Evaluating terrestrial CO$_2$ flux diagnoses and uncertainties from a simple land surface model and its residuals

T. W. Hilton$^1$, K. J. Davis$^1$, and K. Keller$^2$

$^1$Department of Meteorology, Pennsylvania State University, University Park, PA, USA
$^2$Department of Geosciences, Pennsylvania State University, University Park, PA, USA

*now at: Department of Biology, University of New Mexico, Albuquerque, NM, USA

Received: 27 July 2013 – Accepted: 8 August 2013 – Published: 22 August 2013

Correspondence to: T. W. Hilton (hilton@unm.edu)

Published by Copernicus Publications on behalf of the European Geosciences Union.
Abstract

Global terrestrial atmosphere-ecosystem carbon dioxide fluxes are well-constrained by the concentration and isotopic composition of atmospheric carbon dioxide. In contrast, considerable uncertainty persists surrounding regional contributions to the net global flux as well as the impacts of atmospheric and biological processes that drive the net flux. These uncertainties severely limit our ability to make confident predictions of future terrestrial biological carbon fluxes. Here we use a simple light-use efficiency ecosystem model (the Vegetation Photosynthesis Respiration Model, VPRM) driven by remotely-sensed temperature, moisture, and phenology to diagnose North American gross ecosystem exchange (GEE), ecosystem respiration, and net ecosystem exchange (NEE) for the period 2001 to 2006. We optimize VPRM parameters to eddy covariance (EC) NEE observations from 65 North American FluxNet sites. We use a separate set of 27 cross-validation FluxNet sites to evaluate a range of spatial and temporal resolutions for parameter estimation. With these results we demonstrate that different spatial and temporal groupings of EC sites for parameter estimation achieve similar sum of squared residuals values through radically different spatial patterns of NEE. We also derive a regression model to estimate observed VPRM errors as a function of VPRM NEE, temperature, and precipitation. Because this estimate is based on model-observation residuals it is comprehensive of all of the error sources present in modeled fluxes. We find that 1 km interannual variability in VPRM NEE is of similar magnitude to estimated 1 km VPRM NEE errors.

1 Introduction

Terrestrial ecosystems remove roughly 25% of annual anthropogenic fossil fuel carbon dioxide (CO2) via gross primary production (GPP) in excess of respiration (Keeling et al., 1996). More completely, net ecosystem exchange (NEE) – the balance between photosynthesis and heterotrophic respiration – controls the magnitude of atmosphere
to ecosystem CO$_2$ uptake. Diagnosing terrestrial biological carbon dioxide fluxes with confidence is a necessary step toward understanding biological and climatological drivers of these fluxes. Because these fluxes are first-order influences on the accumulation of carbon dioxide in the atmosphere (Denman et al., 2007), understanding their mechanics is necessary to forecast impacts of past and future fossil fuel emissions. In spite of this, atmosphere-based methods to estimate global NEE (e.g., Peters et al., 2007; Janssens et al., 2003) and ground-based approaches (e.g., Potter et al., 2007; Janssens et al., 2003) have produced conflicting estimates, demonstrating substantial uncertainty surrounding these efforts to describe terrestrial carbon cycle mechanics.

Several recent studies demonstrate the discrepancy in diagnosed NEE between “top-down” atmosphere-based approaches and ground-based “bottom-up” approaches rooted in eddy covariance (EC) flux measurements combined with ecosystem models. Janssens et al. (2003) assembled top-down and bottom-up estimates for late 20th Century annual cumulative European NEE and found that the top down estimates are larger by roughly 100 Tg CO$_2$ yr$^{-1}$. This 100 Tg difference is between 30% and 100% of their best-estimate annual total of 100 to 300 Tg yr$^{-1}$. Peters et al. (2007) estimated a North American uptake for 2001 to 2005 of roughly 700 Tg CO$_2$ yr$^{-1}$ using a top-down approach, while Potter et al. (2007) estimated uptake of 100 to 200 Pg CO$_2$ yr$^{-1}$ over the same period using a bottom-up method.

A number of studies have explored approaches to estimate regional NEE using some combination of land surface models and eddy covariance fluxes (bottom-up methods). Potter et al. (2007) chose a set of four North American EC towers to represent characteristic North American ecosystems and used them to evaluate the performance of the NASA-CASA ecosystem model run with a global set of previously published parameter values. They then used the NASA-CASA model to estimate North American annual cumulative NEE. This approach requires no computationally intensive data assimilation (e.g. parameter estimation), but achieves such savings at the cost of considering only a small portion of the NEE observations that are now available.
Xiao et al. (2008) used a modified regression tree to create a model suite to explain observed NEE as a function of a variety of satellite-derived ecological measures. A regression tree is a method to empirically derive a best-fit statistical model based on a set of linear models. They derived the models using data from 42 AmeriFlux EC sites in the coterminous United States, producing a set of empirical models capable of upscaling the tower observations to the continental scale. The model that best explained the observed NEE used a combination of MODIS surface reflectances, enhanced vegetation index (EVI), land surface temperature, and normalized difference water index (NDWI). Though statistical, this model structure is quite similar to light-use efficiency (LUE) based models such as the Vegetation Photosynthesis Respiration Model (VPRM) of Mahadevan et al. (2008).

Beer et al. (2010) compared five different diagnostic gross primary production (GPP) models with sharply contrasting structures, including two machine learning approaches, an NEE–biome region look-up table, and a LUE model. Each of these approaches was then used to estimate regional NEE values and uncertainties. The Beer et al. study explicitly considered many sources of uncertainty in the models considered. For the light-use efficiency model, the authors estimated site-specific parameter probability density functions (PDFs) at a number of FluxNet eddy covariance sites around the globe. The Bayesian framework of the study allowed the authors to consider PDFs from multiple uncertainty sources: parameter value uncertainty as well as driver data uncertainty. By taking random draws from these parameter PDFs, Beer et al. (2010) constructed a population of global GPP estimates driven by their distribution of parameter values. This population then provided confidence intervals for their global GPP estimates.

Each study outlined above presents a framework to use ecosystem modeling to combine the information in eddy covariance flux tower observations with the information contained in an ecosystem model structure and allows estimation of regional biological CO₂ fluxes. These studies exhibit many ways to treat uncertainty sources, ranging in complexity from not including uncertainty to Bayesian consideration of multiple uncer-
The set of available eddy covariance NEE observations has increased dramatically in recent years (http://www.fluxdata.org); none of the above studies, however, take advantage of the wide spatial coverage of these observations except to perform site-specific calibration of model parameters.

Hilton et al. (2013) used North American eddy covariance NEE observations from 65 different locations from the FluxNet project (http://www.fluxdata.org) to optimize parameter values for a simple ecosystem model (VPRM, Mahadevan et al., 2008). The 65 tower locations span North America in both space and plant functional type. We conducted extensive experiments varying the temporal and spatial periods for parameter estimation to determine an optimal strategy, producing nine different spatial and temporal resolutions.

Here we use VPRM and the assimilated data from this extensive tower network to diagnose annual integrated gross ecosystem exchange (GEE), ecosystem respiration ($R$), and NEE for the coterminous United States of America, Alaska, and Canada for the period 2002 to 2006. We use eddy covariance observations from a further 27 FluxNet tower locations to quantitatively cross-validate the parameter optimizations. This rigorous cross-validation analysis is a crucial step in a model-based carbon flux upscaling; without such an exercise it is difficult to measure the spatial accuracy, or lack thereof, of estimating unobserved fluxes using a model and data assimilation. Cross validation would not be possible without the recent growth of eddy covariance observation networks: if only a handful of observation locations exist, as in the recent past, we cannot afford to withhold data from parameter estimation.

We also extend the analysis to derive empirical confidence intervals for NEE diagnoses based on observed NEE residuals and NEE drivers. Because this uncertainty is derived from eddy covariance-observation-model NEE residuals, it considers all of the uncertainty sources that are present in model-based upscaling: eddy covariance observation error, model structural error, model parameterization error, and random natural ecosystem variability. This comprehensive and quantitative uncertainty analysis is also, to our knowledge, unique.
2 Methods

2.1 Land surface model

The Vegetation Photosynthesis and Respiration Model (VPRM, Mahadevan et al., 2008) is a light-use efficiency (LUE)-based land surface model. Ecosystem respiration is treated as a linear function of surface air temperature:

\[ R = \alpha T + \beta, \]  

with slope \( \alpha \) and intercept \( \beta \); \( \beta \) determines the basal rate of respiration that occurs at near-freezing temperatures. Gross Ecosystem Exchange (GEE) is modeled as

\[ \text{GEE} = \lambda \times T_{\text{scale}} \times P_{\text{scale}} \times W_{\text{scale}} \times \text{EVI} \times \frac{1}{1 + \frac{\text{PAR}}{\text{PAR}_0}} \times \text{PAR}, \]  

with PAR denoting photosynthetically active radiation and EVI the satellite-derived enhanced vegetation index (Huete et al., 2002). \( P_{\text{scale}} \) (satellite-derived), \( W_{\text{scale}} \) (satellite-derived), and \( T_{\text{scale}} \) (literature-derived) are scaling terms that take values between 0.0 and 1.0 and attenuate GEE according to phenology, moisture conditions, and temperature, respectively. Parameter \( \lambda \) encodes light use efficiency, and parameter \( \text{PAR}_0 \) encodes the LUE curve half-saturation value. Mahadevan et al. (2008) provide detailed description of VPRM structure and performance. As described more fully in Hilton et al. (2013), the relatively simple structure of VPRM and its small number of parameters make it computationally inexpensive. This makes relatively sophisticated parameter estimation methods possible and makes VPRM a useful tool for diagnosing carbon fluxes, estimating flux uncertainty, and exploring the impacts of model parameterization and model error spatial covariance.

2.2 Land surface model parameterization

Hilton et al. (2013) presented estimated values for \( \lambda \), \( \text{PAR}_0 \), \( \alpha \), and \( \beta \) using data from 65 North American eddy covariance towers (Fig. 1, Table 1). For parameter estimation,
the eddy-covariance data were partitioned in three different ways in space (individual sites, plant functional types (PFTs), and all-sites-together), and three different ways in time (monthly, annual, and all-available-data). This produced nine unique VPRM parameter sets with differing spatial and temporal optimization “resolutions”: single sites – monthly, PFT – annual, etc. Upscaling tower measurements intrinsically requires model parameters that are applicable in spatial locations without tower observations, making the single-site parameters not useful for the task. This leaves the six parameter sets from the PFT and all-sites-together spatial groupings available for upscaling. In addition to those six, PFT–ten-day parameters were calculated as well. The sum of squared errors (SSE) were computed for the seven VPRM parameter sets that are useful for upscaling, using observations from 27 “cross-validation” eddy covariance sites (Table 2; Fig. 2; Sect. 2.3) that were not used for parameter estimation.

2.3 Data

The 2007 FluxNet Synthesis dataset (http://www.fluxdata.org) assembled eddy covariance observations from field sites around the world. The data were gap-filled and assigned quality scores using published methods (Papale et al., 2006; Moffat et al., 2007). The present study uses non-gapfilled NEE from 92 eddy covariance sites from the United States and Canada: 65 flux towers to estimate VPRM parameters (Table 1, Fig. 1), plus 27 “cross-validation” sites (Table 2, Fig. 2). The cross-validation sites were not used to estimate parameters; they are used in the present study to evaluate the performance of the optimized VPRM.

VPRM uses temperature and photosynthetically active radiation (PAR) to drive GEE and respiration. To run VPRM at the continental scale, air temperature and downward surface radiation values were obtained from the reanalysis products of Sheffield et al. (2006). The Sheffield et al. (2006) products attempt to correct known biases (Brotzge, 2004) to the NCEP-NCAR reanalysis products (Kalnay et al., 1996). VPRM was driven with the three-hourly, 1° × 1° product for temperature and PAR.
VPRM is also driven by satellite-derived moisture and phenology. MODIS products MOD13 A2 (enhanced vegetation index (EVI); Huete et al., 2002, 1999), MCD12Q1 (land cover; Friedl et al., 2002; Strahler et al., 1999), MCD12Q2 (vegetation dynamics; Zhang et al., 2003), and MCD43B4 (Bidirectional Reflectance Distribution Function (BRDF) reflectances; Schaaf et al., 2002) provided these drivers. EVI data reported with quality ratings of “lowest quality” and “not useful” (VI quality bits 2–3 equal to 11) were discarded. Gaps from discarded MODIS data were not filled; VPRM fluxes were not calculated in these instances. EVI and BRDF data are reported at one-kilometer, 16 day resolution; land cover and vegetation dynamics are reported at 500-meter, annual resolution and were processed from 500-meter resolution to 1000-meter resolution using software tools provided by the MODIS Land quality assessment group (Roy et al., 2002).

### 2.4 Ecosystem-atmosphere carbon dioxide flux calculation

VPRM gross ecosystem exchange (GEE), ecosystem respiration ($R$), and net ecosystem exchange (NEE) were calculated for the 48 coterminous United States, Alaska, and Canada at three hourly temporal resolution and one kilometer spatial resolution for 2002 to 2006. The MODIS products with 16 day temporal resolution were simply repeated at each three-hourly interval across the 16 days. Three-hourly diagnoses of GEE, $R$, and NEE were integrated to annual values and used to calculate annual anomalies (defined as the annual integrated value minus the 2002 to 2006 mean annual integrated value).

### 2.5 NEE residual spread estimation

This study seeks upscaled NEE diagnoses accompanied by uncertainty estimates. There are several methods of varying complexity available to quantify this uncertainty. A joint Bayesian inversion of VPRM parameters and VPRM NEE variance against eddy covariance NEE observations using a joint likelihood function would extract information
from the available data with maximum mathematical rigor (though is still vulnerable to
aggregation errors in grouping scheme (e.g. plant functional types), as well as errors
in observations and driver data). Statistically rigorous likelihood functions for model
NEE error, however, remain an ongoing research topic, and the calculation itself is
computationally expensive.

The method employed here uses an empirically-derived statistical model to char-
acterize VPRM NEE residual spread – a middle ground between the joint Bayesian
inversion with MCMC and the simple interpolation.

It is known that eddy covariance observation error is proportional to NEE magnitude
itself (Richardson et al., 2006). Typical magnitude for this random EC observation error
is roughly 20 to 30 g C m\(^{-2}\) yr\(^{-1}\) (Richardson et al., 2006; Goulden et al., 1996), roughly
an order of magnitude smaller than the VPRM annual NEE residuals. Random eddy
covariance observation error is a component of VPRM NEE error, making it reasonable
to posit that VPRM NEE residual magnitude is correlated to VPRM NEE magnitude.
Furthermore, the structure of VPRM (Eqs. 1 and 2) assumes that temperature, water
availability, and greenness (EVI) are primary drivers of NEE. It seems reasonable, then,
that VPRM residuals would be affected by these influences as well.

Using non-gapfilled observations from the 65 eddy covariance sites used to esti-
mate VPRM parameter values, annual integrated NEE residuals were calculated as
the integrated sum of non-gapfilled NEE observations minus VPRM NEE, both at EC
site-specific native reporting resolution (generally 30 min, 60 min at a few sites). Resid-
uals were calculated only at time stamps where both quantities were available. With
these integrated residuals squared differences were calculated for each site-year:

\[
\text{NEE}_{\text{sq diff}}' \equiv (\text{NEE}' - \overline{\text{NEE}}')^2.
\]  

(3)

\(\text{NEE}'\) denotes annually integrated VPRM residual, and \(\overline{\text{NEE}}\) denotes the mean of
\(\text{NEE}'\) across all site-years. \(\text{NEE}_{\text{sq diff}}'\) is closely related to statistical variance \(\sigma^2\)
\((\sigma^2 \equiv \sum_{i=1}^{N} (x_i - \overline{x})^2 / (N - 1))\). Estimating \(\text{NEE}_{\text{sq diff}}'\) in terms of known quantities provides

13761
a method to estimate the spread of VPRM NEE errors that can be upscaled along with the flux diagnoses.

Regression models for $\text{NEE}_{\text{sq diff}}$ were derived from subsets of these candidate explanatory variables: VPRM annual integrated NEE, total annual precipitation, annual mean surface air temperature, annual mean EVI, PFT, and year. The set of models consisting of all combinations within the categorical variables, the linear numerical terms, and the quadratic numerical terms was searched exhaustively using the glmulti package (Calcagno, 2011) for $R$ (R Development Core Team, 2007) and the results ranked by Akaike’s Information Criterion (AIC, Akaike, 1976). Annual mean EVI was calculated as the mean of 16 day MODIS EVI values (see Sect. 2.3). Annual total precipitation and annual mean temperature were calculated as the sum and mean, respectively, of the monthly mean $1^\circ \times 1^\circ$ Sheffield et al. (2006) reanalysis products.

3 Results and discussion

3.1 Land surface model parameter set ranking

As described in Sect. 2.2, we ranked the parameter sets that are useful for upscaling (this excludes the three individual-site-based parameter sets) by sum of squared errors (SSE). Figure 3 presents these SSE values, plotted against the number of unique parameter values. The solid curve plots the SSE for the 27 cross-validation sites not used for VPRM parameter estimation, and the dashed curve plots the penalized sum of squared errors (PSSE, e.g. Hilborn and Mangel, 1997) for all 92 sites (the 27 cross-validation sites and the 65 sites used to parameterize VPRM.) PSSE is given by

$$\text{PSSE} = \frac{\text{SSE}}{n_{\text{obs}} - 2n_{\text{pars}}},$$

with SSE the sum of squared errors, $n_{\text{pars}}$ the number of unique model parameter values, and $n_{\text{obs}}$ the number of data points available. Among cross-validation sites
withheld from parameter estimation, we can detect overfitting when the SSE begins to increase with the number of parameters. Because model parameterization, by definition, fits the model to observed data, SSE among parameterization sites should only decrease with additional parameters. PSSE provides a method to detect overfitting among parameterization sites. Figure 3 suggests that the monthly and 10 day VPRM parameter sets overfit the data.

The PFT–all-data VPRM parameters achieved the lowest cross-validation SSE as well as a PSSE only slightly above the lowest PSSE; therefore, those parameters are used for most of the analyses presented here. The five lowest cross-validation SSE values in Fig. 3 are not drastically different from one another, though the penalized SSE values for the two most parsimonious parameter sets (all–all and all–annual) are significantly higher. In combination with the parameter distributions presented in Hilton et al. (2013), this result might suggest that order 100 parameters are optimal for flux upscaling. The two parameter sets considered in Fig. 3 that use parameterization temporal windows shorter than annual (monthly and 10-day) produced notably higher cross-validation SSE values and higher penalized SSE values than the other five parameter sets, suggesting these parameterizations overfit the observations.

### 3.2 VPRM NEE residual evaluation

To evaluate the quality of VPRM NEE diagnoses, Fig. 4 presents the histogram of VPRM NEE residuals calculated at eddy covariance site reporting intervals (30 min at most sites; 60 min at a few sites), calculated using PFT–all-data VPRM parameters. The mean residual of $4.66 \times 10^{-4}$ μmol CO$_2$ m$^{-2}$ s$^{-1}$ corresponds to an annually integrated flux of 0.18 g C m$^{-2}$ yr$^{-1}$, small compared to a typical observed EC annual NEE between 100 and 300 g C m$^{-2}$ yr$^{-1}$. This suggests that the parameter optimization achieved its task of optimizing VPRM to observed fluxes at hourly time scales. A normal distribution with the same mean and standard deviation is overlaid; the observed residuals show a higher peak around their mean but otherwise correspond closely to the normal distribution.
Having demonstrated that the parameter optimization performs well at hourly intervals we turn now to the annual time scale. Figure 5 examines the distribution at the annual timescale, showing the VPRM NEE residuals integrated by site-year. Because NEE gapfilling would introduce a new source of error to VPRM residuals, the annual integrated residuals in Fig. 5 are calculated from non-gapfilled NEE observations. The mean integrated residual value of 1.6 g C m\(^{-2}\) yr\(^{-1}\) demonstrates that VPRM optimization also performed well at the annual scale. This observed residual distribution also follows a normal distribution (overlaid) reasonably well.

Satisfied now that VPRM residuals are small relative to EC-observed fluxes, we investigate whether 1 km VPRM diagnosed annual NEE for North America seems reasonable when compared to EC-observed annual NEE. Figure 6 compares the distribution of annually integrated VPRM NEE diagnoses for the modeling domain (Sect. 2.4) with the distribution of annually integrated NEE observations from the 2007 FluxNet synthesis dataset for the sites in Table 1. To obtain a meaningful comparison to model diagnoses we use the FluxNet synthesis dataset gapfilled NEE observations in this case. The gapfilling used the methods of Papale et al. (2006) and Moffat et al. (2007). VPRM reproduces well the mode of the observed distribution as well as the right-hand tail (sources of CO\(_2\) to the atmosphere). The left tail of the observed NEE distribution contains more density than the VPRM diagnoses, suggesting that VPRM estimates lower sinks of atmospheric CO\(_2\) than the gapfilled FluxNet 2007 Synthesis dataset in some cases. The FluxNet synthesis dataset contains a handful of site-years with sinks of atmospheric CO\(_2\) approaching or even exceeding 1000 g C m\(^{-2}\) yr\(^{-1}\). VPRM was optimized to these data, and the left-side tail of the VPRM diagnosed NEE distributions does contain more mass than the right side.

In all, the VPRM performance summarized by Figs. 4–6 are encouraging for the ability of the parameter estimation process to optimize VPRM to eddy covariance observations at both hourly and annual timescales.
3.3 VPRM fluxes

Figures 7–9 show annually integrated VPRM GEE, \( R \), and NEE, respectively, for 2002. The larger-scale (order > 100 km) spatial patterns are representative of the integrated fluxes for 2003 to 2006 (not shown). NEE is the difference between GEE and \( R \), both much larger in magnitude. This raises detectability issues for NEE: this difference between two larger and roughly equal quantities is easily polluted by errors from GEE and \( R \) estimation. Therefore, rather than focus on integrated annual NEE values or aggregated continental NEE, this study instead focuses on year-to-year NEE differences and NEE differences across different VPRM parameter sets.

The broad spatial patterns in these results largely agree with other analyses (e.g., Beer et al., 2010; Xiao et al., 2011; Running et al., 2004). As we would expect given the prominence of the vegetation index in VPRM structure (Eq. 2), the patterns of strong GPP reflect areas of relatively dense vegetation as measured by vegetation index (Huete et al., 2002) or biomass (Myneni et al., 2001).

The relatively large respiration diagnoses for the southeastern USA in Fig. 8 is also present in the 2003 to 2006 diagnoses. This area, roughly covering the US states of Louisiana, Mississippi, Alabama, Georgia, and South Carolina, is dominated by the mixed forest PFT in the MODIS land cover classification (Sect. 2.3). The three mixed forest eddy covariance sites used for VPRM parameterization are in Wisconsin, USA and Ontario, Canada. Rather than conclude that the mixed forests of the US Gulf Coast are much stronger sources of biological \( \text{CO}_2 \) than other classes of southern forests or more northerly mixed-forests, several alternative explanations seem more likely. First, perhaps the carbon cycle mechanics of northern mixed forests do not describe well the behavior of southerly mixed forests and diagnose erroneously strong respiration when applied in southerly regions. Second, three eddy covariance sites may provide insufficient data to characterize this (or any) PFT. Last, stand age is an important driver of NEE (Litvak et al., 2003), and is ignored by the modeling methods employed here.
Year to year flux differences are reported here as annual anomalies, calculated as (integrated annual flux) minus (mean integrated annual flux, 2002 to 2006). VPRM integrated annual flux anomalies for 2002 to 2006 are shown in Figs. 10 (GEE), 11 \((R)\), and 12 (NEE). Notable in these diagnoses is the much larger variability of GEE as compared to \(R\). That GEE variability is reflected in NEE variability as well. This could be a consequence of VPRM’s structural treatment of respiration as a linear function of temperature (Eq. 1). In contrast VPRM GEE (Eq. 2) considers a number of other variables in addition to temperature. Increased interannual variability (IAV) in GEE may simply reflect that there are more constituent quantities to vary.

Much of the stronger GEE IAV (Fig. 10) occurs in the upper Midwestern USA. 2006, for example, showed a particularly strong VPRM GEE diagnosis centered around the US state of Indiana. This area is dominated by agriculture – the cropland PFT in the MODIS IGBP landcover classification. Within the cropland PFT different agricultural products are known to vary in the strength of their carbon uptake. Corn, for example, has particularly strong atmospheric CO\(_2\) uptake (Lokupitiya et al., 2009). Without parameterizations specific to particular crops, model NEE diagnosis can be poor (Lokupitiya et al., 2009). There are only five agricultural EC sites in the group used to parameterize VPRM (Table 1). This makes it possible that the model parameterization suffers from the same representativeness problem that may cause potentially spurious VPRM respiration spatial structure in the southeastern USA. Many farms rotate crops from one season to the next; for example, corn in year \(y\) followed by soybeans in year \(y+1\). If reflected in remotely sensed ecosystem variables (e.g. vegetation indices or moisture) this sort of rotation could itself cause the GEE interannual variation seen in the VPRM annual anomalies. Similarly, if the cropland VPRM parameter estimation EC sites (US-Ne1, US-Ne2, US-Ne3, US-Bo1, and US-Bo2) were consistently planted with a particular crop during the periods used for parameter estimation, VPRM should not be expected to perform well for different crops. Looking to other potential causes for large year to year changes in the upper Midwestern USA, VPRM \(R\) diagnoses in that region show little year to year variation, removing temperature anomalies as a driver.
of GEE IAV. From the structure of VPRM GEE (Eq. 2) this leaves moisture availability, PAR, and vegetation index as primary candidates for driving GEE variability.

3.4 Estimated spread of VPRM fluxes

Determining whether these flux diagnoses are able to detect meaningful interannual variability (IAV) requires a measure of the variance of annual integrated NEE. Section 2.5 describes the empirical derivation of a statistical model to predict the squared difference between the annual integrated VPRM residual and its mean (NEE′_sq diff, Eq. 3) across site years. Of the candidate models, the best-fitting model (lowest AIC) was

\[
\text{NEE}'_{\text{sq diff}} = 2.66 \times 10^{-1} \text{NEE}^2_{\text{VPRM}} + 
\]

\[
5.72 \text{NEE}_{\text{VPRM}} + 9.86 \times 10^2 \ T + 
\]

\[
3.95 \times 10^{-2} \ pcp^2 + 2.05 \times 10^3.
\]

NEE_{VPRM} is annual integrated VPRM NEE, \( T \) is annual mean temperature (°C), and \( pcp \) is annual total precipitation (mm). The fit achieved a multiple \( R \)-squared of 0.289, with the coefficients significant at \( p < 0.001 \) (NEE^2_{VPRM}), \( p < 0.05 \) (pcp^2), \( p < 0.1 \) (\( T \)), and no significance for NEE_{VPRM} (\( p = 0.12 \)).

The regression model in Eq. (5) was tested at the cross-validation EC sites (Table 2, Fig. 2). Figure 13 (top panel) shows observed vs. predicted NEE′_sq diff with the 95% prediction interval. The prediction intervals at each point are calculated from the regression slope and intercept variances, which are estimated from the residuals of the regression fit. Of 56 site yr in the cross-validation data set, one observation is outside of the 95% prediction interval. The bottom panel of Fig. 13 shows histograms of the observed and predicted values. The distributions are similar, except for predicted values around zero. This highlights a shortcoming of the regression model approach: negative predicted NEE′_sq diff values are possible. This should emphasize that, as with
any regression model, predictions are only valid when the explanatory variables take values within the ranges used to fit the model.

The regression model performs well, then, at the cross-validation eddy covariance sites. Figure 14 shows the square root of estimated $\text{NEE}_{\text{sq diff}}'$ for the modeling area for 2002. The spatial patterns for 2003 to 2006 (not shown) are similar. The estimated VPRM errors are broadly of similar magnitude to the VPRM NEE differences between years (Fig. 12) and the VPRM NEE differences between VPRM parameter sets (Fig. 15).

Though the regression model estimation methods developed here are applied to estimate VPRM NEE error magnitude, the approach is equally applicable to estimating errors in an ecosystem model diagnosis of GEE or $R$; this change would be subject only to quality of the partitioning of EC NEE observations into GEE and $R$.

As noted in Sect. 1, several recent studies have attempted continent-scale carbon flux diagnoses; those diagnoses generally do not report uncertainty. Beer et al. (2010) reported spatial estimates of GPP accompanied by globally aggregated uncertainties. The work presented here reports spatial GPP, $R$, and NEE diagnoses, and further extends the literature by estimating annual continental NEE uncertainty in space.

Beer et al. (2010) estimate GPP uncertainty for their LUE model by randomly resampling from within their population of parameters optimized to eddy covariance observations at each observation site. Because the parameters are optimized to flux observations, these uncertainty estimates include observation errors and model parameterization errors. Driver data uncertainty is quantified by analyzing uncertainty separately for three different reanalysis products.

Because they are based on model-observation differences, the estimates presented here include eddy covariance observation error, model parameterization error, model structural error, and driver data error. This makes these estimates inclusive of a broader range of error sources relative to approaches that focus on propagating specific errors through a model calculation. This inclusiveness sacrifices the possibility of partitioning the estimated error into contributions from constituent sources.
As an empirical approach based upon observed model residuals the regression model approach to uncertainty estimation studies flux diagnosis uncertainty from a foundation independent of the method of Beer et al. (2010). Direct comparison is difficult because the uncertainty estimates presented here quantify NEE errors for North America, while Beer et al. estimate globally aggregated GPP uncertainty. It is simple in concept, however, to extend the methods shown here to GPP uncertainty and to the global scale.

3.5 NEE error covariance nugget

The estimated nugget values from the VPRM NEE error spatial covariance (Hilton et al., 2013) quantify combined eddy covariance observation error and “microscale variation”, that is, the behavior of the difference in VPRM NEE error between two locations that are closer to one another than the closest pairs of towers among the 65 used for covariance parameter estimation. The median estimated seasonal nugget values range from $5.42 \times 10^{-5}$ (individual-site–monthly VPRM parameters) through 0.775 (PFT–all-data parameters) to 0.884 (all-sites–all-data parameters), with units of flux squared: $(\mu $mol CO$_2$ m$^{-2}$ s$^{-1})^2$. Converted to standard deviation and integrated annually (g C m$^{-2}$ yr$^{-1}$) these nuggets are 21.0, 586, and 603.

In units of standard deviation the annual integrated NEE error nugget of 21.0 g C m$^{-2}$ yr$^{-1}$ from the individual-site–monthly VPRM parameters is essentially equal to the annual total eddy covariance random observation error of $\pm$20 g C m$^{-2}$ yr$^{-1}$ estimated by Richardson and Hollinger (2005). The 65 eddy covariance sites used for fitting include 26 pairs that are within 10 kilometers of each other, so there are many data points at small separation distances to quantify the nugget.

At coarser spatial and temporal parameter estimation resolutions (PFT–all-data, all-sites–all-data, etc.) the NEE error standard deviations of roughly 600 g C m$^{-2}$ yr$^{-1}$ calculated from the nuggets are of similar magnitude to the errors estimated from VPRM NEE and climate drivers (Fig. 13) for high-productivity PFTs (e.g. forests, croplands).
These results suggest that when VPRM is optimized to NEE observations at short temporal scales (order one month) the VPRM NEE nugget is dominated by eddy covariance observation error – that is, under these conditions VPRM performs quite well in close proximity (order one kilometer) to an optimization location. Eddy covariance observation error is independent of VPRM optimization spatial and temporal windows, so its contribution to either nugget should remain constant across these windows. The much larger nugget when the temporal optimization window is all available observations therefore suggests that microscale VPRM NEE error increases dramatically from its value when VPRM is optimized monthly for individual sites. This means that VPRM can perform quite poorly even in close proximity to an optimization location in these cases.

Light use efficiency models such as VPRM make climate-driven diagnoses of NEE (Eqs. 1 and 2). Widespread VPRM annual NEE error magnitudes (Fig. 14, this section) on the order of VPRM NEE interannual variability (Fig. 12) imply that climate (or, at least, climate viewed through the prism of VPRM) cannot reliably explain NEE interannual variability.

### 3.6 VPRM parameterization and NEE spatial behavior

The PFT–all-data and all-sites–all-data VPRM parameter sets produce starkly different spatial patterns of growing season NEE in the southeastern USA (Fig. 15). This result highlights a question: what are the most appropriate parameterization time and space windows for a land surface model? In contrast, other results of this work might suggest de-emphasizing this line of inquiry. For example, the total 30 min cross-validation SSE values (Fig. 3) across the five most optimally fitting VPRM parameter sets are nearly equal, suggesting that the choice of parameter optimization spatial and temporal windows is perhaps of secondary importance. In that case, the drastically lesser computational cost makes coarser spatial and temporal windows preferable.

The region of positive NEE in Fig. 15, bottom panel corresponds to the region of large diagnosed respiration discussed in Sect. 3.3. Once again, instead of concluding
that respiration is causing the mixed forests of the southeastern USA to release on the order of $150 \text{gCm}^{-2}\text{yr}^{-1}$ to the atmosphere, the explanations discussed in Sect. 3.3 seem more plausible.

Considered in conjunction with the differing spatial behaviors in Fig. 15, the similar PSSE values among the better-performing parameter sets in Fig. 3 suggest an instance of equifinality: PFT–all-data parameters and all-sites–all-data parameters produce comparable sums of 30 min squared residuals via strongly divergent spatial outcomes.

### 3.7 Caveats

The results reported here were compiled using VPRM, a simple LUE-based ecosystem model, and are therefore most directly informative toward similar models. Questions of spatial and temporal resolution for model parameterization arise for more complicated mechanistic ecosystem models as well. Whether optimizing more complex model structures would result in similar total PSSE values for strongly contrasting spatial and temporal optimization windows (as reported here in Fig. 3) is a question for further analysis. Regardless, this work suggests that applying model parameterizations outside of the climate and ecosystem conditions where the parameter values were optimized can produce suspicious spatial structures such as the widespread flux of CO$_2$ to the atmosphere across the southeastern USA in Fig. 15 (bottom panel) and Fig. 8.

The 27 cross-validation sites (Fig. 2, Table 2) generally have shorter observational records than the sites used for VPRM parameterization. Repeating the cross-validation experiment with different, perhaps randomly selected subsets might be a useful exercise.

In addition, forest stand age since disturbance is a first order determinant of NEE magnitude (Litvak et al., 2003). Structurally, VPRM does not consider stand age (Eqs. 2 and 1), and the work presented here does not attempt to assess disturbance history. For this reason, this work does not emphasize integrated NEE magnitudes or attempt regional NEE aggregation.
Likewise, the NEE residual magnitude statistical model derived here (Sect. 2.5) was fit using observed NEE residuals and observed climatic drivers. While estimating uncertainty directly from observed residuals is a strength of the approach, as with any regression these results cannot produce meaningful estimates where the driver variables depart the range of values used for fitting.

4 Conclusions

This work presents high-resolution diagnoses of North American NEE and NEE interannual variability accompanied by NEE error estimates, all derived from a simple light use efficiency-based ecosystem model (VPRM). Several different model optimization spatial and temporal resolutions achieve similar fits when evaluated by total sum of squared errors at cross-validation sites and penalized sum of squared errors at the model parameterization sites. Cross validation is useful for identifying parameterizations that overfit assimilated data, however. Cross-validation SSE eliminated two of seven parameterizations we considered for upscaling. Penalized SSE from the parameterization sites eliminated another three. This sort of method to evaluate model parameter sets is computationally inexpensive and would make a welcome addition to future flux diagnoses.

Two of our model parameterizations achieved similar cross-validation SSE, but reached their NEE diagnoses through starkly contrasting spatial distributions of NEE. Modeling efforts that do not consider multiple spatial and temporal parameterization resolutions risk missing the structural uncertainty that this equifinality reveals, and that radically different fluxes across space may not be readily distinguishable when viewed through the lens of aggregated model errors.

The results here demonstrate that modeled annual integrated flux magnitude, annual mean surface temperature, and annual total precipitation provide reasonable (and computationally inexpensive) empirical predictors of NEE error magnitude. Estimated NEE errors are of equal magnitude to diagnosed NEE interannual variability. That a climate-
driven ecosystem model cannot reliably separate year-to-year differences in model NEE from model error suggests that NEE interannual variability has important drivers outside of large-scale climate, or, alternatively, that the present network of North American eddy covariance NEE observation sites provide insufficient constraints on NEE and NEE error to reveal a strong climate–NEE interannual variability connection.

Acknowledgements. Data were processed using SciPy (Jones et al., 2001) and Matplotlib (Hunter, 2007) as well as the $R$ language and platform for statistical computing (R Development Core Team, 2007), using the gstat (Pebesma, 2004), geoR (Ribeiro Jr., and Diggle, 2001), and DEoptim (Ardia and Mullen, 2009) packages. Funding for this research was provided by the NOAA Office of Global Programs and the US Department of Energy Terrestrial Carbon Processes Program. We wish to thank the many agencies that provided support for eddy covariance tower construction and maintenance. The Metolius AmeriFlux research was supported by the Office of Science (BER), US Department of Energy, Grant No. DE-FG02-06ER64318. The Metolius old-aged ponderosa pine study was supported by NASA (grant #NAG5-7531), and the Office of Science (BER), US Department of Energy (grant #FG0300ER63014). Data collection for the US-ARM site was supported by the Office of Biological and Environmental Research of the US Department of Energy under contract DE-AC02-05CH11231 as part of the Atmospheric Radiation Measurement Program. Research at the Morgan Monroe State Forest site was supported by the Office of Science (BER), US Department of Energy, Grant No. DE-FG02-07ER64371.

This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (US Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program (DE-FG02-04ER63917 and DE-FG02-04ER63911)), FluxNet-Canada (supported by CFCAS, NSERC, BIOCAP, Environment Canada, and NRCan), GreenGrass, LBA, NECC, USCCC. We acknowledge the financial support to the eddy covariance data harmonization provided by CarboEuropeIP, FAO-GTOS-TCO, iLEAPS, Max Planck Institute for Biogeochemistry, National Science Foundation, University of Tuscia, Université Laval and Environment Canada and US Department of Energy and the database development and technical support from Berkeley Water Center, Lawrence Berkeley National Laboratory, Microsoft Research eScience, Oak Ridge National Laboratory, University of California – Berkeley, University of Virginia.
References


Brotzge, J.: A two-year comparison of the surface water and energy budgets between two OASIS sites and NCEP-NCAR reanalysis data, J. Hydrometeorol., 5, 311–326, 2004. 13759


Fischer, M., Billesbach, D., Berry, J., Riley, W., and Torn, M.: Spatiotemporal variations in growing season exchanges of CO$_2$, H$_2$O, and sensible heat in agricultural fields of the Southern Great Plains, Earth Interact., 11, 1–21, 2007. 13785


13776


13777
Lee, J. T., Richardson, A. D., Rodrigues, C., Scott, N., Achuatavarier, D., and Walsh, J.: Spatial and


13778


13782


Table 1. 65 North American eddy covariance sites used to parameterize VPRM and calculate VPRM flux errors. PFTs are taken from the International Geosphere-Biosphere Programme (IGBP) land cover classification scheme (Loveland and Belward, 1997). The PFT classifications are taken from literature citations or investigator descriptions where available, and otherwise derived from MODIS 1 km land surface classifications. Data are from the 2007 FluxNet Synthesis Dataset.

<table>
<thead>
<tr>
<th>Site Code</th>
<th>Site Name</th>
<th>Latitude (°N)</th>
<th>Longitude (°E)</th>
<th>Land Cover</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA-Ca1</td>
<td>British Columbia – Campbell River – Mature Forest Site</td>
<td>49.870</td>
<td>−125.340</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Humphreys et al. (2006)</td>
</tr>
<tr>
<td>CA-Ca2</td>
<td>British Columbia – Campbell River – Clearcut Site</td>
<td>49.870</td>
<td>−125.290</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Humphreys et al. (2006)</td>
</tr>
<tr>
<td>CA-Ca3</td>
<td>British Columbia – Campbell River – Young Plantation Site</td>
<td>49.520</td>
<td>−124.900</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Humphreys et al. (2006)</td>
</tr>
<tr>
<td>CA-Let</td>
<td>Lethbridge</td>
<td>49.710</td>
<td>−112.940</td>
<td>10 – Grasslands</td>
<td>Flanagan and Adkinson (2011a)</td>
</tr>
<tr>
<td>CA-Mer</td>
<td>Eastern Peatland – Mer Bleue</td>
<td>45.410</td>
<td>−75.520</td>
<td>11 – Permanent Wetlands</td>
<td>Lafleur et al. (2003)</td>
</tr>
<tr>
<td>CA-NS2</td>
<td>UCI-1930 burn site</td>
<td>55.910</td>
<td>−98.520</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Goulden et al. (2006)</td>
</tr>
<tr>
<td>CA-NS3</td>
<td>UCI-1964 burn site</td>
<td>55.910</td>
<td>−98.380</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Goulden et al. (2006)</td>
</tr>
<tr>
<td>CA-NS4</td>
<td>UCI-1964 burn site wet</td>
<td>55.910</td>
<td>−98.380</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Goulden et al. (2006)</td>
</tr>
<tr>
<td>CA-NS5</td>
<td>UCI-1981 burn site</td>
<td>55.860</td>
<td>−98.490</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Goulden et al. (2006)</td>
</tr>
<tr>
<td>CA-NS6</td>
<td>UCI-1989 burn site</td>
<td>55.920</td>
<td>−98.960</td>
<td>1 – Deciduous Broadleaf Forest</td>
<td>Goulden et al. (2006)</td>
</tr>
<tr>
<td>CA-Oas</td>
<td>Sask – SSA Old Aspen</td>
<td>53.630</td>
<td>−106.200</td>
<td>4 – Deciduous Broadleaf Forest</td>
<td>Black et al. (2000)</td>
</tr>
<tr>
<td>CA-Obs</td>
<td>Sask – SSA Old Black Spruce</td>
<td>53.990</td>
<td>−105.120</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Bergeron et al. (2007)</td>
</tr>
<tr>
<td>CA-Qcu</td>
<td>Quebec Boreal Cutover Site</td>
<td>49.270</td>
<td>−74.040</td>
<td>7 – Open Shrublands</td>
<td>Giasson et al. (2006)</td>
</tr>
<tr>
<td>CA-Qto</td>
<td>Quebec Mature Boreal Forest Site</td>
<td>49.690</td>
<td>−74.340</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Bergeron et al. (2007)</td>
</tr>
<tr>
<td>CA-SF2</td>
<td>Sask – Fire 1989</td>
<td>54.250</td>
<td>−105.880</td>
<td>6 – Closed Shrublands</td>
<td>Mkhabela et al. (2009)</td>
</tr>
<tr>
<td>CA-SF3</td>
<td>Sask – Fire 1998</td>
<td>54.090</td>
<td>−106.010</td>
<td>6 – Closed Shrublands</td>
<td>Mkhabela et al. (2009)</td>
</tr>
<tr>
<td>CA-WP1</td>
<td>Western Peatland – LaBiche-Black Spruce/Larch Fen</td>
<td>54.960</td>
<td>−112.460</td>
<td>11 – Permanent Wetlands</td>
<td>Flanagan and Syed (2011b)</td>
</tr>
<tr>
<td>US-Blo</td>
<td>Blodgett Forest – California</td>
<td>38.900</td>
<td>−120.630</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Goldstein et al. (2000)</td>
</tr>
<tr>
<td>US-Br1</td>
<td>Delta Junction 1920 control site</td>
<td>63.920</td>
<td>−145.370</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Liu et al. (2005)</td>
</tr>
<tr>
<td>US-Goo</td>
<td>Goodwin Creek – Mississippi</td>
<td>34.250</td>
<td>−89.970</td>
<td>10 – Grasslands</td>
<td>Wilson and Meyers (2007)</td>
</tr>
<tr>
<td>US-Ha1</td>
<td>Harvard Forest EMS Tower – Massachusetts (HFR1)</td>
<td>42.540</td>
<td>−72.170</td>
<td>4 – Deciduous Broadleaf Forest</td>
<td>Urbanski et al. (2007)</td>
</tr>
<tr>
<td>Site Code</td>
<td>Site Name</td>
<td>Latitude (°N)</td>
<td>Longitude (°E)</td>
<td>Land Cover</td>
<td>Reference</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>---------------</td>
<td>----------------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>US-Ho1</td>
<td>Howland Forest (main tower) – Maine</td>
<td>45.200</td>
<td>-68.740</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Hollinger et al. (1999)</td>
</tr>
<tr>
<td>US-Ho2</td>
<td>Howland Forest (west tower) – Maine</td>
<td>45.210</td>
<td>-68.750</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Hollinger et al. (2004)</td>
</tr>
<tr>
<td>US-Los</td>
<td>Lost Creek – Wisconsin</td>
<td>46.080</td>
<td>-89.980</td>
<td>6 – Closed Shrublands</td>
<td>Sulman et al. (2009)</td>
</tr>
<tr>
<td>US-MOz</td>
<td>Missouri Ozark Site</td>
<td>38.740</td>
<td>-92.200</td>
<td>4 – Deciduous Broadleaf Forest</td>
<td>Gu et al. (2006)</td>
</tr>
<tr>
<td>US-Ne1</td>
<td>Mead – irrigated continuous maize site – Nebraska</td>
<td>41.100</td>
<td>-96.290</td>
<td>12 – Croplands</td>
<td>Verma et al. (2005)</td>
</tr>
<tr>
<td>US-Ne2</td>
<td>Mead – irrigated maize-soybean rotation site – Nebraska</td>
<td>41.100</td>
<td>-96.280</td>
<td>12 – Croplands</td>
<td>Verma et al. (2005)</td>
</tr>
<tr>
<td>US-Ne3</td>
<td>Mead – rainfed maize-soybean rotation site – Nebraska</td>
<td>41.180</td>
<td>-96.440</td>
<td>12 – Croplands</td>
<td>Verma et al. (2005)</td>
</tr>
<tr>
<td>US-PFa</td>
<td>Park Falls/WLEF- Wisconsin</td>
<td>45.950</td>
<td>-90.270</td>
<td>5 – Mixed Forest</td>
<td>Davis et al. (2003)</td>
</tr>
<tr>
<td>US-Ton</td>
<td>Tonzi Ranch – California</td>
<td>38.430</td>
<td>-120.970</td>
<td>8 – Woody Savannas</td>
<td>Ma et al. (2007)</td>
</tr>
<tr>
<td>US-Var</td>
<td>Vaira Ranch – Ione – California</td>
<td>38.410</td>
<td>-120.950</td>
<td>10 – Grasslands</td>
<td>Ma et al. (2007)</td>
</tr>
</tbody>
</table>
Table 2. 27 North American eddy covariance cross-validation sites.

<table>
<thead>
<tr>
<th>Site Code</th>
<th>Site Name</th>
<th>Latitude (' N)</th>
<th>Longitude (' E)</th>
<th>Land Cover</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA-SF1</td>
<td>Sask – Fire 1977</td>
<td>54.490</td>
<td>-105.820</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Amiro et al. (2006)</td>
</tr>
<tr>
<td>CA-TP2</td>
<td>Ontario – Turkey Point Young White Pine</td>
<td>42.770</td>
<td>-80.460</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Peichl and Arain (2006)</td>
</tr>
<tr>
<td>US-NC1</td>
<td>NC Clearcut</td>
<td>35.810</td>
<td>-76.710</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Sun et al. (2010)</td>
</tr>
<tr>
<td>US-NC2</td>
<td>NC Lobolly Plantation</td>
<td>35.800</td>
<td>-76.670</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Sun et al. (2010)</td>
</tr>
<tr>
<td>CA-TP3</td>
<td>Ontario – Turkey Point Middle-aged White Pine</td>
<td>42.710</td>
<td>-80.360</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Peichl and Arain (2006)</td>
</tr>
<tr>
<td>CA-TP4</td>
<td>Ontario – Turkey Point Mature White Pine</td>
<td>42.710</td>
<td>-80.360</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Peichl and Arain (2006)</td>
</tr>
<tr>
<td>CA-TP1</td>
<td>Ontario – Turkey Point Seedling White Pine</td>
<td>42.660</td>
<td>-80.560</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Peichl and Arain (2006)</td>
</tr>
<tr>
<td>CA-Man</td>
<td>BOREAS NSA – Old Black Spruce</td>
<td>55.880</td>
<td>-98.480</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Dunn et al. (2007)</td>
</tr>
<tr>
<td>US-LPH</td>
<td>Little Prospect Hill – Massachusetts</td>
<td>42.540</td>
<td>-72.180</td>
<td>5 – Mixed Forest</td>
<td>Borken et al. (2006)</td>
</tr>
<tr>
<td>US-W5</td>
<td>Mixed young jack pine (MYJP)</td>
<td>46.650</td>
<td>-91.090</td>
<td>1 – Evergreen Needleleaf Forest</td>
<td>Noormets et al. (2007)</td>
</tr>
<tr>
<td>US-W6</td>
<td>Pine barrens #1 (PB1)</td>
<td>46.620</td>
<td>-91.300</td>
<td>7 – Open Shrublands</td>
<td>Noormets et al. (2007)</td>
</tr>
<tr>
<td>US-W8</td>
<td>Young hardwood clearcut (YHW)</td>
<td>46.720</td>
<td>-91.250</td>
<td>4 – Deciduous Broadleaf Forest</td>
<td>Noormets et al. (2007)</td>
</tr>
<tr>
<td>CA-WP2</td>
<td>Poor Fen</td>
<td>55.540</td>
<td>-112.330</td>
<td>11 – Permanent Wetlands</td>
<td>Adkinson et al. (2011)</td>
</tr>
</tbody>
</table>
Fig. 1. The 65 eddy covariance flux tower sites from the FluxNet network (http://www.fluxdata.org) that provide observations for VPRM parameterization and VPRM flux residual calculation. ENF: evergreen needleleaf forest, DBF: deciduous broadleaf forest, MF: mixed forest, CS: closed shrubland, OS: open shrubland, WS: woody savanna, Gr: grassland, Wet: permanent wetland, Crop: cropland.
Fig. 2. 27 eddy covariance flux tower “cross-validation” sites from the 2007 FluxNet synthesis dataset that were not used for VPRM parameterization or for VPRM NEE error covariance parameter estimation. Plant functional type abbreviations: ENF: evergreen needleleaf forest, DBF: deciduous broadleaf forest, MF: mixed forest, CS: closed shrubland, OS: open shrubland, WS: woody savanna, Gr: grassland, Wet: permanent wetland, Crop: cropland.
Fig. 3. VPRM sum of squared errors vs. number of unique parameter values. Shown are the seven VPRM parameter sets available for upscaling. Parameter sets are labeled as [space grouping]-[time grouping] used for VPRM parameter estimation. Left vertical axis shows sum of squared errors (SSE) for the 27 cross-validation sites not used to estimate VPRM parameters (Fig. 2, Table 1); right vertical axis shows penalized sum of squared errors (PSSE) for the 27 cross-validation sites combined with the 65 sites used to parameterize VPRM (Fig. 1, Table 1). Note the log scale on the horizontal axis.
Fig. 4. Histogram of VPRM NEE residuals at eddy covariance site reporting temporal resolution (30 min or 60 min), PFT–all-data VPRM parameters. NEE residuals are calculated as observed NEE (non-gapfilled) minus VPRM NEE. A normal distribution probability density function with the same mean and standard deviation is overlaid.
Fig. 5. Histogram of VPRM annual integrated NEE residuals, PFT–all-data VPRM parameters. These are the residuals in Fig. 4 integrated by site-year. The residuals are calculated as 
\[(\text{annual integrated observed NEE, non-gapfilled}) \text{ minus (VPRM annual integrated NEE)}\].
Fig. 6. Histograms of annually integrated NEE. Observations are the 2007 FluxNet synthesis gapfilled annual NEE.
Fig. 7. 2002 annual integrated VPRM GEE, g C m$^{-2}$ yr$^{-1}$. PFT–all-data VPRM parameters.
Fig. 8. 2002 annual integrated VPRM respiration, g C m\(^{-2}\) yr\(^{-1}\). PFT–all-data VPRM parameters.
Fig. 9. 2002 annual integrated VPRM NEE, g C m$^{-2}$ yr$^{-1}$. PFT–all-data VPRM parameters.
Fig. 10. Annual anomaly, annual integrated VPRM GEE, calculated using PFT–all-data VPRM parameters. Units are gC m\(^{-2}\) yr\(^{-1}\). Anomalies are calculated as annual integrated VPRM GEE minus the 2002 to 2006 mean annual integrated VPRM GEE. Thus negative values denote lesser than average atmosphere to ecosystem CO\(_2\) flux.
Fig. 11. Annual anomaly, annual integrated VPRM respiration, calculated using PFT–all-data VPRM parameters. Units are gC m\(^{-2}\) yr\(^{-1}\). Anomalies are calculated as annual integrated VPRM \(R\) minus the 2002 to 2006 mean annual integrated VPRM \(R\). Thus negative values denote greater than average atmosphere to ecosystem CO\(_2\) flux.
Fig. 12. Annual anomaly, annual integrated VPRM NEE, calculated using PFT–all-data VPRM parameters. Units are g C m\(^{-2}\) yr\(^{-1}\). Anomalies are calculated as annual integrated VPRM NEE minus the 2002 to 2006 mean annual integrated VPRM NEE. Thus negative values denote greater than average atmosphere to ecosystem CO\(_2\) flux.
Fig. 13. Results from an empirical regression model (Eq. 5) for VPRM NEE residual spread. Units are (gCm$^{-2}$ yr$^{-1}$)$^2$. Top panel shows observed values vs. predicted values at 27 cross-validation sites (Table 2). Observed values are outside of the 95% prediction interval where the solid line falls outside of the dashed lines. One of 56 predicted values (2%) is outside the 95% prediction interval. The bottom panel shows histograms for observed values and model-predicted values.
Fig. 14. 2002 estimated square root of NEE residual squared difference (Eq. 3), calculated using PFT–all-data VPRM parameters. Units are gCm$^{-2}$yr$^{-1}$. Estimates are calculated by a statistical model (Eq. (5), Sect. 2.5) with explanatory variables annual integrated VPRM NEE, annual total precipitation, and annual mean surface temperature. 2003 to 2006 estimated annual errors (not shown) show similar spatial patterns.
Fig. 15. 2004 VPRM June-July-August integrated NEE. Top panel shows all-sites–all-data parameters, bottom panel shows PFT–all-data parameters. Units are g C m$^{-2}$ yr$^{-1}$. 