, X., Huete, A. R., Ni, W., Miura, T. (2000). Optical–biophysical relationships of
706 vegetation spectra without background contamination. *Remote Sensing of Environment, 74*,
707 609– 620.

708 Gilmanov, T.G., Verma, S.B., Sims, P.L., Meyers, T.P., Bradford, J.A., Burba, G.G., Suyker,
709 A.E. (2003). Gross primary production and light response parameters of four Southern
710 Plains ecosystems estimated using long-term CO_2-flux tower measurements. *Global*
711 *Biogeochem. Cycles, 17*, 1071, doi:10.1029/2002GB002023.
712 Göckede, M., Foken, T., Aubinet, M. et al. (2008). Quality control of CarboEurope flux data -
713 Part 1: Coupling footprint analyses with flux data quality assessment to evaluate sites in
714 forest ecosystems. *Biogeosciences, 5*, 433-450.
715 Gough, C.M., Vogel, C.S., Schmid, H.P., Su, H.-B., Curtis, P.S. (2008). Multi-year
716 convergence of biometric and meteorological estimates of forest carbon
717 storage. *Agricultural and Forest Meteorology, 148*, 158-170.
718 Gu, L., Meyers, T., Pallardy, S.G. et al. (2007). Influences of biomass heat and biochemical
719 energy storages on the land surface fluxes and radiative temperature. *Journal of*
720 *Geophysical Research, D02107*, doi:10.1029/2006JD007425.
721 Gu, L., Meyers, T., Pallardy, S.G. et al. (2006). Direct and indirect effects of atmospheric
722 conditions and soil moisture on surface energy partitioning revealed by a prolonged
723 drought at a temperate forest site. *Journal of Geophysical Research, D16102*,
724 doi:10.1029/2006JD007161.
725 Hagen, S.C., Braswell, B.H., Linder, E., Frolking, S., Richardson, A.D., Hollinger, D.Y.
726 (2006). Statistical uncertainty of eddy-flux based estimates of gross ecosystem carbon
727 exchange at Howland Forest, Maine. *Journal of Geophysical Research, 111*, D08S03,
728 doi:10.1029/2005JD006154.

729 Heinsch, F.A., Zhao, M., Running, S.W. (2006). Evaluation of remote sensing based terrestrial
730 productivity from MODIS using regional tower eddy flux network observations. *IEEE*
731 *Transactions on Geosciences and. Remote Sensing, 44*, 1908-1925.
732 Hoblen, B.N. (1986). Characteristics of maximum value composite images from temporal
733 VHRR data. *International Journal of Remote Sensing, 7*, 1417–1434.
734 Hollinger, D.Y., Goltz, S.M., Davidson, E.A., Lee, J.T., Tu, K.,
735 Valentine, H.T. (1999). Seasonal patterns and environmental control of carbon
736 dioxide and water vapor exchange in an ecotonal boreal forest. *Global*
737 *Change Biology, 5*, 891-902
738 Hollinger, D.Y., Aber, J., Dail, B. et al. (2004). Spatial and temporal variability in
739 forest-atmosphere CO_2 exchange. *Global Change Biology, 10*, 1689-1706.
740 Hollinger, S.E., Bernacchi, C.J., Meyers, T.P. (2005). Carbon budget of mature no-till
741 ecosystem in North Central Region of the United States. *Agricultural and Forest*
742 *Meteorology, 130*, 59-69, 2005.
743 Huang, C., Townshend, J.R.G. (2003). A stepwise regression tree for nonlinear approximation:
744 applications to estimating subpixel land cover. *International Journal of Remote Sensing,*
745 *24*, 75-90.
746 Huete, A., Didan, K., Miura, T., Rodríguez, E.P., Gao, X., & Ferreira, L.G. (2002). Overview
747 of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote*
748 *Sensing of Environment, 83*, 195-213.
749 Irvine, J., Law, B.E., Hibbard, K.A. (2007). Postfire carbon pools and fluxes in semiarid
750 ponderosa pine in Central Oregon. *Global Change Biology, 13*, 1748-1760.

751 Jackson, T.J., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., Doriaswamy, P., Hunt,
752 E.R. (2004). Vegetation water content mapping using Landsat data derived normalized
753 difference water index fro corn and soybeans. *Remote Sensing Environment, 92,* 475-482.
754 Jenkins, J.P., Richardson, A.D., Braswell, B.H., Ollinger, S.V., Hollinger, D.Y., Smith, M.-L.
755 (2007). Refining light-use efficiency calculations for a deciduous forest canopy using
756 simultaneous tower-based carbon flux and radiometric measurements. *Agricultural and*
757 *Forest Meteorology, 143,* 64-79.
758 Law, B.E. (2006). Carbon dynamics in response to climate and disturbance: recent progress
759 from multiscale measurements and modeling in AmeriFlux. In S. Yamamoto, ed. Plant
760 Responses to Air Pollution and Global Change. Springer, Tokyo, Japan.
761 Law, B.E., Sun, O.J., Campbell, J., Tuyl, S.V., & Thornton, P.E. (2003). Changes in carbon
762 storage and fluxes in a chronosequence of ponderosa pine. *Global Change Biology,* 9, 510-
763 524.
764 Law, B.E., Turner, D., Campbell, J., Sun, O.J., Van Tuyl, S., Ritts, W.D., & Cohen, W.B.
765 (2004). Disturbance and climate effects on carbon stocks and fluxes across Western
766 Oregon USA. *Global Change Biology, 10,* 1429-1444.
767 Law, B.E., Turner, D., Lefsky, M., Campbell, J., Guzy, M., Sun, O., Van Tuyl, S., Cohen, W.
768 (2006). Carbon fluxes across regions: Observational constraints at multiple scales. In J.
769 Wu, B. Jones, H. Li, O. Loucks, eds. *Scaling and Uncertainty Analysis in Ecology:*
770 *Methods and Applications.* Springer, USA. Pages 167-190.
771 Lipson, D.A., Wilson, R.F., Oechel, W.C. (2005). Effects of Elevated Atmospheric CO_2 on
772 Soil Microbial Biomass, Activity, and Diversity in a Chaparral Ecosystem. *Applied and*
773 *Environmental Microbiology, 71,* 8573-8580.

774 Lloyd, J., Farquhar, G.D. (1996). The CO_2 dependence of photosynthesis, plant growth
775 responses to elevated atmospheric CO_2 concentrations, and their interaction with soil
776 nutrient status: I. General principles and forest ecosystems. *Functional Ecology, 10*, 4 – 32.
777 Lutes, D.C., Keane, J.F., Caratti, C.H., Key, C.H., Benson, N.C., & Gangi, L.J. (2004).
778 *FIREMON: Fire Effects Monitoring and Inventory System* (US Department of Agriculture
779 Forest Service, Rocky Mountain Research Station, Ogden, UT), Vol RMRS-GTR-164-CD,
780 p 400.
781 Ma, S., Baldocchi, D.D., Xu, L., & Hehn, T. (2007). Inter-annual variability in carbon dioxide
782 exchange of an oak/grass savanna and open grassland in California. *Agricultural Forest*
783 *Meteorology, 147*, 157-171.
784 Mäkelä, A., Pulkkinen, M.A., Kolari, P. et al. (2008). Developing an empirical model of stand
785 GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer
786 sites in Europe. *Global Change Biology, 14*, 92-108.
787 McGarvey, R.C., Martin, T.A., & White, T.L. (2004). Integrating within-crown variation in net
788 photosynthesis in loblolly and slash pine families. *Tree Physiology, 24*, 1209-1220.
789 Minasny, B., McBratney, A.B. (2008). Regression rules as a tool for predicting soil properties
790 from infrared reflectance spectroscopy. Chemometrics and Intelligent Laboratory Systems,
791 94, 72-79.
792 Monson, R.K., Turnipseed, A.A., Sparks, J.P., Harley, P.C., Scott-Denton, L.E., Sparks, K., &
793 Huxman, T.E. (2002). Carbon sequestration in a high-elevation, subalpine forest. *Global*
794 *Change Biology, 8*, 459-478.

795 Moffat, A.M., Papale, D., Reichstein, M. et al. (2007). Comprehensive comparison of gap-
796 filling techniques for eddy covariance net carbon fluxes. *Agricultural Forest Meteorology*,
797 *147*, 209-232.
798 Myneni, R.B., Dong, J., Tucker, C.J., Kaufmann, R.K., Kauppi, P.E., Liski, J., Zhou, L.,
799 Alexeyev, V., Hughes, M.K. (2001). A large carbon sink in the woody biomass of northern
800 forests *Proc. Natl. Acad. Sci. U.S.A., 98*, 14784-14789.
801 Myneni, R.B., Hoffman, S., Knyazikhin, Y. et al. (2002). Global products of vegetation leaf
802 area and fraction absorbed PAR from year one of MODIS data. *Remote Sensing of*
803 *Environment*, *83*, 214-231.
804 Noormets, A., Chen, J., Crow, T.R., 2007. Age-dependent changes in ecosystem carbon fluxes
805 in managed forests in northern Wisconsin, USA. *Ecosystems* 10, 187-203.
806 Noormets, A., Desai, A.R., Cook, B.D., Euskirchen, E.S., Ricciuto, D.M., Davis, K.J., Bolstad,
807 P.V., Schmid, H.P., Vogel, C.V., Carey, E.V., Su, H.B., Chen, J., 2008a. Moisture
808 sensitivity of ecosystem respiration: Comparison of 14 forest ecosystems in the Upper
809 Great Lakes Region, USA. *Agricultural and Forest Meteorology* 148, 216-230.
810 Noormets, A., McNulty, S.G., DeForest, J.L., Sun, G., Li, Q., Chen, J., 2008b. Drought during
811 canopy development has lasting effect on annual carbon balance in a deciduous temperate
812 forest. *New Phytologist*: doi: 10.1111/j.1469-8137.2008.02501.x.
813 Noormets, A., McNulty, S.G., Gavazzi, M.J., Sun, G., Domec, J.C., King, J., Chen, J., 200X.
814 High drought tolerance of carbon fluxes in a coastal plain loblolly pine forest. *Global*
815 *Change Biology*, submitted.
816 Oak Ridge National Laboratory Distributed Active Archive Center (ORNL DAAC) (2006).
817 MODIS subsetted land products, Collection 4. Available on-line

818 (http://www.daac.ornl.gov/MODIS/modis.html) from ORNL DAAC, Oak Ridge,
819 Tennessee, U.S.A. Accessed Month 06, 2007.
820 Oren, R., Ewers, B.E., Todd, P., Phillips, N., & Katul, G. (1998). Water balance delineates the
821 soil layer in which moisture affects canopy conductance. *Ecological Applications, 8*, 990-
822 1002.
823 Oren, R., Hsieh, C.-I., Stoy, P., Albertson, J., McCarthy, H.R., Harrell, P., & Katul, G.G.
824 (2006). Estimating the uncertainty in annual net ecosystem carbon exchange: spatial
825 variation in turbulent fluxes and sampling errors in eddy-covariance measurements. *Global*
826 *Change Biology, 12*, 883-896.
827 Papale, D., & Valentini, A. (2003). A new assessment of European forests carbon exchange by
828 eddy fluxes and artificial neural network spatialization. *Global Change Biology, 9*, 525-
829 535.
830 Pataki, D.E., & Oren, R. (2003). Species difference in stomatal control of water loss at the
831 canopy scale in a bottomland deciduous forest. *Advances in Water Research, 26*, 1267-
832 1278.
833 Penuelas, J., Gamon, J.A., Griffin, K.L., & Field, C.B. (1993). Assessing community type,
834 plant biomass, pigment composition, and photosynthetic efficiency of aquatic vegetation
835 from spectral reflectance. *Remote Sensing of Environment, 46*, 110-118.
836 Persson, P., Hall-Könyves, K. Sjöström, G., Pinzke, S. (1993). NOAA/AVHRR data for crop
837 productivity estimation in Sweden. *Adv. Space Res., 13*, 111-116.
838 Peters, W., Jacobson, A.R., Sweeney, C. et al. (2007). An atmospheric perspective on North
839 American carbon dioxide exchange: carbontracker. *Proc Natl Acad Sci USA, 104*, 18925-
840 18930 (2007).

841 Potter, C., Klooster, S., Huete, A., & Genovese, V. (2007). Terrestrial carbon sinks for the
842 United States predicted from MODIS satellite data and ecosystem modeling. *Earth*
843 *Interactions, 11*, 1-21.
844 Powell, T.L., Gholz, H.L., Clark, K.L., Starr, G., Cropper, W.P., Martin, T.A. (2008). Carbon
845 exchange of a mature, naturally-regenerated pine forest in north Florida. *Global Change*
846 *Biology*, 14, 2523-2538, doi: 10.1111/j.1365-2486.2008.01675.x.
847 Prince, S.D., Goward, S.N. (1995). Global primary production: a remote sensing approach.
848 *Journal of Biogeography, 22*, 815-835.
849 PRISM Climate Group, Oregon State University, http://www.prismclimate.org, created 4 Feb
850 2004.
851 Rahman, A.F., Sims, D.A., Cordova, V.D., El-Masri, B.Z. (2005). Potential of MODIS EVI
852 and surface temperature for directly estimating per-pixel ecosystem C fluxes. *Geophysical*
853 *Research Letters, 32*, L19404, doi:10.1029/2005GL024127.
854 Ranson, K.J., Daughtry, C.S.T., Biehl, L.L., Bauer, M.E. (1985). Sun-view angle effects on
855 reflectance factors of corn canopies. *Remote Sensing of Environment, 18*, 147-161.
856 Reichstein, M., Falge, E., Baldocchi, D. et al. (2005). On the separation of net ecosystem
857 exchange into assimilation and ecosystem respiration: review and improved algorithm.
858 *Global Change Biology, 11*, 1424-1439.
859 Richardson, A.D., Hollinger, D.Y., Dail, D.B., Lee, J.T., Munger, J.W., O'Keefe, J. (2008a).
860 Influence of spring phenology on seasonal and annual carbon balance in two contrasting
861 New England forests. *Global Change Biology*, under review.

862 Richardson, A.D., Mahecha, M.D., Falge, E. et al. (2008b). Statistical properties of random
863 CO_2 flux measurement uncertainty inferred from model residuals. *Agricultural and Forest*
864 *Meteorology, 148*, 38-50.
865 Ruimy, A., Jarvis, P.G., Baldocchi, D.D., & Saugier, B. (1995). CO2 fluxes over plant
866 canopies and solar radiation: a review. *Advances in Ecological Research, 26*, 1-68.
867 RuleQuest, 2008. http://www.rulequest.com. Visited on 10/18/2007.
868 Running, S.W., Baldocchi, D.D., Turner, D.P., Gower, S.T., Bakwin, P.S., & Hibbard, K.A.
869 (1999). A global terrestrial monitoring network integrating tower fluxes, flask sampling,
870 ecosystem modeling and EOS satellite data. *Remote Sensing of Environment, 70*, 108-127.
871 Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., & Hashimoto, H. (2004).
872 A continuous satellite-derived measure of global terrestrial primary production.
873 *BioScience, 54*, 547-560.
874 Ryan, M.G., Binkley, D., Fownes, J.H., Giardina, C.P., Senock, R.S. (2004). An experimental
875 test of the causes of forest growth decline with stand age. *Ecological Monographs, 74*, 393-
876 414.
877 Schmid, H.P. (1994). Source areas for scalars and scalar fluxes. *Boundary Layer Meteorology,*
878 *67*, 293-318.
879 Schmid, H.P. (2002). Footprint modeling for vegetation atmospheric exchange studies.: A
880 review and perspective. *Agricultural and Forest Meteorology, 113*, 159-183.
881 Schmid, H.P., Grimmond, C.S.B., Cropley, F., Offerle, B., & Su, H.B. (2000). Measurements
882 of CO_2 and energy fluxes over a mixed hardwood forest in the mid-western United States.
883 *Agricultural and Forest Meteorology, 103*, 357-374.

884 Schmidt, K.S., & Skidmore, A.K. (2003). Spectral discrimination of vegetation types in a
885 coastal wetland. *Remote Sensing of Environment, 85*, 92-108.
886 Sellers, P.J., Randall, D.A., Betts, A.K. et al. (1997). Modeling the exchanges of energy, water,
887 and carbon between continents and the atmosphere. *Science, 275*, 502– 509.
888 Shroyer, J.P., Thompson, C., Brown, R., Ohlenbach, P.D., Fjell, D.L., Staggenborg, S.,
889 Duncan, S., & Kilgore, G.L. (1996). Kansas crop planting guide. Publication, Vol. L-818
890 (pp. 2). Manhattan, KS: Kansas State University.
891 Sims, D.A., Rahman, A.F., Cordova, V.D. et al. (2005). Midday values of gross CO_2 flux and
892 light use efficiency during satellite overpasses can be used to directly estimate eight-day
893 mean flux. *Agricultural and Forest Meteorology, 131*, 1-12.
894 Stoy, P.C., Katul, G.G., Siqueira, M.B.S., Juang, J.Y., Novick, K.A., Uebelherr, J.M., Oren, R.
895 (2006). An evaluation of models for partitioning eddy covariance-measured net ecosystem
896 exchange into photosynthesis and respiration. *Agricultural and Forest Meteorology, 141*,
897 2-18.
898 Thompson, J.R., Spies, T.A., Ganio, L.M. (2007). Reburn severity in managed and unmanaged
899 vegetation in a large wildfire. *Proceedings of the National Academy of Sciences of the*
900 *United States of America, 104*, 10743-10748.
901 Tucker, C.J., Vanpraet, C.L., Sharman, M.J., Van Ittersum, G. (1985). Satellite remote sensing
902 of total herbaceous biomass production in the Senegalese Sahel: 1980-1984. *Remote Sens.*
903 *Environ., 17*, 233-249.
904 Urbanski, S., Barford, C., Wofsy, S. et al. (2007). Factors controlling CO_2 exchange on
905 timescales from hourly to decadal at Harvard Forest. *Journal of Geophysical Research,*
906 *112*, G02020, doi:10.1029/2006JG000293.

907 Verma, S.B., Dobermann, A., & Cassman, K.G. (2005). Annual carbon dioxide exchange in
908 irrigated and rainfed maize-based agroecosystems. *Agricultural and Forest Meteorology, 131*, 77-96.
909
910 Vermote, E.F., Vermeulen, A. (1999). MODIS Algorithm Technical Background Document –
911 Atmospheric Correction Algorithm: Spectral Reflectances (MOD09), Version 4.0.
912 http://modis.gsfc.nasa.gov/data/atbd/atbd_mod08.pdf.
913 Vermote, E.F., Kotchenova, S.Y. (2008). MOD09 (Surface Reflectance) User's Guide, Version
914 1.1, March, 2008, http://modis-sr.ltdri.org.
915 Wan, Z., Zhang, Y., Zhang, Q., & Li, Z.-L. (2002). Validation of the land-surface temperature
916 products retrieved from Terra Moderate Resolution Imaging Spectroradiometer data.
917 *Remote Sensing of Environment, 83*, 163-180.
918 Waring, R.H., & Franklin, J.F. (1979). Evergreen coniferous forests of the Pacific Northwest.
919 *Science, 204*, 1380-1386.
920 Waring, R.H., Landsberg, J.J., & Williams, M. (1998), Net primary production of forests: A
921 constant fraction of gross primary production? *Tree Physiology, 18*, 129– 134.
922 Watts, C.J., Scott, R.L., Garatuza-Payan, J., Rodriguez, J.C., Prueger, J., Kustas, W., Douglas,
923 M. (2007). Changes in vegetation condition and surface fluxes during NAME 2004.
924 *Journal of Climate, 20*, 1810-1820.
925 Wofsy, S.C., Goulden, M.L., Munger, J.W., Fan, S.-M., Bakwin, P.S., Daube, B.C., Bassow,
926 S.L., & Bazzaz, F.A. (1993). Net exchange of CO_2 in a mid-latitude forest. *Science, 260*,
927 1314-1317.

928 Wylie, B.K., Fosnight, E.A., Gilmanov, T.G. et al. (2007). Adaptive data-driven models for
929 estimating carbon fluxes in the Northern Great Plains. *Remote Sensing of Environment,*
930 *106*, 399-413.
931 Xiao, J., Zhuang, Q., Baldocchi, D.D. et al. (2008). Estimation of net ecosystem carbon
932 exchange for the conterminous United States by combining MODIS and AmeriFlux data.
933 *Agricultural and Forest Meteorology, 148*, 1827-1847.
934 Xiao, J., Zhuang, Q., Liang, E., McGuire, A.D., Moody, A., Kicklighter, D.W., Shao, X.,
935 Melillo, J.M. (2009). Twentieth-century droughts and their impacts on terrestrial carbon
936 cycling in China. *Earth Interactions, 13*, Paper No. 10, 1-31, DOI: 10.1175/2009EI275.1
937 Xu, L., & Baldocchi, D.D. (2004). Seasonal variation in carbon dioxide exchange over a
938 Mediterranean annual grassland in California. *Agricultural and Forest Meteorology, 123*,
939 79-96.
940 Yang, L., Huang, C., Homer, C., Wylie, B.K., Coan, M.J. (2003). An approach for mapping
941 large-area impervious surfaces: synergistic use of Landsat-7 ETM+ and high spatial
942 resolution imagery. *Canadian Journal of Remote Sensing, 29*, 230-240.
943 Yang, F., Ichii, K., White, M.A., Hashimoto, H., Michaelis, A.R., Votava, P., Zhu, A.-X.,
944 Huete, A., Running, S.W., & Nemani, R.R. (2007). Developing a continental-scale measure
945 of gross primary production by combining MODIS and AmeriFlux data through support
946 machine approach. *Remote Sensing of Environment, 110*, 109-122.
947 Zhang, L., Wylie, B., Loveland, T., Fosnight, E., Tieszen, L.L., Ji, L., & Gilmanov, T. (2007).
948 Evaluation and comparison of gross primary production estimates for the Northern Great
949 Plains grasslands. *Remote Sensing of Environment, 106*, 173-189.

950 Zhang, Y., Yu, Q., Jiang, J., Tang, Y. (2008). Calibration of Terra/MODIS gross primary
951 production over an irrigated cropland on the North China Plain and an alpine meadow on
952 the Tibetan Plateau. *Global Change Biology, 14*, 757-767.
953 Zhao, M., Heinsch, F.A., Nemani, R.R., & Running, S.W. (2005). Improvements of the
954 MODIS terrestrial gross and net primary production global data set. *Remote Sensing of*
955 *Environment, 95*, 164-175.
956 Zhao, M., Running, S.W., & Nemani, R.R. (2006). Sensitivity of moderate resolution imaging
957 spectroradiometer (MODIS) terrestrial primary production to the accuracy of
958 meteorological reanalyses. *Journal of Geophysical Research, 111*, G01002,
959 doi:10.1029/2004JG000004.
960
961
962
963
964
965
966
967
968
969
970
971
972

973 Table 1. Site descriptions including name, latitude, longitude, vegetation type, years of data

974 available, and references for each flux site.

Site	State	Lat	Lon	Vegetation type	Year	References
Audubon Research Ranch (ARR)	AZ	31.59	-110.51	Grasslands	2002-2006	
Santa Rita Mesquite (SRM)	AZ	31.82	-110.87	Savannas	2004-2006	Watts et al. 2007
Walnut Gulch Kendall Grasslands (WGK)	AZ	31.74	-109.94	Grasslands	2004-2006	
Sky Oaks Old Stand (SOO)	CA	33.37	-116.62	Shrublands	2004-2006	Lipson et al. 2005
Sky Oaks Young stand (SOY)	CA	33.38	-116.62	Shrublands	2001-2006	Lipson et al. 2005
Tonzi Ranch (TR)	CA	38.43	-120.97	Savannas	2001-2006	Ma et al. 2007
Vaira Ranch (VR)	CA	38.41	-120.95	Grasslands	2001-2006	Xu and Baldocchi 2004
Niwot Ridge Forest (NRF)	CO	40.03	-105.55	Evergreen forests	2000-2003	Monson et al. 2002
Kennedy Space Center -Scrub Oak (KSC)	FL	28.61	-80.67	Shrublands	2000-2006	Dore et al. 2003
Austin Cary - Slash Pine (AC)	FL	29.74	-82.22	Evergreen forests	2001-2005	Powell et al. 2008
Bondville (Bon)	IL	40.01	-88.29	Croplands	2001-2006	Hollinger et al. 2005
FNAL agricultural site (FAg)	IL	41.86	-88.22	Croplands	2005-2006	
FNAL Prairie site (FPr)	IL	41.84	-88.24	Grasslands	2004-2006	
Morgan Monroe State Forest (MMS)	IN	39.32	-86.41	Deciduous forests	2000-2005	Schmid et al. 2000
Harvard Forest EMS Tower (HFE)	MA	42.54	-72.17	Deciduous forests	2000-2004	Urbanski et al. 2007
Harvard Forest Hemlock Site (HFH)	MA	42.54	-72.18	Evergreen forests	2004	
Little Prospect Hill (LPH)	MA	42.54	-72.18	Deciduous forests	2002-2005	
Howland forest (HF)	ME	45.20	-68.74	Evergreen forests	2000-2004	Hollinger et al., 1999, 2004
Howland forest (west tower) (HFW)	ME	45.21	-68.75	Deciduous forests	2000-2004	Hollinger et al., 1999, 2004
Sylvania Wilderness Area (SWA)	MI	46.24	-89.35	Mixed forests	2002-2006	Desai et al., 2005
Univ. of Mich. Biological Station (UMB)	MI	45.56	-84.71	Mixed forests	2000-2003	Gough et al., 2008
Missouri Ozark (MO)	MO	38.74	-92.20	Deciduous forests	2004-2006	Gu et al. 2006, 2007
Goodwin Creek (GC)	MS	34.25	-89.97	Grasslands	2002-2006	
Fort Peck (FPe)	MT	48.31	-105.10	Grasslands	2000-2006	
Duke Forest loblolly pine (DFP)	NC	35.98	-79.09	Evergreen forests	2001-2005	Oren et al. 1998, 2006

Duke Forest hardwoods (DFH)	NC	35.97	-79.10	Deciduous forests	2003-2005	Pataki and Oren, 2003
North Carolina loblolly pine (NCP)	NC	35.80	-76.67	Evergreen forests	2005-2006	Noormets et al. submitted
Mead irrigated continuous maize site (MIC)	NE	41.17	-96.48	Croplands	2001-2005	Verma et al. 2005
Mead irrigated rotation (MIR)	NE	41.16	-96.47	Croplands	2001-2005	Verma et al. 2005
Mead rainfed (MR)	NE	41.18	-96.44	Croplands	2001-2005	Verma et al. 2005
Bartlett Experimental Forest (BEF)	NH	44.06	-71.29	Deciduous forests	2004-2005	Jenkins et al., 2007
Toledo Oak Openings (TOO)	OH	41.55	-83.84	Savannas	2004-2005	Noormets et al. 2008b
ARM Oklahoma (ARM)	OK	36.61	-97.49	Croplands	2003-2006	
Metolius intermediate aged ponderosa pine (MI)	OR	44.45	-121.56	Evergreen forests	2003-2005	Law et al. 2003; Irvine et al. 2007
Metolius new young pine (MN)	OR	44.32	-121.61	Evergreen forests	2004-2005	Law et al. 2003; Irvine et al. 2007
Brookings (Bro)	SD	44.35	-96.84	Grasslands	2004-2006	
Freeman Ranch Mesquite Juniper (FRM)	TX	29.95	-98.00	Savannas	2004-2006	
Wind River Crane Site (WRC)	WA	45.82	-121.95	Evergreen forests	2000-2004	Falk et al., 2008
Lost Creek (LC)	WI	46.08	-89.98	Deciduous forests	2000-2005	
Willow Creek (WC)	WI	45.81	-90.08	Deciduous forests	2000-2006	Cook et al., 2004
Wisconsin intermediate hardwood (WIH)	WI	46.73	-91.23	Deciduous forests	2003	Noormets et al. 2008a
Wisconsin mature red pine (MRP)	WI	46.74	-91.17	Evergreen forests	2002-2005	Noormets et al. 2007

975

976

977

978

979

980

981

982

983

Table 2. Spatially averaged (spatial mean) and integrated (spatial total) annual GPP over the period 2001-2006 for each region: Northeast (NE), North Central (NC), Southeast (SE), South Central (SC), Rocky Mountain (RM), Pacific Northwest (PNW), and Pacific Southwest (PSW).

GPP	NE	NC	SE	SC	RM	PNW	PSW	US
Spatial Mean ($g\ C\ m^{-2}\ yr^{-1}$)	1604.20	1212.55	2033.91	1457.97	438.20	1007.35	927.77	1103.92
Spatial Total ($Pg\ C\ yr^{-1}$)	0.67	2.00	1.00	1.93	0.81	0.36	0.29	7.06

1000 Table 3. Spatially averaged (spatial mean) and integrated (spatial total) annual GPP over the
1001 period 2001-2006 for each vegetation type: evergreen forests (EF), deciduous forests (DF),
1002 mixed forests (MF), shrublands (Sh), savannas (Sa), grasslands (Gr), and croplands (Cr).

GPP	EF	DF	MF	Sh	Sa	Gr	Cr	All
Spatial Mean ($g\ C\ m^{-2}\ yr^{-1}$)	1431.26	1774.74	1447.31	303.78	1500.96	589.24	1500.38	1103.92
Spatial Total ($Pg\ C\ yr^{-1}$)	0.85	0.83	0.67	0.33	0.63	0.83	2.92	7.06

Figure captions:

Fig. 1. Scatterplots of observed 8-day GPP versus predicted 8-day GPP. (a) Our estimate ($y = 0.85x + 0.37$, $R^2 = 0.74$, $p < 0.0001$). (b) MODIS GPP product (Running et al. 2004) ($y = 0.50x + 1.01$, $R^2 = 0.58$, $p < 0.0001$). For each plot, the solid line is the 1:1 line, and the dashed line is the regression line.

Fig. 2. Examples of time series plots of observed (open circles) and predicted (solid circles) 8-day GPP (g C m^{-2} day^{-1}) for each AmeriFlux site over the period 2005-2006: (1) evergreen forests – AC (FL) and MRP (WI); (2) deciduous forests – DFH (NC); (3) mixed forests – SWA (MI); (4) shrublands – KSC (FL); (5) savannas – SRM (AZ), TR (CA), and VR (CA); (6) grasslands – ARR (AZ); (7) croplands – MIC (NE). For x-axis, the starting dates (month/day) of every two 8-day intervals are provided in parentheses under interval numbers. Dashed lines are used to separate 2005 from 2006. Site abbreviations are used here, and their full names are given in Table 1.

Fig. 3. Scatterplot of observed mean 8-day GPP versus predicted mean 8-day GPP across the AmeriFlux sites. Error bars are standard errors (defined as the standard deviation divided by the square root of the number of observations) of the observed and predicted 8-day mean GPP. The solid line indicates the 1:1 line, and the dashed line indicates the regression line ($y = 0.95 * x + 0.21$, $R^2 = 0.84$, $p < 0.0001$). Site abbreviations are used here, and their full names are given in Table 1.

Fig. 4. Examples of leave-one-out cross-validation scatterplots with observed 8-day GPP versus predicted 8-day GPP: (1) evergreen forests – HF ($y = 0.85x+1.00$, $R^2 = 0.91$, $p < 0.0001$) and NRF ($y = 0.67x + 0.63$, $R^2 = 0.63$, $p < 0.0001$); (2) deciduous forests – HFE ($y = 0.97x + 0.24$, $R^2 = 0.87$, $p < 0.0001$) and WC ($y = 0.73x + 0.62$, $R^2 = 0.86$, $p < 0.0001$); (3)

mixed forests – SWA ($y = 0.89x + 0.03$, $R^2 = 0.89$, $p < 0.0001$) and UMB ($y = 0.92x + 0.66$, $R^2 = 0.93$, $p < 0.0001$); (4) shrublands – KSC ($y = 0.28x + 2.35$, $R^2 = 0.14$, $p < 0.0001$); (5) savannas – FRM ($y = 0.83x + 1.94$, $R^2 = 0.49$, $p < 0.001$); (6) grasslands – GC ($y = 0.76x + 2.03$, $R^2 = 0.55$, $p < 0.0001$); (7) croplands – MR ($y = 1.04x + 0.81$, $R^2 = 0.86$, $p < 0.0001$). The solid line indicates the 1:1 line, and the dashed line indicates the regression line. Site abbreviations are used here, and their full names are given in Table 1.

Fig. 5. Leave-one-out cross-validation scatterplot of observed mean 8-day GPP versus predicted mean 8-day GPP across the AmeriFlux sites. Error bars are standard errors of the observed and predicted 8-day GPP. The solid line indicates the 1:1 line, and the dashed line indicates the regression line ($y = 0.73 * x + 1.07$, $R^2 = 0.69$, $p < 0.0001$). Site abbreviations are used here, and their full names are given in Table 1.

Fig. 6. Monthly GPP (g C m^{-2} mo^{-1}) for the conterminous U.S. from January through December in 2005.

Fig. 7. Spatially averaged and integrated 8-day GPP for each biome across the conterminous U.S. over the period 2001-2006. (a) Spatially averaged 8-day GPP (g C m^{-2} day^{-1}). (b) Spatially integrated 8-day GPP (Tg C day^{-1}).

Fig. 8. Average annual GPP (g C m^{-2} yr^{-1}) of the conterminous U.S. over the period 2001-2006. The gray lines indicate state boundaries. The black lines indicate boundaries of geographical regions: Northeast (NE), Southeast (SE), North Central (NC), South Central (SC), Rocky Mountain (RM), Pacific Northwest (PNW), and Pacific Southwest (PSW).

Fig. 9. Annual GPP (g C m^{-2} yr^{-1}) for the conterminous U.S. for 2005. (a) Our estimate. (b) The MODIS GPP product (MOD17A3; Running et al. 2004).

Fig. 10. Annual GPP anomalies (g C m^{-2} yr^{-1}) and annual precipitation anomalies (mm) for the conterminous U.S. for 2002 and 2006. The anomalies of annual GPP were relative to the 6-year period from 2001 to 2006, and the anomalies of annual precipitation were relative to the 30-year period from 1971 to 2000 taken from the PRISM climate database (PRISM Climate Group 2004).

Fig. 11. Impact of the Biscuit Fire (> 2000 km^2) in Oregon on annual GPP in 2004. (a) Burned area. (b) The dots represent fire detections from Terra MODIS and Aqua MODIS MODIS (USDA Forest Service MODIS Active Fire Mapping Program, http://activefiremaps.fs.fed.us). (c) Burn severity based on the difference normalized burn ratio (dNBR; Lutes et al. 2004) from Landsat Thematic Mapper (TM) data acquired before and immediately after the fire. (d) Annual GPP in 2003 (g C m^{-2} yr^{-1}).

Fig. 12. Impact of Hurricane Katrina on annual GPP in 2006. The units are g C m^{-2} yr^{-1}. The white lines indicate the isotachs, including tropical storm, hurricane category 1, and hurricane category 2.

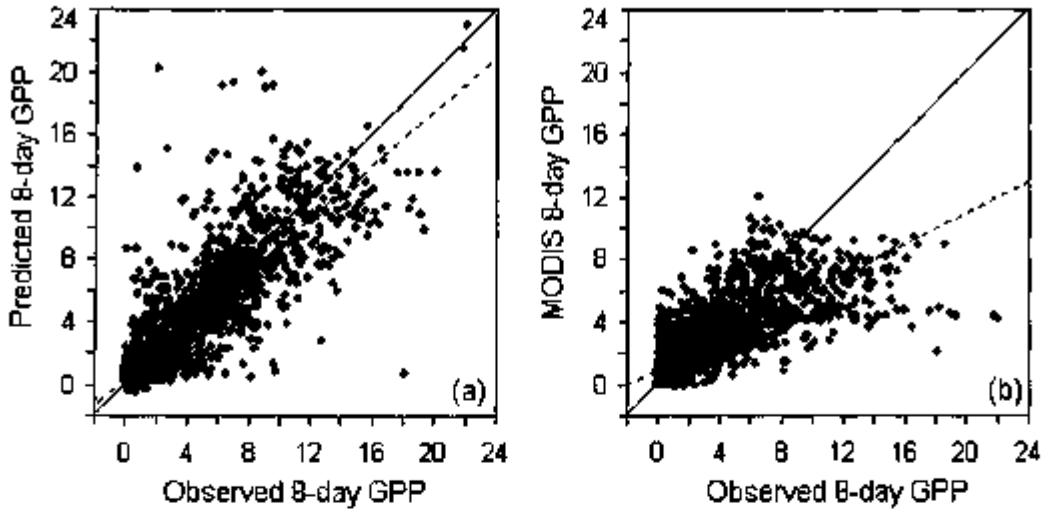

Fig. 1. Scatterplots of observed 8-day GPP versus predicted 8-day GPP. (a) Our estimate ($y = 0.85x + 0.37$, $R^2 = 0.74$, $p < 0.0001$). (b) MODIS GPP product (Running et al. 2004) ($y = 0.50x + 1.01$, $R^2 = 0.58$, $p < 0.0001$). For each plot, the solid line is the 1:1 line, and the dashed line is the regression line.

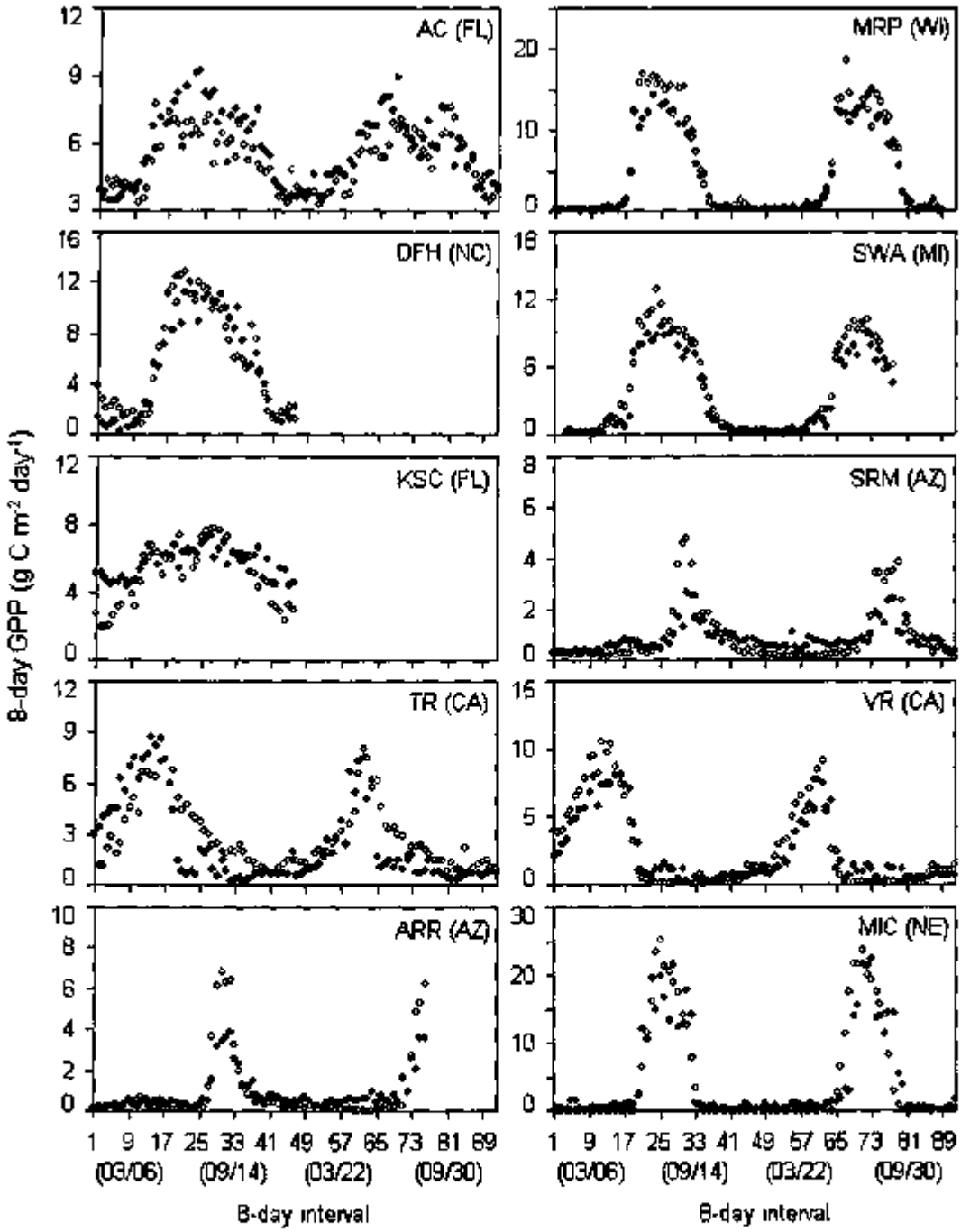

Fig. 2. Examples of time series plots of observed (open circles) and predicted (solid circles) 8-day GPP (g C m^2 day^{-1}) for each AmeriFlux site over the period 2005-2006: (1) evergreen forests – AC (FL) and MRP (WI); (2) deciduous forests – DFH (NC); (3) mixed forests – SWA (MI); (4) shrublands – KSC (FL); (5) savannas – SRM (AZ), TR (CA), and VR (CA); (6)

grasslands – ARR (AZ); (7) croplands – MIC (NE). For x-axis, the starting dates (month/day) of every two 8-day intervals are provided in parentheses under interval numbers. Dashed lines are used to separate 2005 from 2006. Site abbreviations are used here, and their full names are given in Table 1.

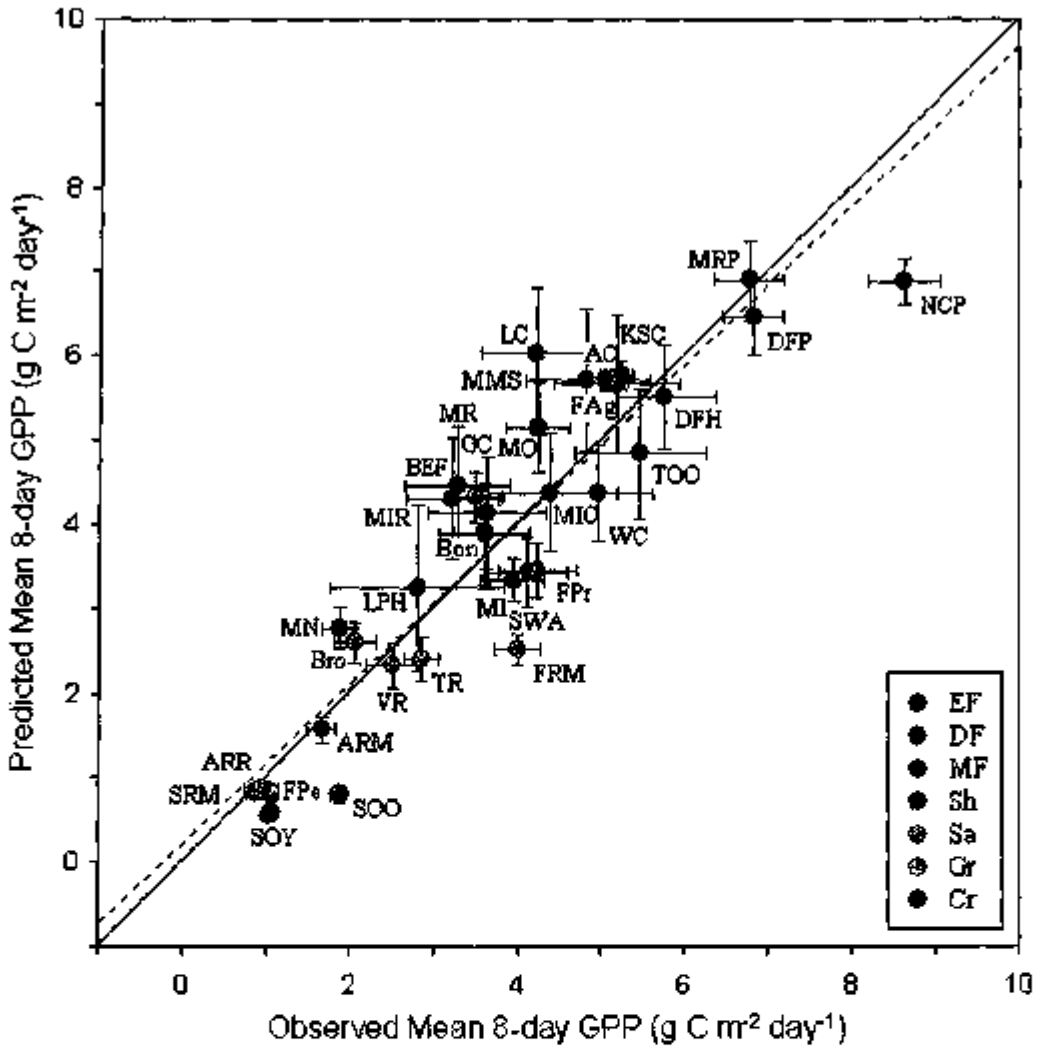

Fig. 3. Scatterplot of observed mean 8-day GPP versus predicted mean 8-day GPP across the AmeriFlux sites. Error bars are standard errors (defined as the standard deviation divided by the square root of the number of observations) of the observed and predicted 8-day mean GPP. The solid line indicates the 1:1 line, and the dashed line indicates the regression line ($y = 0.95 * x + 0.21$, $R^2 = 0.84$, $p < 0.0001$). Site abbreviations are used here, and their full names are given in Table 1.

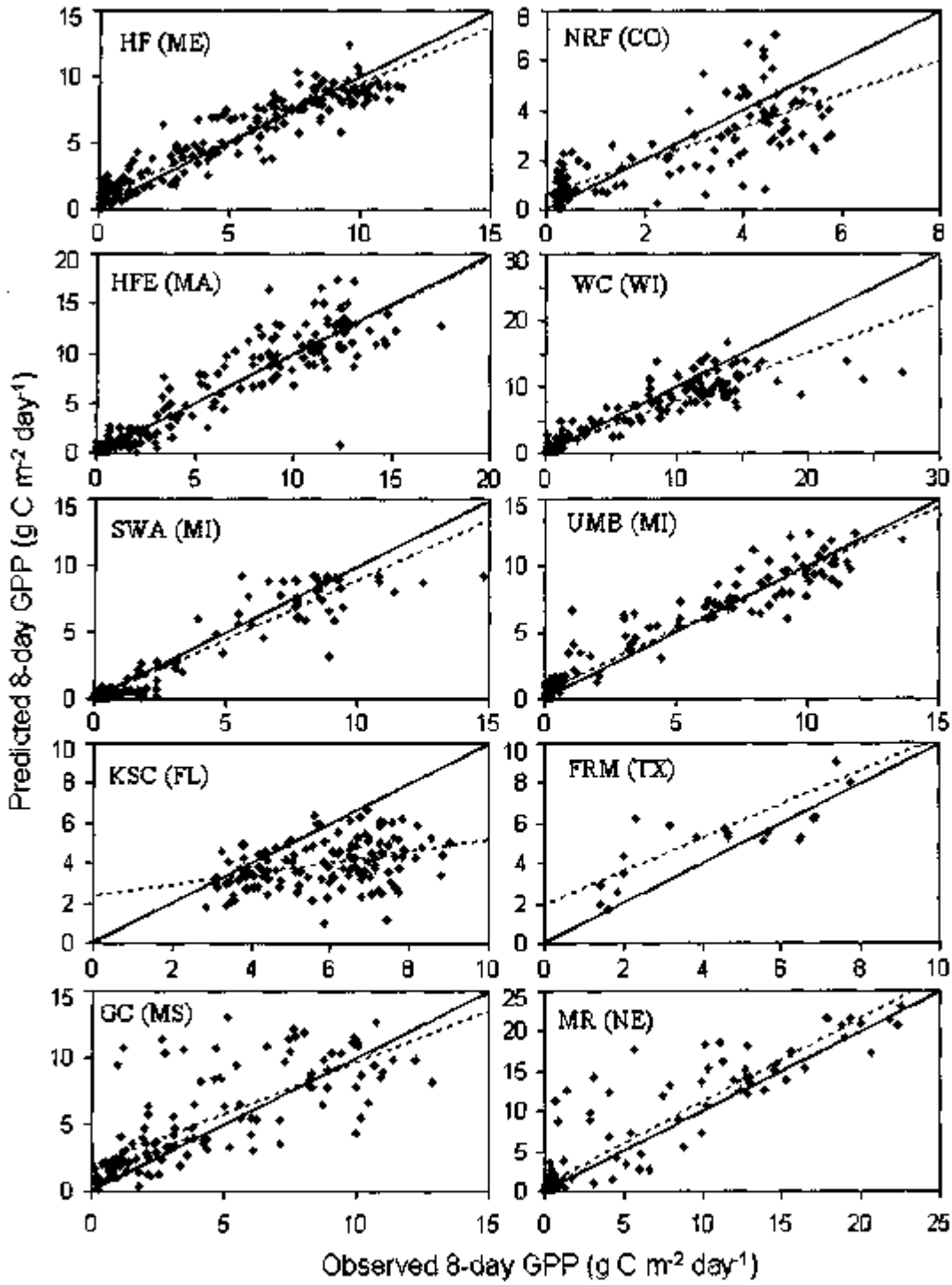

Fig. 4. Examples of leave-one-out cross-validation scatterplots with observed 8-day GPP versus predicted 8-day GPP: (1) evergreen forests – HF ($y = 0.85x+1.00$, $R^2 = 0.91$, $p <$

0.0001) and NRF ($y = 0.67x + 0.63$, $R^2 = 0.63$, $p < 0.0001$); (2) deciduous forests – HFE ($y = 0.97x + 0.24$, $R^2 = 0.87$, $p < 0.0001$) and WC ($y = 0.73x + 0.62$, $R^2 = 0.86$, $p < 0.0001$); (3) mixed forests – SWA ($y = 0.89x + 0.03$, $R^2 = 0.89$, $p < 0.0001$) and UMB ($y = 0.92x + 0.66$, $R^2 = 0.93$, $p < 0.0001$); (4) shrublands – KSC ($y = 0.28x + 2.35$, $R^2 = 0.14$, $p < 0.0001$); (5) savannas – FRM ($y = 0.83x + 1.94$, $R^2 = 0.49$, $p < 0.001$); (6) grasslands – GC ($y = 0.76x + 2.03$, $R^2 = 0.55$, $p < 0.0001$); (7) croplands – MR ($y = 1.04x + 0.81$, $R^2 = 0.86$, $p < 0.0001$). The solid line indicates the 1:1 line, and the dashed line indicates the regression line. Site abbreviations are used here, and their full names are given in Table 1.

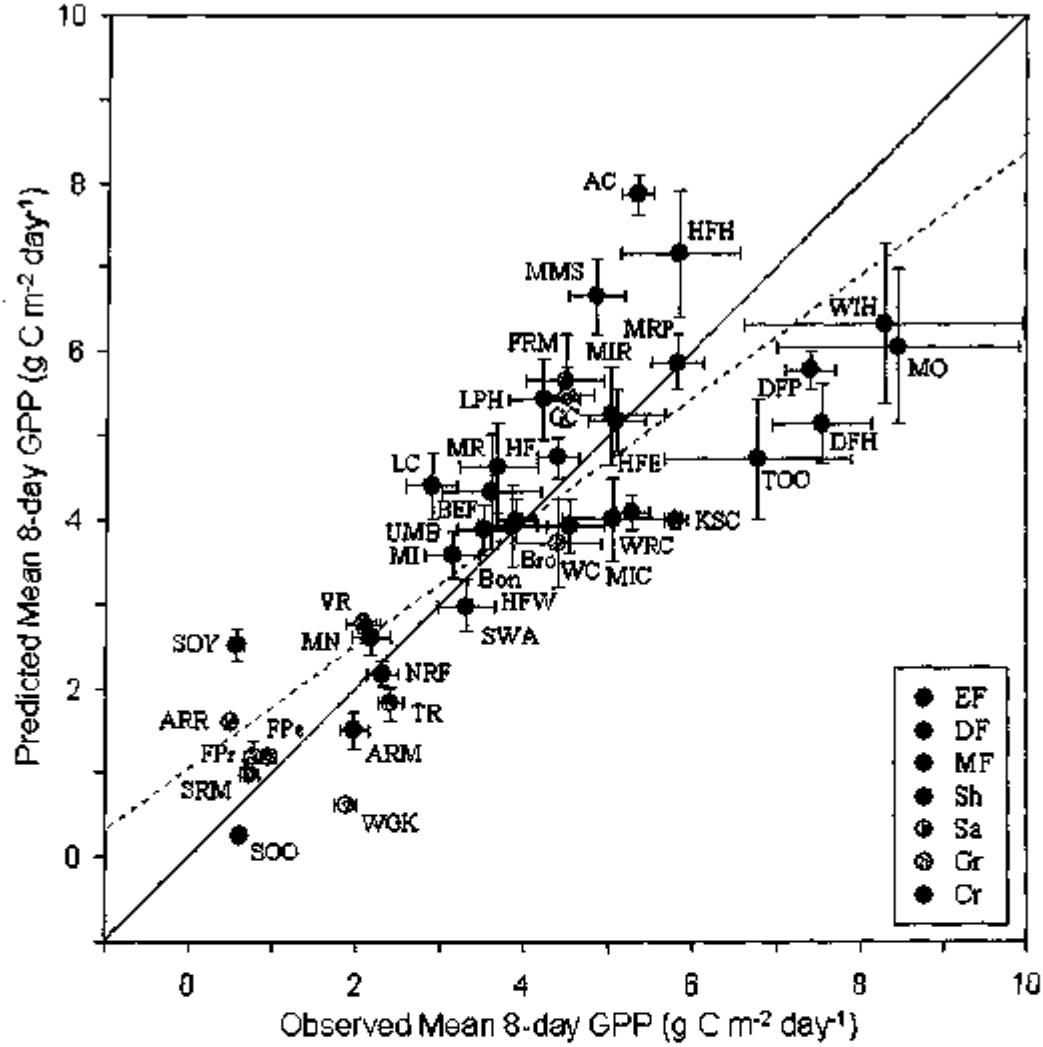

Fig. 5. Leave-one-out cross-validation scatterplot of observed mean 8-day GPP versus predicted mean 8-day GPP across the AmeriFlux sites. Error bars are standard errors of the observed and predicted 8-day GPP. The solid line indicates the 1:1 line, and the dashed line indicates the regression line ($y = 0.73 * x + 1.07$, $R^2 = 0.69$, $p < 0.0001$). Site abbreviations are used here, and their full names are given in Table 1.

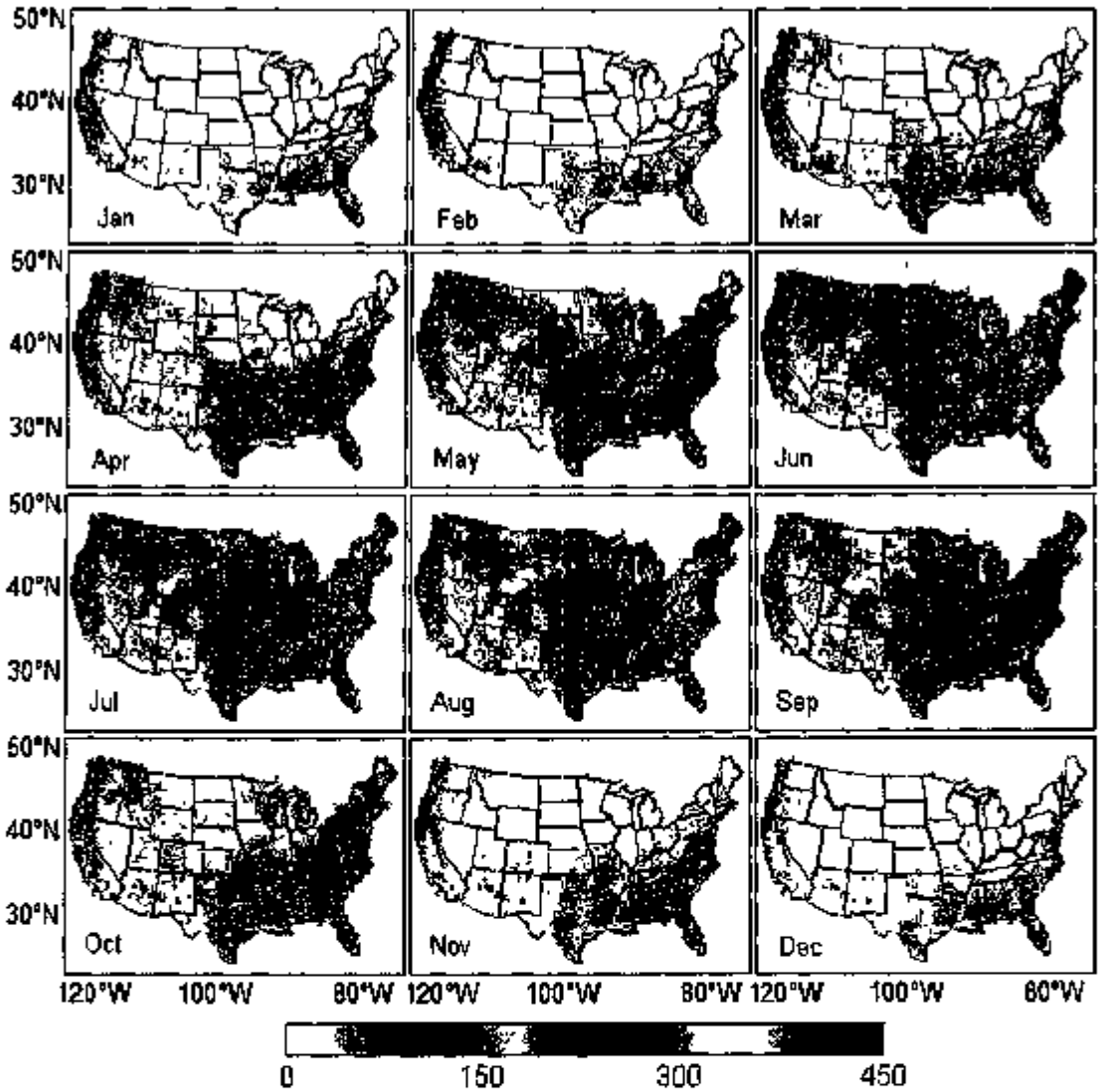

Fig. 6. Monthly GPP (g C m^{-2} mo^{-1}) for the conterminous U.S. from January through December in 2005.

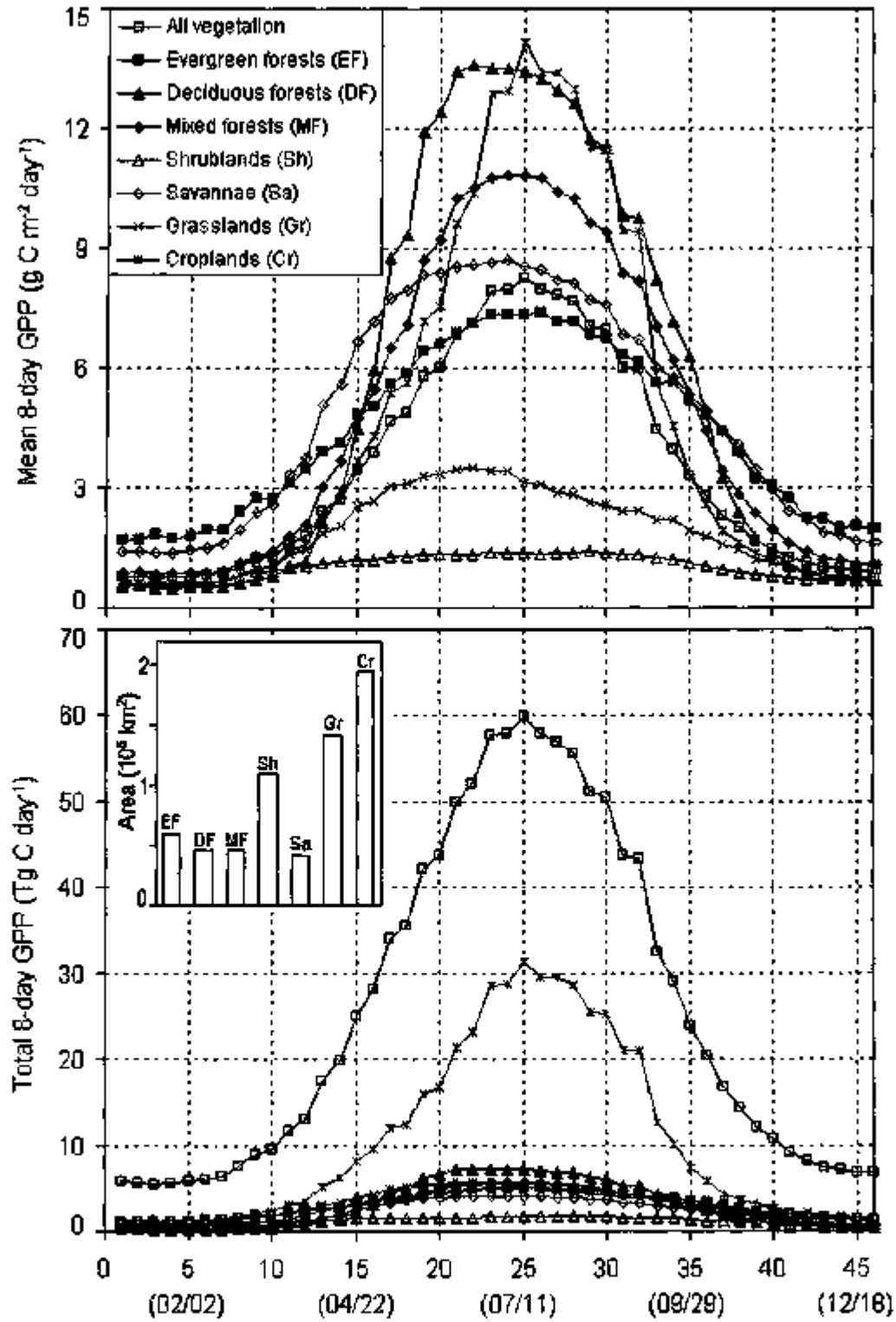
All vegetation
Evergreen forests (EF)
Deciduous forests (DF)
Mixed forests (MF)
Shrublands (Sh)
Savannae (Sa)
Grasslands (Gr)
Croplands (Cr)
Mean 8-day GPP (g C m-2 day-1)
Total 8-day GPP (Tg C day-1)
Area (10^6 km^2)
EF
DF
MF
Sh
Sa
Gr
Cr
(02/02)
(04/22)
(07/11)
(09/29)
(12/18)

Fig. 7. Spatially averaged and integrated 8-day GPP for each biome across the conterminous U.S. over the period 2001-2006. (a) Spatially averaged 8-day GPP (g C m^{-2} day^{-1}). (b) Spatially integrated 8-day GPP (Tg C day^{-1}).

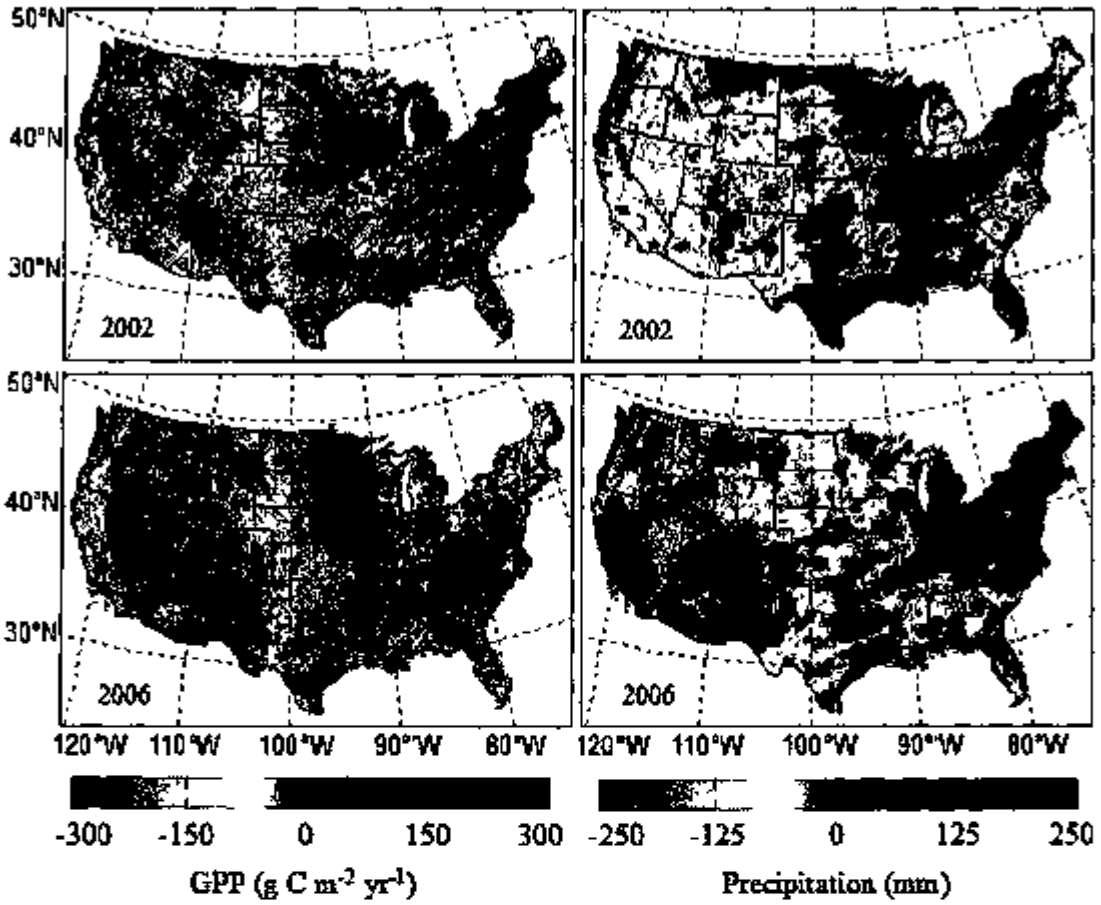

Fig. 10. Annual GPP anomalies (g C m^{-2} yr^{-1}) and annual precipitation anomalies (mm) for the conterminous U.S. for 2002 and 2006. The anomalies of annual GPP were relative to the 6-year period from 2001 to 2006, and the anomalies of annual precipitation were relative to the 30-year period from 1971 to 2000 taken from the PRISM climate database (PRISM Climate Group 2004).

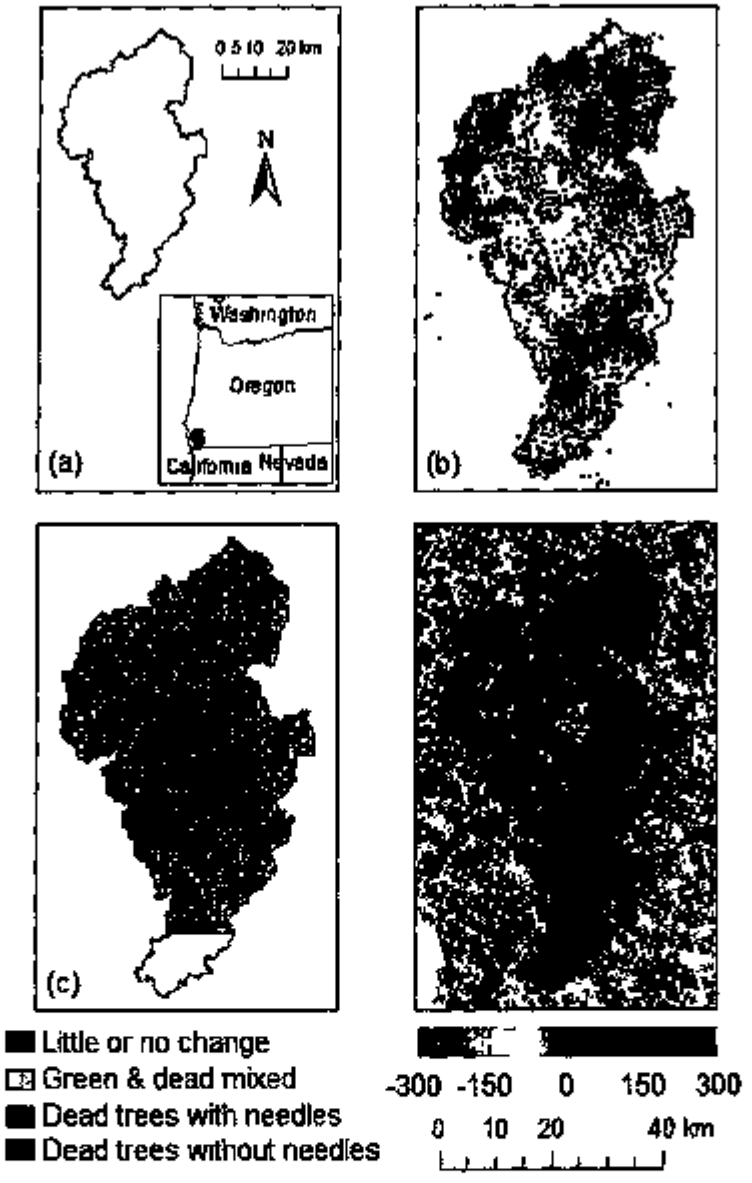

Fig. 11. Impact of the Biscuit Fire (> 2000 km^2) in Oregon on annual GPP in 2004. (a) Burned area. (b) The dots represent fire detections from Terra MODIS and Aqua MODIS MODIS (USDA Forest Service MODIS Active Fire Mapping Program, http://activefiremaps.fs.fed.us). (c) Burn severity based on the difference normalized burn ratio (dNBR; Lutes et al. 2004) from Landsat Thematic Mapper (TM) data acquired before and immediately after the fire. (d) Annual GPP in 2003 (g C m^{-2} yr^{-1}).

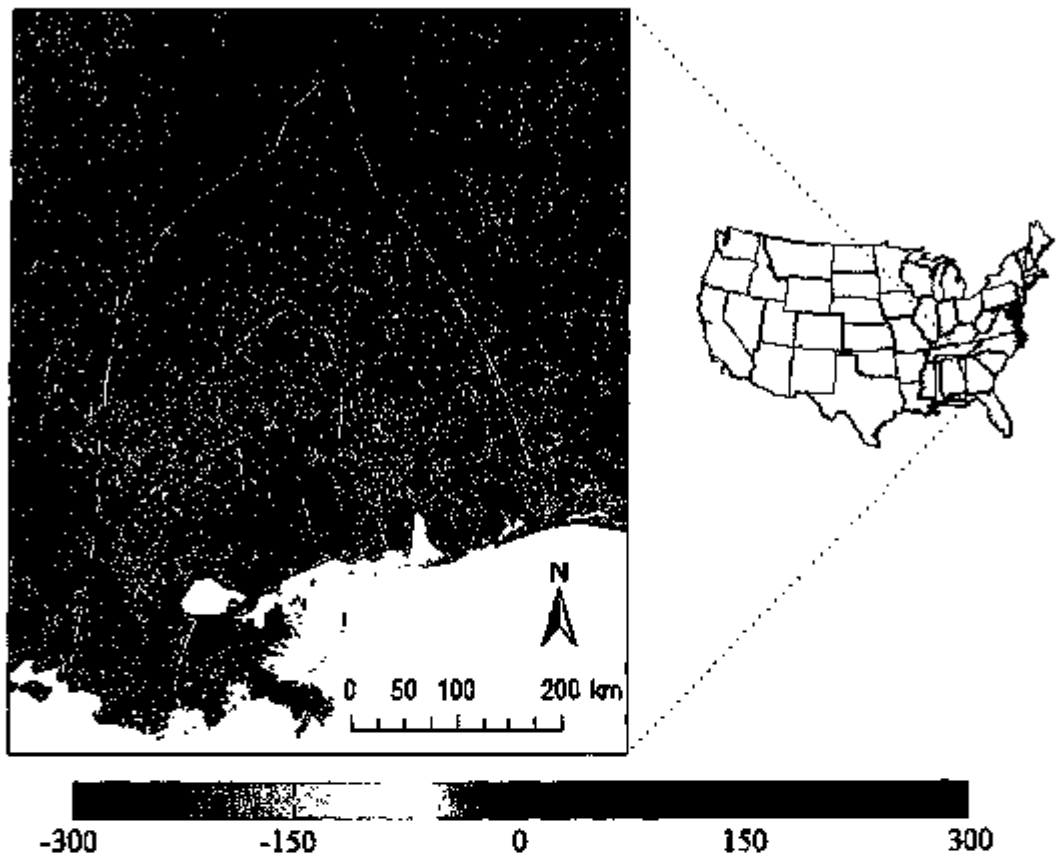

Fig. 12. Impact of Hurricane Katrina on annual GPP in 2006. The units are g C m^{-2} yr^{-1}. The white lines indicate the isotachs, including tropical storm, hurricane category 1, and hurricane category 2.