Biogeosciences Discuss., 9, 7073–7116, 2012
www.biogeosciences-discuss.net/9/7073/2012/
doi:10.5194/bgd-9-7073-2012

Biogeosciences
Discussions

Improving terrestrial CO_2 flux diagnosis using spatial structure in land surface model residuals

T. W. Hilton[1,*], **K. J. Davis**[1], **K. Keller**[2], **and N. M. Urban**[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
[**] now at: Energy Security Center, Los Alamos National Laboratory, Los Alamos, NM, USA

Received: 31 January 2012 – Accepted: 17 May 2012 – Published: 15 June 2012

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

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

Abstract

We evaluate spatial structure in North American CO_2 flux observations using a simple diagnostic land surface model. The Vegetation Photosynthesis Respiration Model (VPRM) calculates net ecosystem exchange (NEE) using locally observed temperature and photosynthetically active radiation (PAR) along with satellite-derived phenology and moisture. We use observed NEE from a group of 65 North American eddy covariance tower sites spanning North America to estimate VPRM parameters for these sites. We investigate spatial coherence in regional CO_2 fluxes at several different time scales by using geostatistical methods to examine the spatial structure of model data–model residuals. We find that persistent spatial structure does exist in the data-model residuals at a length scale of approximately 1000 km. This spatial structure defines a flux-tower-based VPRM residual covariance matrix. The residual covariance matrix is useful in constructing prior fluxes for atmospheric CO_2 concentration inversion calculations, as well as for constructing a VPRM North American CO_2 flux map optimized to eddy covariance observations. Finally, the estimated VPRM parameter values do not separate clearly by plant functional type (PFT). This calls into question whether PFTs partition ecosystems by carbon cycle participation when the viewing lens is a simple model.

1 Introduction

The rapid carbon dioxide (CO_2) accumulation in Earth's atmosphere in the second half of the 20th century (Conway et al., 2009) has been partially offset by natural biogeochemical processes. Without these buffers, atmospheric CO_2 could accumulate twice as fast: of the roughly 7 Pg of carbon humans release each year by burning fossil fuels, only roughly half remains in the atmosphere as carbon dioxide (Denman et al., 2007). The rest is absorbed by oceans through air-sea gas exchange or by terrestrial and marine flora through net primary production (NPP; Denman et al., 2007). Terrestrial biological fluxes of CO_2 through photosynthesis and respiration constituted a net sink from the atmosphere of two to three Pg C per year during the 1990s (Le Quéré et al., 2009), and they exhibit higher interannual variability than oceanic fluxes (Bousquet et al., 2000; Le Quéré et al., 2009). Understanding terrestrial fluxes is crucial to understanding and predicting the increase in atmospheric CO_2 caused by anthropogenic emissions.

Though the net global flux of CO_2 to the atmosphere is well-constrained (Tans and Conway, 2005), continental biological CO_2 fluxes are not well characterized, and their drivers are, so far, poorly understood. Diagnostic skill at interannual timescales is poor: land surface models consistently fail to capture observed CO_2 flux interannual variability (e.g., Friend et al., 2007; Ricciuto et al., 2008; Prentice et al., 2000). Predictive skill is also poor: a sampling of terrestrial flux models project terrestrial sink strengths for the year 2100 that vary widely in magnitude and sign (Friedlingstein et al., 2006).

There are a number of sources of information available to constrain terrestrial fluxes. These include direct eddy covariance flux observations, observed atmospheric CO_2 concentrations coupled with atmospheric transport models, and land surface models.

Land surface models integrate ecological and meterological drivers into a quantitative biological carbon flux estimate for some land region. They are useful because they can be used to extrapolate over large scales. Direct observation footprints of even the most spatially dense CO_2 flux observation networks cover only a tiny fraction of the land areas they span. For example, even with seven eddy covariance (EC) towers in a roughly 50 km by 75 km area, Goulden et al. (2006) estimate that they directly observe less than 0.01 % of that space. Land surface models estimate fluxes where direct observations do not exist. Improving model diagnoses of the magnitudes and drivers of terrestrial fluxes is a necessary step toward improving overall predictive skill.

Atmospheric inversion calculations (e.g., Rayner et al., 1999; Gurney et al., 2002; Rödenbeck et al., 2003; Peters et al., 2005, 2007) offer one approach to use observed atmospheric CO_2 concentrations coupled to an atmospheric transport model to further constrain terrestrial CO_2 flux diagnoses. This approach usually divides the planet into regions, and within each region typically solves for a correction to a prior flux estimate from a land surface model (e.g. Gurney et al., 2002). In regions of the world where CO_2 concentration observations are scarce there is little information with which to correct the prior flux, and the resulting flux estimations are therefore heavily dependent on the prior. Rödenbeck et al. (2003) introduced a prescribed isotropic spatial covariance to the prior, choosing a spatial correlation scale of 1275 km based on the average scale of autocorrelation among four different land surface models examined by McGuire et al. (2001). Peters et al. (2005, 2007) propagate inferred surface fluxes forward through time instead of using a prior flux estimate calculated offline for each time step before beginning the inversion calculations. They also use an ensemble Kalman filter to estimate a surface flux spatial covariance matrix. This relies on the assumption that flux errors are independent at weekly time scales and at spatial regions of 25 % to 50 % of each continent (Peters et al., 2005), an assumption that is conventional, though most likely not strictly accurate (Peters et al., 2005).

Jacobson et al. (2007a,b) describe an inversion approach that does not rely on prior fluxes on the grounds that modeled regional prior fluxes must either be assumed to be independent or treated as spatially correlated with an explicit spatial structure. In reality they are often correlated, though the quantitative correlations are unknown (Jacobson et al., 2007b, auxiliary materials). Jacobson et al. (2007b) show that this assumption of independence results in overconfident flux estimates. By eschewing prior flux estimates the Jacobson et al. (2007a,b) study avoids this pitfall, but at the cost of ignoring the knowledge of ecosystem behavior encapsulated in the flux model: the resulting posterior flux uncertainties are much larger than when modeled prior fluxes are included. That is, removing the information provided by a land surface model removes a significant constraint from the estimation.

Ideally, a flux diagnosis method would integrate all available sources of information. Here we focus on extracting information from a land surface model, while minimizing the overconfidence-producing assumptions demonstrated by Jacobson et al. (2007a,b). Model-data residuals are the combined effects of flux observation errors, model structural error, and natural variability uncaptured by the flux model. The spatial behavior of model residuals sheds light on the processes that drive fluxes. In the following analyses we will use that information to produce a data-derived quantification of the model-data residuals and their spatial structure that may be used to constrain the fluxes.

Terrestrial flux drivers (meteorological and ecological) do not appreciably vary at scales of, say, one centimeter; therefore model residuals should be correlated within some distance, however small. That distance is an upper bound on the area that a model result for a single spatial point can illuminate. Without a quantitative method for determining a correlation structure for model residuals, it is convenient to assume model residuals are independent and identically distributed (i.i.d.) in space and time. For example, Pacala et al. (2001), Peylin et al. (2002), Gurney et al. (2002) and Peters et al. (2005) adopt this assumption in their inversions. In fact, inversions that solve for corrections to regional prior fluxes intrinsically assume that prior flux residuals are correlated within the time scale and spatial scale of the inversion (Rödenbeck et al., 2003; Michalak et al., 2004). If they are not, the inversion applies a uniform correction to a group of uncorrelated residuals, creating a source of error (Chevallier et al., 2006). As noted, Rödenbeck et al. (2003) impose a prior flux uncertainty spatial correlation length scale of 1275 km. They base that distance on the autocorrelation length scales of the four models used by McGuire et al. (2001). This depends on the assumption that the NEE range among those four models is representative of flux model uncertainty (Rödenbeck et al., 2003). Furthermore, Michalak et al. (2004) point out that spatial structure, if existent, contains information that constrains fluxes and suggests weights for fluxes to identify and remove redundant information.

We can improve on existing flux diagnoses by deriving a residual covariance matrix to characterize the spatial behavior of flux model residual correlation. A necessary (and independently useful) prerequisite for estimating a model's residual covariance matrix is an estimation of the spatial scale at which the model's residuals are correlated. Here

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

we present an analysis of the residual covariance matrix of VPRM (Mahadevan et al., 2008), a simple land surface model.

We test the hypothesis that VPRM model residuals are spatially correlated at length scales smaller than the North American continent but larger than an individual EC tower footprint. Analyzing the spatial scale of VPRM residual correlation will provide that length scale. The Ameriflux and Fluxnet Canada networks of EC towers provide observations that allow us to directly analyze the spatial behavior of VPRM residuals. If that correlation length scale proves larger than the tower footprints, it will prove that the network of EC flux towers in North America has sufficient spatial span and density and has collected enough data across time to empirically define a flux model residual covariance matrix.

2 Methods

2.1 Land surface model

The Vegetation Photosynthesis and Respiration Model (VPRM) of Mahadevan et al. (2008) is a simple diagnostic terrestrial flux model. In spite of its simplicity, VPRM captures daily and annual cycles in CO_2 fluxes reasonably well (Mahadevan et al., 2008). VPRM structure and skill are described in great detail by Mahadevan et al. (2008). Here we provide a brief overview of the model structure.

VPRM models net ecosystem exchange (NEE) as the sum of a photosynthetic component (gross ecosystem exchange, GEE) and an ecosystem respiration component. GEE is modeled via the equation:

$$\mathrm{GEE} = \lambda \cdot T_{\mathrm{scale}} \cdot P_{\mathrm{scale}} \cdot W_{\mathrm{scale}} \cdot \mathrm{EVI} \cdot \frac{1}{1 + \mathrm{PAR}/\mathrm{PAR}_0} \cdot \mathrm{PAR}. \quad (1)$$

PAR is observed photosynthetically active radiation and EVI is the satellite-derived enhanced vegetation index (Huete et al., 2002). P_{scale} and W_{scale} are satellite-derived dimensionless scaling terms describing phenology and canopy moisture. The value of the third dimensionless scaling term T_{scale} is taken from literature and describes the relationship between photosynthesis and temperature. P_{scale}, W_{scale}, and T_{scale} may vary in both time and splace. PAR_0 and λ are model parameters.

Respiration (R) is modeled as a linear function of observed air temperature (T):

$$R = \alpha \cdot T + \beta \tag{2}$$

with parameters α and β.

NEE is the difference between the photosynthetic flux and the respiration flux:

$$\text{NEE} = R - \text{GEE}. \tag{3}$$

Within Eqs. (1), (2), and (3), λ governs the slope of the light-response curve (the relationship between photosynthetic CO_2 flux and PAR). α defines the slope of the respiration response to temperature. PAR_0 defines a half-saturation value for photosynthesis. That is, it specifies a PAR value at which further increases in PAR no longer enhance photosynthesis, as other limiting factors become dominant. VPRM places a PFT-specific floor T_{low}, $1\,^{\circ}\text{C} \leq T_{\text{low}} \leq 5\,^{\circ}\text{C}$, on surface temperatures. Temperatures below T_{low} are raised to T_{low} when calculating respiration. β thus specifies a minimal level of respiration that occurs regardless of air temperature.

In its simplicity, VPRM offers two important advantages over more complex models. First, it has only four user-defined parameters and is computationally inexpensive. This makes parameter estimation via data assimilation methods that do not require parametric assumptions computationally tractable. Second, as inputs, VPRM requires only air temperature, photosynthetically active radiation (PAR), and satellite-derived vegetation and moisture indices. It can thus be run globally, with no need to compile temporally-filled meteorological driver data. These advantages make VPRM a useful tool both for producing diagnostic regional flux maps, and also for evaluating spatial scales of model residuals in the manner of Chevallier et al. (2006).

2.2 Data

To constrain VPRM parameter values and examine NEE residuals we use data from 65 North American eddy covariance flux towers. Figure 1 shows the sites on a map, and Table 1 lists the sites and dominant plant functional type (PFT). These data are part of the 2007 Fluxnet Synthesis Dataset (http://www.fluxdata.org). For each site, this dataset contains CO_2 flux, air temperature, and PAR observations at 30-min intervals, as well as many other quantities not needed for VPRM.

The Fluxnet dataset contains gap-filled NEE, as well as non-filled NEE. Structurally, VPRM does not consider driver data or flux results from previous time steps (Eq. 3). VPRM simply does not report an NEE at timesteps where the required driver data are not available. In light of this, and to reduce potential residuals due to gap filling, we use the non-filled data.

The 65 observation sites cover nine of the 17 PFTs of the International Geosphere-Biosphere Programme (IGBP) land cover classification scheme (Loveland and Belward, 1997): evergreen needleleaf forest (27 sites), deciduous broadleaf forest (8 sites), mixed forest (3 sites), closed shrublands (7 sites), open shrublands (2 sites), woody savannas (1 site), grasslands (7 sites), permanent wetlands (4 sites), and croplands (6 sites). The eight PFTs not represented are: deciduous needleleaf forest, evergreen broadleaf forest, savannas, water, cropland/natural vegetation mosaic, urban and built-up, snow and ice, and barren or sparsely vegetated.

Site phenology, land surface water, enhanced vegetation index (Huete et al., 2002), and land surface cover type are calculated from reflectances measured by the NOAA MODIS instrument, orbiting with the NASA Terra satellite since 2000 and the NASA Aqua satellite since 2002. Oak Ridge National Laboratory extracts MODIS data for many Fluxnet tower sites and makes them available on the world wide web (ORNL DAAC, 2007).

For the present study, we use MODIS Collection 5 data (ORNL DAAC, 2010). Collection 5 data offer improved processing algorithms from Collection 4 (Friedl et al., 2010;

Didan and Huete, 2006). Site phenology is from dataset M*D12Q2 (Strahler et al., 1999a), reflectances are from dataset M*D43A4 (Strahler et al., 1999b), vegetation indices are from dataset M*D13A2 (Huete et al., 2002, 1999), and IGBP land surface cover types are from dataset M*D12Q1 (Loveland and Belward, 1997; Strahler et al., 1999a). The “*” in M*D is either “O”, representing the data from the Terra satellite, or “Y” representing the data from the Aqua satellite. We considered only MODIS data of “best” quality, as indicated by each MODIS product’s associated quality QA flags.

We examine the time period 2000 to 2006, bounded in 2000 by the MODIS instrument launch and in 2006 by eddy covariance flux availability.

2.3 VPRM parameter estimation

VPRM has four user-estimated parameters that may depend on the location being simulated: λ, PAR_0, α, and β. In this section we describe how we estimated these parameter values.

We seek the parameter values that cause VPRM NEE to match observed NEE as closely as possible. We chose to minimize the sum of squared errors (SSE; we define VPRM residuals as $\mathrm{NEE}_{\mathrm{VPRM}}$ minus $\mathrm{NEE}_{\mathrm{observed}}$). If the residuals are normally distributed with a constant standard deviation (i.e., homoskedastic), minimizing SSE is equivalent to a maximum likelihood estimate (Hilborn and Mangel, 1997).

In reality, flux model residuals are neither independent nor identically distributed. A double exponential distribution describes EC observation error better than the normal distribution (Richardson et al., 2006). EC observation error is also proportional to NEE magnitude and wind speed (Richardson et al., 2006). Thus the strong daily and seasonal cycles of NEE cause EC observation errors to be temporally autocorrelated as well as heteroskedastic. The combined impact of land surface model structural error, incorrect parameter values, and natural variability – microscale variations in climate, ecosystem behavior, etc. – may also exhibit statistically significant autocorrelation (Ricciuto et al., 2008). We can approximate distributions for each of those error sources from published literature; therefore the full likelihood function may be written out as the integrated product of likelihood functions for several different statistical distributions. Reducing that integral to an anlytical solution, however, is beyond the scope of this study.

NEE varies on a number of different time scales (e.g. daily, annual) and space scales (e.g. local land-use and PFT heterogeneity, larger regions that experience similar climate patterns). An ideal land surface model parameter estimation method would allow parameter values to vary at space and time scales matching the ecological variations in NEE. Optimizing parameter values in short time intervals and small spatial windows would run the risk of overfitting as well as incur unnecessary computational cost. We chose to examine three temporal and three spatial windows for SSE minimization – in time: monthly, annual, and all available data; and, in space: individual sites, sites grouped by PFT, and all sites together. This approach yields nine different parameter sets, ranging from four to more than 21 000 parameter values. Table 2 summarizes the nine parameter sets.

To search for parameter values that minimize SSE we used differential evolution (DE) (Price et al., 2006). DE is a genetic optimization algorithm that is both fast and more reliable in identifying a global optimum compared to gradient-based minimization algorithms. We used the DEoptim package (Ardia and Mullen, 2009) for the R language and platform for statistical computing (R Development Core Team, 2007).

2.4 Quantifying spatial structure

The spatial covariance structure quantifies the spatial structure (or lack thereof) for an arbitrary function of space. The semivariogram offers a concise visual summary of the covariance structure. The spatial functions of interest here are VPRM NEE residuals, VPRM NEE, and observed NEE. This section defines the semivariogram and describes its typical behavior for geophysical quantities.

The semivariogram (γ) is generically defined (Cressie, 1993) as:

$$\gamma(h) = \frac{1}{2}\mathrm{var}(Z(s_i) - Z(s_j)), \tag{4}$$

where s_i and s_j are two locations in space, h is the distance between s_i and s_j, var denotes variance, and Z is some function of location–air temperature, VPRM NEE residual, etc. If s_i and s_j are near one another, one might expect $Z(s_i)$ and $Z(s_j)$ to have similar values, causing γ to be correspondingly small. As h increases, $Z(s_i)$ and $Z(s_j)$ typically diverge, and the value of γ increases. At some sufficiently large h, $Z(s_i)$ and $Z(s_j)$ can become independent, causing γ to level off. The value of h where the leveling-off occurs is known as the range, and the value of γ at this leveling-off is known as the sill. The range estimates the length scale of spatial correlation in Z. These easily-visualized semivariogram features are formal parameters (range, ϕ; sill or variance, σ^2) of the covariance function.

In the same way that the population mean provides a statistical estimator for a population's expected value, there is a statistical estimator to calculate an empirical semivariogram ($\hat{\gamma}$) from a set of observed data (Cressie, 1993):

$$\hat{\gamma}(h) \equiv \frac{1}{2|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2, \tag{5}$$

where $N(h)$ is the number of location pairs separated by distance h and the $\hat{\gamma}$ notation distinguishes the estimated semivariogram from the theoretical definition of Eq. (4); other terms are defined above. The separation distance h may be a precise distance for a single pair of locations, or may be an aggregated separation distance for a number of pairs of locations.

In this study we use the "robust" semivariogram estimator of Cressie and Hawkins (1980). This estimator includes a correction term for non-normally distributed data, and also reduces the impact of outlying data:

$$\overline{\gamma}(h) \equiv \frac{\left(\frac{1}{|N(h)|} \sum_{N(h)} |Z(s_i) - Z(s_j)|^{0.5}\right)^4}{2\left(0.457 + \frac{0.494}{|N(h)|}\right)} \tag{6}$$

Purely mathematically, Eq. (4) requires the semivariogram to equal zero at $h = 0$, because $Z(s_i) - Z(s_i) = 0$. In practice, measurement errors cause repeated measurements at a single location to differ. Moreover, measurements are not made at infinitesimally small separation distances. There is no information about $\gamma(h)$ at distances below the minimum separation distance h present in the data. This unknown behavior at small h is sometimes called microscale variation. When microscale variation or measurement error are present, $\hat{\gamma}(h)$ does not approach zero as h approaches zero. The value of $\hat{\gamma}(h = 0)$ is known as the semivariogram nugget, denoted by τ^2. Together, the sill, range, and nugget characterize the semivariogram and yield much information about the spatial structure of Z.

In addition to providing the length scale of spatial correlation for Z, the semivariogram also specifies the spatial covariance of Z. Specifically,

$$\mathrm{cov}(Z(s_1), Z(s_2)) = \frac{1}{2}\mathrm{var}(Z(s_1)) + \frac{1}{2}\mathrm{var}(Z(s_2)) - \frac{1}{2}\mathrm{var}(Z(s_1) - Z(s_2)) \qquad (7)$$

with cov denoting covariance, expresses the spatial covariance of VPRM residuals in terms of available quantities: the first two terms on the right side of Eq. (7) are the variance within individual sites and the last term is the semivariogram.

Covariance parameters ϕ, σ^2, and τ^2 may be estimated directly from spatial data via maximum likelihood estimation (MLE) by maximizing the log-likelihood function (Diggle and Ribeiro Jr., 2007):

$$\begin{aligned} L(\beta, \tau^2, \sigma^2, \phi) = -\,0.5 \Big\{ & n \log(2\pi) + \log\Big\{ \left| \sigma^2 R(\phi) + \tau^2 I \right| \Big\} \\ & + (y - D\beta)^T (\sigma^2 R(\phi) + \tau^2 I)^{-1} (y - D\beta) \Big\} \end{aligned} \qquad (8)$$

with the covariance matrix $\sigma^2 R(\phi) + \tau^2 I$ expressed in terms of ϕ (range), σ^2 (sill), and τ^2 (nugget). The residual matrix $(y - D\beta)$ is the difference between observations y and a model structure given by $D\beta$ with model explanatory variables D and model parameters β. MLE is the preferred approach for formal covariance parameter estimation

(Diggle and Ribeiro Jr., 2007) in large part because it considers the full set of available data rather than relying on the summary provided by an empirical semivariogram.

Fitting a parametric covariance function to observed VPRM residuals provides three key outcomes: (i) the range for VPRM NEE residuals; (ii) covariance of VPRM NEE residuals at arbitrary separation distances – that is, a residual covariance matrix; and (iii), via kriging, a VPRM NEE map that explicitly considers VPRM NEE residuals (Cressie, 1993). Figure 2 shows semivariograms for two common parametric covariance functions. When a spatial field has no spatial correlation, its semivariogram looks like the pure nugget covariance function. The exponential covariance function is one example of a model describing a spatial field that is correlated in space to a certain distance and uncorrelated beyond that distance.

For the nine VPRM parameterizations of Sect. 2.3 we calculated seasonal mean VPRM residuals. We defined seasons as December-January-February (DJF), March-April-May (MAM), June-July-August (JJA), and September-October-November (SON). Within each season we maximized the negative log likelihood (Eq. 8) to estimate covariance parameters for both the pure nugget as well as exponential covariance functions (Fig. 2). We then compared the pure nugget and exponential fits using AIC (Akaike, 1976). This experiment determines whether the observed VPRM NEE residuals are better described as covarying in space at some length scale (the exponential covariance model) or as spatially independent even at minimal distances (the pure nugget model). We follow this experiment with two pseudodata experiments to assess the tendency of 65 observation locations spread across North America and our AIC test to choose the exponential covariance function when no spatial covariance is present, or to choose the pure nugget covariance model when the underlying field was generated from an exponential covariance model. The rest of this paper describes our parametrization of VPRM and our analysis of VPRM NEE residual spatial structure.

3 Results

3.1 VPRM parametrization

As described in Sect. 2.3, we calculated VPRM parameter values for nine different groupings of those sites in space and time (Table 2), conditioned on observations from 65 North American eddy covariance sites (Table 1, Fig. 1).

Figure 3 shows the distribution of VPRM parameter values when estimated monthly within each PFT. Parameter distributions across PFTs for the other eight parameter sets in Table 2 are nearly identical to Fig. 3. The parameter distributions are similar to those of Mahadevan et al. (2008). Most striking in Fig. 3 is the failure of the parameterization to distinguish among plant functional types.

It is perhaps unexpected that VPRM parameters do not cluster by plant functional type. For example, one might expect that the model parameter estimates of a boreal needleleaf forest should be different from a cropland, for example. There is evidence that light-use efficiency (LUE) is not consistent within PFTs (Ruimy et al., 1994; Schwalm et al., 2006), particularly at daily timescales (Schwalm et al., 2006). Another recent study assumes that maximum LUE is constant across PFTs (Yuan et al., 2007). Schwalm et al. (2006) also suggest intra-PFT LUE varies less at annual time-scales than at daily scales. The values of λ (the VPRM LUE parameter) in Fig. 3, relatively invariant across different PFTs, differ from the results of Schwalm et al. (2006). VPRM respiration parameters α and β also do not vary much across PFTs; this is consistent with previous studies indicating that PFTs are not predictive of soil respiration (Raich and Tufekciogul, 2000; Bond-Lamberty et al., 2004).

The similar parameter values in Fig. 3 could be a consequence of VPRM's simplicity; perhaps a two-equation model which takes climatology and phenology from satellite observations is only able to separate landscapes into "green – photosynthesizing" and "brown – not photosynthesizing." These results offer hints; investigating the question rigorously would require parameter PDFs to ascertain whether the differences in Fig. 3 are significant. That investigation should also compare model fluxes across different parameterizations. It is possible, for example, that the remote sensing data that drive VPRM are sufficient to separate the NEE of different plant functional types without large parameter differences. The question is intriguing, however. If PFTs truly are not important for NEE diagnosis and prediction, the task of estimating model parameters becomes much simpler: land surfaces may then be simply classified as “green” and “not-green.”

3.2 VPRM NEE residual spatial structure

Qualitatively inspecting the shape of an empirical semivariogram gives an intuitive sense for a function’s spatial covariance. Figure 4 plots binned semivariograms for June-July-August (JJA) mean VPRM NEE residuals. There is one curve for each of the nine VPRM parameter sets considered (Table 2); each point shows the mean semivariance within a 300 km bin.

The nugget is small for the site-specific parameter sets (black curves), and varies from 1 to 3 $(\mu\text{mol}\,\text{CO}_2\,\text{m}^{-2}\,\text{s}^{-1})^2$, for the other six parameter sets. In units of standard deviation, these six nuggets equal roughly 2.0 $\mu\text{mol}\,\text{CO}_2\,\text{m}^{-2}\,\text{s}^{-1}$. The nugget represents the combined influence of variations at spatial scales smaller than the minimum separation distance as well as the contributors to VPRM residuals (EC observation error, VPRM structural error, natural variability; see Sect. 2.3).

In general, the semivariances for each parameter set increase from separation distances of 0 km to roughly 800 km, and level off or decrease thereafter. This suggests that VPRM NEE residuals are correlated at distances up to 800 km. We quantify this in the following results.

We are interested primarily in the parameters (range, sill, nugget) of the covariance function that best describes the VPRM residuals. To estimate these parameters, we employ maximum likelihood estimation (MLE, described by Eq. 8).

We fit both pure nugget and exponential covariance functions (whose characteristic semivariograms are shown in Fig. 2) to each of the nine sets of VPRM residuals summarized by the binned semivariograms in Fig. 4. Within each VPRM parameter set we selected either the best-fit pure nugget covariance function or the best-fit exponential covariance function using Akaike's Information Criterion (Akaike, 1974). AIC balances goodness of fit (more parameters) against parsimony (fewer parameters). Table 3 shows seasonal range values for the covariance function that optimally fit the observed VPRM NEE residuals. Blank entries show instances where the pure nugget function optimally fit the observed residuals; this indicates that no spatial correlation is present in the VPRM residuals. Among the 2000 to 2006 observed residuals, 92 of 252 (7 yr × 4 seasons × 9 PFTs) were best described by the exponential covariance function. Of those, the median range is 402 km.

To interpret this result, we must test the adequacy of 65 observation locations across North America (Fig. 1) to detect spatial correlations across hundreds of kilometers. The maximum distance between towers in this group of 65 is 6557 km (US-Atq – US-KS1). To quantitatively test the detection capacity of the data set we generated 1000 Gaussian Random Fields (GRFs) on a 6500 by 6500 grid. Each GRF had an imposed exponential covariance structure with a specified range of 402 km (equal to the median VPRM NEE error seasonal covariance range reported in Table 3.) We sampled each GRF at 65 randomly-generated locations and estimated exponential and pure nugget covariance function parameters for each sample set using MLE in the same manner that we estimated range values for the VPRM NEE residuals (Table 3). Of the 1000 GRFs, AIC chose the exponential covariance function for only 74. Of those 74, the median estimated covariance range is 936 km; the median estimated covariance range across all 1000 GRFs is 313 km. This distribution of estimated range values is similar to the distribution estimated from the real VPRM NEE residual observations; Fig. 5 plots the two distributions side by side. These results suggest that the estimated range values for VPRM NEE residuals (Table 3) are consistent with a scenario where VPRM NEE residuals have an exponential covariance structure with a range of roughly 400 km, and that 65 observation locations in the United States and Canada are minimally adequate for detecting that structure.

We also must consider the possibility of spurious MLE results: that the observed VPRM NEE residual realization may occasionally be better fit by an exponential covariance structure when the complete spatial field has no true structure. We generated another set of 1000 GRFs, each containing 65 points within a 6500 by 6500 grid, and each specified to have a pure nugget covariance structure. We calculated MLE covariance function parameters for these 1000 fields. AIC chose the exponential function over the pure nugget for only 25 of the 1000 fields, suggesting that we might expect a data set like our VPRM NEE residuals to produce a spurious exponential covariance structure in only a small minority of realizations considered.

These results suggest quantitatively that JJA mean VPRM NEE residuals are spatially correlated at a length scale on the order of 400 km.

Anomalies from the 2000 to 2006 means for annual cumulative VPRM NEE, annual cumulative observed NEE, and annual cumulative VPRM NEE residuals displayed similar spatial scales (Fig. 6). This analysis tests the hypothesis that while NEE itself varies significantly at spatial scales on the order of 10 km (e.g., Desai et al., 2008), NEE interannual variability (IAV) is driven by phenomena that operate at much larger scales. If so, then we should see spatial correlation in annual cumulative $\mathrm{NEE}_{\mathrm{obs}}$ anomalies. If VPRM is able to capture that large-scale variation, then annual cumulative $\mathrm{NEE}_{\mathrm{VPRM}}$ anomalies will show similar spatial correlation. Any spatial structure that exists in $\mathrm{NEE}_{\mathrm{obs}}$ anomalies that VPRM fails to capture should appear in VPRM NEE residual anomalies.

As with the VPRM NEE residual semivariograms, we chose optimal anomaly covariance structures by AIC. Of the seven years examined, $\mathrm{NEE}_{\mathrm{obs}}$ anomalies show correlation at scales of roughly 1000 km only for 2006 (Fig. 6). This rate of detection is consistent with that of the pseudodata experiment (Fig. 5), in which we were able to detect a known exponential covariance structure in only 62 of 1000 attempts. This could indicate that large-scale structure does not consistently exist. It could also suggest that NEE interannual variability could be shaped by larger-scale drivers than is NEE itself, and that our flux tower spatial density is insufficient to consistently detect it in a noisy NEE signal. This seems reasonable; land use, which influences NEE, is markedly diverse throughout the study area. Also, disturbance events that heavily influence NEE (e.g. fire, insects, tree harvest) usually do not impact 500-km stretches of land surface. VPRM is strongly driven by climate variables (Eqs. 1, 2), so spatial structure in VPRM NEE interannual variability could simply reflect large-scale spatial structure in climatic interannual variability. Though VPRM no doubt contains structural error, it is an attempt to combine climatic terms as ecological research suggests they influence NEE. Therefore, we believe it makes sense to investigate this combined effect of several climate terms (that is, VPRM NEE) rather than attempt to explain NEE interannual variability by searching for spatial coherence in a number of climate variables individually.

Because VPRM NEE residuals are simply the difference between $\mathrm{NEE}_{\mathrm{obs}}$ and $\mathrm{NEE}_{\mathrm{VPRM}}$, the spatial behaviors of these three quantities are interrelated. Where spatial structure exists in observations, we expect it to be partitioned among $\mathrm{NEE}_{\mathrm{VPRM}}$ and VPRM NEE residuals. Results in Fig. 6 from all nine VPRM parameter sets show strong spatial structure in VPRM NEE residuals. This structure occurs at length scales similar to the length scale exhibited by $\mathrm{NEE}_{\mathrm{obs}}$. Sill and nugget values for $\mathrm{NEE}_{\mathrm{VPRM}}$ and VPRM NEE residuals are also of similar magnitude to the sill and nugget for $\mathrm{NEE}_{\mathrm{obs}}$. VPRM NEE residuals are the combination of NEE observation error, VPRM structural error, and natural variability. Because of its correlation to NEE magnitude (Richardson et al., 2006), we expect the NEE observation error component of VPRM residuals to reflect whatever spatial structure is present in NEE itself. It therefore makes sense that the spatial structure present in $\mathrm{NEE}_{\mathrm{obs}}$ is not partitioned exclusively into $\mathrm{NEE}_{\mathrm{VPRM}}$ or VPRM NEE residuals, but appears in both.

The covariance sill value provides an estimate of variance. The sill values (Fig. 7) for the annual anomalies of annual cumulative VPRM NEE, annual cumulative observed NEE, and annual cumulative VPRM NEE residuals display standard deviations (Fig. 7, right-side axis) on the order of the annual cumulative NEE typically observed by an eddy covariance site. This suggests that annual VPRM errors at a single location in space are on the order of the flux at that point. If annual VPRM errors are indeed spatially correlated at length scales of 500 km to 1000 km, as suggested by Fig. 6, then spatially aggregating VPRM NEE at that length scale should provide a method to reduce the VPRM error variance.

4 Discussion

Our findings are relevant to both land surface model upscaling as well as atmospheric inversion studies, though several important uncertainties should guide consideration of our results.

4.1 Caveats

Several caveats accompany these implications. The structural simplicity of VPRM allows us to conduct parameter estimations that use many thousands of model evaluations. The designers of VPRM achieve that simplicity by abstracting the broadest drivers of NEE out of what is in reality a complex ecology and by considering only short-term drivers of NEE. Longer-term drivers, such as carbon pools (e.g., Curtis et al., 2002) and disturbance histories (e.g., Thornton et al., 2002), are known to be first-order drivers as well. These simplifications caution us against attempting detailed ecological interpretation of the VPRM NEE results and VPRM residuals.

In addition, the carbon cycle community's understanding of the statistical properties of land surface model NEE residuals remains rudimentary. Several studies have explored the distribution of NEE observation error (e.g., Richardson et al., 2008, 2006). Richardson et al. (2006) find the observational error to exhibit a double exponential distribution Observation error, however, is but one component of NEE model residuals. In the absence of a rigorous likelihood function that integrates all of the sources of uncertainty that contribute to NEE model residuals, we have used the mathematically simple sum of squared NEE residuals to estimate VPRM parameters. Implementing a statistically proper likelihood function is non-trivial and is the subject of ongoing research.

Our spatial analysis of VPRM residuals compared a pure nugget model with no spatial covariance to an isotropic exponential covariance function that treated all land surface classifications as equal. It is possible that VPRM residuals covary differently in the East-West direction than North-South, or that plant functional types, disturbance history, or some other land surface descriptor is important. The present spatial density of eddy covariance observations limits our ability to test these.

4.2 Implications

It is critically important to quantitatively tailor NEE model parameter estimates to the domain in which the model is to be run; generic parameter values can reproduce observed NEE poorly (Ricciuto, 2006). Good NEE simulation is crucial to calculating accurate model errors, which are in turn crucial to detecting model error spatial structure.

Our finding that VPRM does not resolve different PFTs through its parameter values can be viewed in at least two different lights. First, studies wishing to provide first-order regional NEE estimates via a low-complexity land surface model may not need to distinguish among PFTs for parameterization on pure statistical grounds. This could lead to considerable savings in computation time and CPU resources. Second, PFTs are commonly assumed to partition land into sections with functionally different participations in the carbon cycle. Our results suggest that PFTs may not be the most useful predictor of a land area's carbon cycle dynamics, and that alternative partitioning schemes may be more skillful. Stand age and disturbance history are interesting "land surface NEE descriptor" alternatives to PFTs. Thornton et al. (2002) used the BiomeBGC model to explore the impacts of distubrance history, PFT, site climate, atmospheric CO_2 concentration, and nitrogen deposition on NEE variability among seven evergreen sites spanning North America, and concluded that of those, disturbance history dominated. Goulden et al. (2006) examined seven eddy covariance sites within 50 km of each other that were recovering from burn disturbances that occurred 0, 5, 14, 22, 39, ~ 73, and ~ 153 yr previously. They found that mid-growing season EVI and CO_2 fluxes took roughly 50 yr following a burn disturbance to become approximately interannually constant. That 50-yr period included transition from primarily deciduous species to primarily black spruce. These results and others suggest that disturbance history could be at least important as climate and plant functional type to understanding NEE for large areas.

The results of Goulden et al. (2006) suggest, at least for boreal evergreen forests, a satellite record on the order of 50 to 100 yr or longer could be necessary before stand age and recovery from disturbance can be widely and directly described by remote sensing. Recent landsat products have begun to assemble landscape disturbance records beginning in the 1980s (Huang et al., 2010), offering an opportunity to assess these influences at larger scales.

The spatial length scale of land surface model NEE residual covariance bears directly on atmospheric inversion calculations. Inversions seek to use observed atmospheric CO_2 concentrations to refine estimated biological CO_2 fluxes within a region of defined boundaries, with the estimated fluxes typically coming from models. Intrinsic to the method is the assumption that prior flux errors are correlated within each region treated as a separate unknown (Rödenbeck et al., 2003; Michalak et al., 2004). Moreover, this correlation must be assumed to exist at both the time scale of the inversion as well as the spatial scale of the inversion regions. Our results indicate strongly that this implicit assumption is valid at seasonal time scales (Table 3) and, for annual anomalies, for annual time scales. The relevant spatial scale is approximately 400 km. This length scale is smaller than the scale of 1275 km estimated by Rödenbeck et al. (2003), and is based on eddy covariance flux measurements rather than land surface model comparison. Our length scale also contrasts starkly with the conclusion of minimal spatial covariance presented by Chevallier et al. (2006). Potentially incorrect prior flux error covariance assumptions are but one source of error that an inversion must consider. Scarcity of well-calibrated CO_2 concentration observations, for example, pushes inversion calculations toward regions larger than 1000 km (e.g., Butler et al., 2010). The North American Carbon Program's Mid-Continental Intensive (MCI) region is a notable exception to this scarcity of CO_2 concentration observations (Lauvaux et al., 2011), and presents an opportunity to investigate the impacts of prior flux error covariance assumptions more deeply. Solving for too many regions in an inversion (that is, too many unknowns) risks overfitting the data, and solving for too few risks oversimplifying the inversion and producing over-confident results. We suggest that inversion calculations should optimally use regions with spatial scales on the order of 400 km.

5 Conclusions

Using observed NEE from 65 North American eddy covariance sites for the years 2000 through 2006, we make point estimates of parameter values for VPRM, a simple land surface model. We then estimate and analyze covariance structures of VPRM NEE residuals in the interest of quantifying spatial structure in the residuals.

PFTs demonstrate little skill as land surface classifications for model parameter estimation. This may allow large-region model studies to partition land surfaces into a “photosynthetically active or not” dichotomy, thereby simplifying model parameterization.

The semivariogram analyses presented here demonstrate that VPRM NEE residuals are spatially correlated at length scales well beyond individual tower footprints but well short of continental scales. Depending on the model parameterization, that length scale lies somewhere between 100 km and 900 km, with a median value of roughly 400 km. This result is consistent at both seasonal and interannual time scales, and demonstrates that the North American EC tower network is minimally sufficient to define a VPRM residual covariance matrix. This information will allow us to construct a map of VPRM North American CO_2 fluxes, optimized to eddy covariance observations.

Our estimated covariance functions for model NEE residuals prove that the North American flux tower observation network is adequate for determining a land surface model residual covariance matrix.

Acknowledgements. The authors with to thank M. Haran of the Penn State Department of Statistics for providing advice on the geostatistical methods. Data were processed using 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 # 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

Akaike, H.: A new look at the statistical model identification, IEEE T. Automat. Contr., 19, 716–723, 1974. 7088

Akaike, H.: An information criterion (AIC), Math. Sci., 14, , 14, 5–9, 1976. 7085

Anthoni, P., Unsworth, M., Law, B., Irvine, J., Baldocchi, D., Tuyl, S., and Moore, D.: Seasonal differences in carbon and water vapor exchange in young and old-growth ponderosa pine ecosystems, Agr. Forest Meteorol., 111, 203–222, 2002. 7107

Ardia, D. and Mullen, K.: DEoptim: Differential Evolution Optimization in R, R package version 2.0-1, available online: http://CRAN.R-project.org/package=DEoptim (last access: 1 June 2012), 2009. 7082, 7095

Bergeron, O., Margolis, H., Black, T., Coursolle, C., Dunn, A., Barr, A., and Wofsy, S.: Comparison of carbon dioxide fluxes over three boreal black spruce forests in Canada, Glob. Change Biol., 13, 89–107, 2007. 7106

Black, T., Chen, W., Barr, A., Arain, M., Chen, Z., Nesic, Z., Hogg, E., Neumann, H., and Yang, P.: Increased carbon sequestration by a boreal deciduous forest in years with a warm spring, Geophys. Res. Lett., 27, 1271–1274, 2000. 7106

Bond-Lamberty, B., Wang, C., and Gower, S.: A global relationship between the heterotrophic and autotrophic components of soil respiration?, Glob. Change Biol., 10, 1756–1766, doi:10.1111/j.1365-2486.2004.00816.x, 2004. 7086

Bousquet, P., Peylin, P., Ciais, P., Le Quéré, C., Friedlingstein, P., and Tans, P. P.: Regional changes in carbon dioxide fluxes of land and oceans since 1980, Science, 290, 1342–1347, 2000. 7075

Bracho, R., Powell, T., Dore, S., Li, J., Hinkle, C., and Drake, B.: Environmental and biological controls on water and energy exchange in Florida scrub oak and pine flatwoods ecosystems, J. Geophys. Res., 113, G02004, doi:10.1029/2007JG000469, 2008. 7107

Bracho, R. G., Starr, G., Gholz, H., Martin, T. A., Cropper, J. W. P., and Loescher, H. W.: Controls on carbon dynamics by ecosystem structure and climate for Southeastern US slash pine plantations, Ecol. Monogr., 82, 101–128, http://dx.doi.org/10.1890/11-0587.1, doi: 10.1890/11-0587.1, 2012. 7107

Butler, M. P., Davis, K. J., Denning, A. S., and Kawa, S. R.: Using continental observations in global atmospheric inversions of CO_2: North American carbon sources and sinks, Tellus B, 62, 550–572, doi:10.1111/j.1600-0889.2010.00501.x, 2010. 7093

Chevallier, F., Viovy, N., Reichstein, M., and Ciais, P.: On the assignment of prior errors in Bayesian inversions of CO_2 surface fluxes, Geophys. Res. Lett., 33, L13802, doi:10.1029/2006GL026496, 2006. 7077, 7079, 7093

Conway, T., Lang, P., and Masarie, K.: Atmospheric carbon dioxide dry air mole fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 1968–2008, Version: 2009-07-15, available online: ftp://ftp.cmdl.noaa.gov/ccg/co2/flask/event/ (last access: 1 June 2012), 2009. 7074

Cook, B. D., Davis, K. J., Wang, W. Q., Desai, A., Berger, B. W., Teclaw, R. M., Martin, J. G., Bolstad, P. V., Bakwin, P. S., Yi, C. X., and Heilman, W.: Carbon exchange and venting anomalies in an upland deciduous forest in northern Wisconsin, USA, Agr. Forest Meteorol., 126, 271–295, doi:10.1016/j.agrformet.2004.06.008, 2004. 7107

Cressie, N.: Statistics for Spatial Data, Wiley & Sons, New York, USA, 2nd Edn., 1993. 7082, 7083, 7085

Cressie, N. and Hawkins, D.: Robust estimation of the variogram: I, Math. Geol., 12, 115–125, 1980. 7083

Curtis, P. S., Hanson, P. J., Bolstad, P., Barford, C., Randolph, J. C., Schmid, H. P., and Wilson, K. B.: Biometric and eddy-covariance based estimates of annual carbon storage in five Eastern North American deciduous forests, Agr. Forest Meteorol., 113, 3–19, doi:10.1016/S0168-1923(02)00099-0, 2002. 7091

Davis, K., Bakwin, P. S., Yi, C., Berger, B., Zhao, C., Teclaw, R., and Isebrands, J.: The annual cycles of CO_2 and H_2O exchange over a northern mixed forest as observed from a very tall tower, Glob. Change Biol., 9, 1278–1293, 2003. 7107

Denman, K., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P., Dickinson, R., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P., Wofsy, S., and Zhang, X.: Couplings between changes in the climate system and biogeochemistry, in: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and Miller, H., Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007. 7074

Desai, A., Bolstad, P., Cook, B., Davis, K., and Carey, E.: Comparing net ecosystem exchange of carbon dioxide between an old-growth and mature forest in the upper Midwest, USA, Agr. Forest Meteorol., 128, 33–55, 2005. 7107

Desai, A., Noormets, A., Bolstad, P., Chen, J., Cook, B., Davis, K., Euskirchen, E., Gough, C., Martin, J., Ricciuto, D. M., Schmid, H., Tang, J., and Wang, W.: Influence of vegetation and seasonal forcing on carbon dioxide fluxes across the Upper Midwest, USA: implications for regional scaling, Agr. Forest Meteorol., 148, 288–308, 2008. 7089

Didan, K. and Huete, A.: MODIS Vegetation Index Product Series Collection 5 Change Summary, Terrestrial Biophysics and Remote Sensing (TBRS) Lab, University of Arizona, available at: http://modland.nascom.nasa.gov/QA_WWW/forPage/MOD13_VI_C5_Changes_Document_06_28_06.pdf (last access: 1 June 2012), 2006. 7081

Diggle, P. and Ribeiro Jr., P.: Model-Based Geostatistics, Springer, New York, NY, USA, 2007. 7084, 7085

Fischer, M., Billesbach, D., Berry, J., Riley, W., and Torn, M.: Spatiotemporal variations in growing season exchanges of CO_2, H_2O, and sensible heat in agricultural fields of the Southern Great Plains, Earth Interact., 11, 1–21, 2007. 7106

Flanagan, L., Wever, L., and Carlson, P.: Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland, Glob. Change Biol., 8, 599–615, 2002. 7106

Friedl, M., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X.: MODIS collection 5 global land cover: Algorithm refinements and characterization of new datasets, Remote Sens. Environ., 114, 168–182, doi:10.1016/j.rse.2009.08.016, 2010. 7080

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Bala, G., John, J., Jones, C., Joos, F., Kato, T., Kawamiya, M., Knorr, W., Lindsay, K., Matthews, H. D., Raddatz, T., Rayner, P., Reick, C., Roeckner, E., Schnitzler, K.-G., Schnur, R., Strassmann, K., Weaver, A. J., and Yoshikawa, C. Z. N.: Climate-carbon cycle feedback analysis: Results from the C[4]MIP model intercomparison, J. Climate, 19, 3337–3353, doi:10.1175/JCLI3800.1, 2006. 7075

Friend, A., Arneth, A., Kiang, N., Lomas, M., Ogee, J., Rodenbeck, C., Running, S., Santaren, J., Sitch, S., Viovy, N., Woodward, F., and Zaehle, S.: FLUXNET and modelling the global carbon cycle, Glob. Change Biol., 13, 610–633, 2007. 7075

Giasson, M., Coursolle, C., and Margolis, H.: Ecosystem-level CO_2 fluxes from a boreal cutover in Eastern Canada before and after scarification, Agr. Forest Meteorol., 140, 23–40, 2006. 7106

Goldstein, A., Hultman, N., Fracheboud, J., Bauer, M., Panek, J., Xu, M., Qi, Y., Guenther, A., and Baugh, W.: Effects of climate variability on the carbon dioxide, water, and sensible heat fluxes above a ponderosa pine plantation in the Sierra Nevada (CA), Agr. Forest Meteorol., 101, 113–129, 2000. 7106

Gough, C., Vogel, C., Schmid, H., Su, H., and Curtis, P.: Multi-year convergence of biometric and meteorological estimates of forest carbon storage, Agr. Forest Meteorol., 148, 158–170, 2008. 7107

Goulden, M., Winston, G., McMillan, A., Litvak, M., Read, E., Rocha, A., and Rob Elliot, J.: An eddy covariance mesonet to measure the effect of forest age on land-atmosphere exchange, Glob. Change Biol., 12, 2146–2162, 2006. 7075, 7092, 7093, 7106

Gu, L., Meyers, T., Pallardy, S., Hanson, P., Yang, B., Heuer, M., Hosman, K., Riggs, J., Sluss, D., and Wullschleger, S.: Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site, J. Geophys. Res.-Atmos., 111, D16102, doi:10.1029/2006JD007161, 2006. 7107

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.-W.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, 2002. 7075, 7076, 7077

Hadley, J. and Schedlbauer, J.: Carbon exchange of an old-growth eastern hemlock (*Tsuga canadensis*) forest in Central New England, Tree Physiol., 22, 1079–1092, 2002. 7106

Harazono, Y., Mano, M., Miyata, A., Zulueta, R., and Oechel, W.: Inter-annual carbon dioxide uptake of a wet sedge tundra ecosystem in the Arctic, Tellus B, 55, 215–231, 2003. 7106

Hilborn, R. and Mangel, M.: The ecological detective: confronting models with data, Princeton University Press, Princeton, NJ, USA, 1997. 7081

Hollinger, D., Goltz, S., Davidson, E., Lee, J., Tu, K., and Valentine, H.: Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest, Glob. Change Biol., 5, 891–902, 1999. 7107

Hollinger, D. Y., Aber, J., Dail, B., Davidson, E. A., Goltz, S. M., Hughes, H., Leclerc, M. Y., Lee, J. T., Richardson, A. D., Rodrigues, C., Scott, N., Achuatavarier, D., and Walsh, J.: Spatial and temporal variability in forest-atmosphere CO_2 exchange, Glob. Change Biol., 10, 1689–1706, doi:10.1111/j.1365-2486.2004.00847.x, 2004. 7107

Howard, E., Gower, S., Foley, J., and Kucharik, C.: Effects of logging on carbon dynamics of a jack pine forest in Saskatchewan, Canada, Glob. Change Biol., 10, 1267–1284, 2004. 7106

Huang, C., Goward, S., Masek, J., Thomas, N., Zhu, Z., and Vogelmann, J.: An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks, Remote Sens. Environ., 114, 183–198, 2010. 7093

Huete, A., Justice, C., and van Leeuwen, W.: MODIS vegetation index (MOD 13) algorithm theoretical basis document, Version 3, available online at http://modis.gsfc.nasa.gov/data/atbd/atbd_mod13.pdf (last access: 1 June 2012), 1999. 7081

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., and Ferreira, L.: Overview of the radiometric and biophysical performance of the MODIS vegetation indices, Remote Sens. Environ., 83, 195–213, 2002. 7078, 7080, 7081

Humphreys, E., Black, T., Morgenstern, K., Cai, T., Drewitt, G., Nesic, Z., and Trofymow, J.: Carbon dioxide fluxes in coastal Douglas-fir stands at different stages of development after clearcut harvesting, Agr. Forest Meteorol., 140, 6–22, 2006. 7106

Jacobson, A., Fletcher, S., Gruber, N., Sarmiento, J., and Gloor, M.: A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes, Global Biogeochem. Cy., 21, GB1019, doi:10.1029/2006GB002703, 2007a. 7076

Jacobson, A., Mikaloff Fletcher, S., Gruber, N., Sarmiento, J., and Gloor, M.: A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 2. Regional results, Global Biogeochem. Cy., 21, GB1020, doi:10.1029/2006GB002703, 2007b. 7076

Lafleur, P., Roulet, N., Bubier, J., Frolking, S., and Moore, T.: Interannual variability in the peatland-atmosphere carbon dioxide exchange at an ombrotrophic bog, Global Biogeochem. Cy., 17, 1036, doi:10.1029/2002GB001983, 2003. 7106

Lauvaux, T., Schuh, A. E., Uliasz, M., Richardson, S., Miles, N., Andrews, A. E., Sweeney, C., Diaz, L. I., Martins, D., Shepson, P. B., and Davis, K. J.: Constraining the CO_2 budget of the corn belt: exploring uncertainties from the assumptions in a mesoscale inverse system, Atmos. Chem. Phys., 12, 337–354, doi:10.5194/acp-12-337-2012, 2012. 7093

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., Conway, T. J., Doney, S. C., Feely, R. A., Foster, P., Friedlingstein, P., Gurney, K., Houghton, R. A., House, J. I., Huntingford, C., Levy, P. E., Lomas, M. R., Majkut, J., Metzl, N., Ometto, J. P., Peters, G. P., Prentice, I. C., Randerson, J. T., Running, S. W., Sarmiento, J. L., Schuster, U., Sitch, S., Takahashi, T., Viovy, N., van der Werf, G. R., and Woodward, F. I.: Trends in the sources and sinks of carbon dioxide, Nat. Geosci., 2, 831–836, 2009. 7075

Liu, H., Randerson, J., Lindfors, J., and Chapin III, F.: Changes in the surface energy budget after fire in boreal ecosystems of interior Alaska: an annual perspective, J. Geophys. Res, 110, D13101, doi:10.1029/2004JD005158, 2005. 7106

Loveland, T. and Belward, A.: The IGBP-DIS global 1 km land cover data set, DISCover: first results, Int. J. Remote Sens., 18, 3289–3295, 1997. 7080, 7081, 7106

Luo, H., Oechel, W., Hastings, S., Zulueta, R., Qian, Y., and Kwon, H.: Mature semiarid chaparral ecosystems can be a significant sink for atmospheric carbon dioxide, Glob. Change Biol., 13, 386–396, 2007. 7107

Ma, S., Baldocchi, D., Xu, L., and Hehn, T.: Inter-annual variability in carbon dioxide exchange of an oak/grass savanna and open grassland in California, Agr. Forest Meteorol., 147, 157–171, 2007. 7107

Mahadevan, P., Wofsy, S., Matross, D., Xiao, X., Dunn, A., Lin, J., Gerbig, C., Munger, J., Chow, V., and Gottlieb, E.: A satellite-based biosphere parameterization for net ecosystem CO_2 exchange: Vegetation Photosynthesis and Respiration Model (VPRM), Global Biogeochem. Cy., 22, GB2005, doi:10.1029/2006GB002735, 2008. 7078, 7086

McCaughey, J., Pejam, M., Arain, M., and Cameron, D.: Carbon dioxide and energy fluxes from a boreal mixedwood forest ecosystem in Ontario, Canada, Agr. Forest Meteorol., 140, 79–96, 2006. 7106

McGuire, A., Prentice, I., Ramankutty, N., Reichenau, T., Schloss, A., Tian, H., Williams, L., and Wittenberg, U.: Carbon balance of the terrestrial biosphere in the twentieth century: Analyses of CO_2, climate and land use effects with four process-based ecosystem models, Global Biogeochem. Cy., 15, 183–206, 2001. 7076, 7077

Meyers, T. and Hollinger, S.: An assessment of storage terms in the surface energy balance of maize and soybean, Agr. Forest Meteorol., 125, 105–115, 2004. 7106

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res.-Atmos., 109, D14109, doi:10.1029/2003JD004422, 2004. 7077, 7093

Mkhabela, M., Amiro, B., Barr, A., Black, T., Hawthorne, I., Kidston, J., McCaughey, J., Orchansky, A., Nesic, Z., Sass, A., Shashkov, A., and Zha, T.: Comparison of carbon dynamics and water use efficiency following fire and harvesting in Canadian boreal forests, Agr. Forest Meteorol., 149, 783–794, 2009. 7106

Monson, R., Turnipseed, A., Sparks, J., Harley, P., Scott-Denton, L., Sparks, K., and Huxman, T.: Carbon sequestration in a high-elevation, subalpine forest, Glob. Change Biol., 8, 459–478, 2002. 7107

Oechel, W., Vourlitis, G., Hastings, S., Zulueta, R., Hinzman, L., and Kane, D.: Acclimation of ecosystem CO_2 exchange in the Alaskan Arctic in response to decadal climate warming, Nature, 406, 978–981, 2000. 7106

ORNL DAAC: MODIS subsetted land products, Collection 4, available at: http://www.daac.ornl.gov/MODIS/modis.html (last access: 7 December 2007), from ORNL DAAC, Oak Ridge, Tennessee, USA, 2007. 7080

ORNL DAAC: MODIS subsetted land products, Collection 5, available at: http://www.daac.ornl.gov/MODIS/modis.html (last access: 12 November 2010), from ORNL DAAC, Oak Ridge, Tennessee, USA, 2010. 7080

Pacala, S. W., Hurtt, G. C., Baker, D., Peylin, P., Houghton, R. A., Birdsey, R. A., Heath, L., Sundquist, E. T., Stallard, R. F., Ciais, P., Moorcroft, P., Caspersen, J. P., Shevliakova, E., Moore, B., Kohlmaier, G., Holland, E., Gloor, M., Harmon, M. E., Fan, S.-M., Sarmiento, J. L., Goodale, C. L., Schimel, D., and Field, C. B.: Consistent land-and atmosphere-based US carbon sink estimates, Science, 292, 2316–2320, 2001. 7077

Pebesma, E.: Multivariable geostatistics in S: the gstat package, Comput. Geosci., 30, 683–691, 2004. 7095

Peters, W., Miller, J., Whitaker, J., Denning, A., Hirsch, A., Krol, M., Zupanski, D., Bruhwiler, L., and Tans, P.: An ensemble data assimilation system to estimate CO_2 surface fluxes from atmospheric trace gas observations, J. Geophys. Res, 110, D24304, doi:10.1029/2005JD006157, 2005. 7075, 7076, 7077

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, Proc. Natl. Acad. Sci., 104, 18925–18930, 2007. 7075, 7076

Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P.: Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO_2 data, J. Geophys. Res, 107, 4385, doi:10.1029/2001JD000857, 2002. 7077

Powell, T., Bracho, R., Li, J., Dore, S., Hinkle, C., and Drake, B.: Environmental controls over net ecosystem carbon exchange of scrub oak in Central Florida, Agr. Forest Meteorol., 141, 19–34, 2006. 7107

Powell, T., Gholz, H., Clark, K., Starr, G., Cropper, W., and Martin, T.: Carbon exchange of a mature, naturally regenerated pine forest in North Florida, Glob. Change Biol., 14, 2523–2538, 2008. 7107

Prentice, I., Heimann, M., and Sitch, S.: The carbon balance of the terrestrial biosphere: ecosystem models and atmospheric observations, Ecol. Appl., 10, 1553–1573, 2000. 7075

Price, K., Storn, R., and Lampinen, J.: Differential Evolution – A Practical Approach to Global Optimization, Springer, Berlin, Germany, ISBN number: 3540209506, 2006. 7082

R Development Core Team: R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, available at: http://www.R-project.org (last access: 4 June 2012), 2007. 7082, 7094

Raich, J. and Tufekciogul, A.: Vegetation and soil respiration: correlations and controls, Biogeochemistry, 48, 71–90, doi:10.1023/A:1006112000616, 2000. 7086

Rayner, P. J., Enting, I. G., Francey, R. J., and Langenfelds, R.: Reconstructing the recent carbon cycle from atmospheric CO_2, $\delta^{13}C$ and O_2/N_2 observations, Tellus B, 51, 213–232, doi:10.1034/j.1600-0889.1999.t01-1-00008.x, 1999. 7075

Ribeiro Jr., P. and Diggle, P.: geoR: a package for geostatistical analysis, R-NEWS, 1, 14–18, available online: http://CRAN.R-project.org/doc/Rnews/ (last access: 1 June 2012), 2001. 7095

Ricciuto, D.: Diagnosing uncertainty and improving predictions of terrestrial CO_2 fluxes at multiple scales through data assimilation, PhD thesis, The Pennsylvania State University, 2006. 7092

Ricciuto, D., Butler, M., Davis, K., Cook, B., Bakwin, P., Andrews, A., and Teclaw, R.: A Bayesian synthesis inversion of simple respiration and GEP models with eddy covariance data in a northern Wisconsin forest: determining the causes of interannual variability, Agr. Forest Meteorol., 148, 309–327, 2008. 7075, 7081

Richardson, A., Hollinger, D., Burba, G., Davis, K., Flanagan, L., Katul, G., Munger, J., Ricciuto, D., Stoy, P., Suyker, A., Verma, S., and Wofsy, S.: A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, Agr. Forest Meteorol., 136, 1–18, 2006. 7081, 7090, 7091

Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A. M., Papale, D., Reichstein, M., Stauch, V. J., Braswell, B. H., Churkina, G., Kruijt, B., and Hollinger, D. Y.: Statistical properties of random CO_2 flux measurement uncertainty inferred from model residuals, Agr. Forest Meteorol., 148, 38–50, 2008. 7091

Rödenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO_2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport, Atmos. Chem. Phys., 3, 1919–1964, doi:10.5194/acp-3-1919-2003, 2003. 7075, 7076, 7077, 7093

Ruimy, A., Saugier, B., and Dedieu, G.: Methodology for the estimation of terrestrial net primary production from remotely sensed data, J. Geophys. Res., 99, 5263–5283, 1994. 7086

Schmid, H., Grimmond, C., Cropley, F., Offerle, B., and Su, H.: Measurements of CO_2 and energy fluxes over a mixed hardwood forest in the Mid-Western United States, Agr. Forest Meteorol., 103, 357–374, 2000. 7107

Schwalm, C. R., Black, T. A., Amiro, B. D., Arain, M. A., Barr, A. G., Bourque, C. P.-A., Dunn, A. L., Flanagan, L. B., Giasson, M.-A., Lafleur, P. M., Margolis, H. A., McCaughey, J. H., Orchansky, A. L., and Wofsy, S. C.: Photosynthetic light use efficiency of three biomes across an east-west continental-scale transect in Canada, Agr. Forest Meteorol., 140, 269–286, doi:10.1016/j.agrformet.2006.06.010, 2006. 7086

Stoy, P., Katul, G., Siqueira, M., Juang, J., Novick, K., Uebelherr, J., and Oren, R.: An evaluation of models for partitioning eddy covariance-measured net ecosystem exchange into photosynthesis and respiration, Agr. Forest Meteorol., 141, 2–18, 2006. 7106

Strahler, A., Muchoney, D., Borak, J., Friedl, M., Gopal, S., Lambin, E., and Moody, A.: MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD) MODIS Land Cover and Land-Cover Change, USGS, NASA, http://modis.gsfc.nasa.gov/data/atbd/atbd_mod12.pdf (last accessed 4 June 2012), 1999a. 7081

Strahler, A. H., Lucht, W., Schaaf, C. B., Tsang, T., Gao, F., Li, X., Muller, J. P., Lewis, P., and Barnsley, M.: MODIS BRDF/albedo product: algorithm theoretical basis document version 5.0, http://modis.gsfc.nasa.gov/data/atbd/atbd_mod09.pdf (last accessed 4 June 2012), 1999b. 7081

Sulman, B. N., Desai, A. R., Cook, B. D., Saliendra, N., and Mackay, D. S.: The impact of a declining water table on observed carbon fluxes at a northern temperate wetland, Biogeosciences Discuss., 6, 2659–2696, doi:10.5194/bgd-6-2659-2009, 2009. 7107

Syed, K., Flanagan, L., Carlson, P., Glenn, A., and Van Gaalen, K.: Environmental control of net ecosystem CO_2 exchange in a treed, moderately rich fen in Northern Alberta, Agr. Forest Meteorol., 140, 97–114, 2006. 7106

Tans, P. and Conway, T.: Monthly atmospheric CO_2 mixing ratios from the NOAA CMDL Carbon Cycle Cooperative Global Air Sampling Network, 1968–2002, in: Trends: a Compendium of Data on Global Change, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge, Tennessee, USA, available online: http://cdiac.ornl.gov/trends/co2/cmdl-flask/cmdl-flask.html (last access: 4 June 2012), 2005. 7075

Thomas, C. K., Law, B. E., Irvine, J., Martin, J. G., Pettijohn, J. C., and Davis, K. J.: Seasonal hydrology explains interannual and seasonal variation in carbon and water exchange in a semi-arid mature ponderosa pine forest in Central Oregon, J. Geophys. Res, 114, G04006, doi:10.1029/2009JG001010, 2009. 7107

Thornton, P., Law, B., Gholz, H., Clark, K., Falge, E., Ellsworth, D., Goldstein, A., Monson, R., Hollinger, D., Falk, M., Chen, J., and Sparks, J.: Modeling and measuring the effects of disturbance history and climate on carbon and water budgets in evergreen needleleaf forests, Agr. Forest Meteorol., 113, 185–222, 2002. 7091, 7092

Urbanski, S., Barford, C., Wofsy, S., Kucharik, C., Pyle, E., Budney, J., McKain, K., Fitzjarrald, D., Czikowsky, M., and Munger, J.: Factors controlling CO_2 exchange on timescales from hourly to decadal at Harvard Forest, J. Geophys. Res, 112, G02020, doi:10.1029/2006JG000293, 2007. 7106

Verma, S. B., Dobermann, A., Cassman, K. G., Walters, D. T., Knops, J. M., Arkebauer, T. J., Suyker, A. E., Burba, G. G., Amos, B., Yang, H., Ginting, D., Hubbard, K. G., Gitelson, A. A., and Walter-Shea, E. A.: Annual carbon dioxide exchange in irrigated and rainfed maize-based agroecosystems, Agr. Forest Meteorol., 131, 77–96, doi:10.1016/j.agrformet.2005.05.003, 2005. 7107

Wilson, T. and Meyers, T.: Determining vegetation indices from solar and photosynthetically active radiation fluxes, Agr. Forest Meteorol., 144, 160–179, doi:10.1016/j.agrformet.2007.04.001, 2007. 7106

Yuan, W., Liu, S., Zhou, G., Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L. L., Baldocchi, D., Bernhofer, C., Gholz, H., Goldstein, A. H., Goulden, M. L., Hollinger, D. Y., Hu, Y., Law, B. E., Stoy, P. C., Vesala, T., and Wofsy, S. C.: Deriving a light use efficiency model from eddy covariance flux data for predicting daily gross primary production across biomes, Agr. Forest Meteorol., 143, 189–207, doi:10.1016/j.agrformet.2006.12.001, 2007. 7086

Zha, T., Barr, A. G., Black, T. A., McCaughey, J. H., Bhatti, J., Hawthorne, I., Krishnan, P., Kidston, J., Saigusa, N., Shashkov, A., and Nesic, Z.: Carbon sequestration in boreal jack pine stands following harvesting, Glob. Change Biol., 15, 1475–1487, doi:10.1111/j.1365-2486.2008.01817.x, 2009. 7106

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.

Site Code	Site Name	Latitude	Longitude	Land Cover	Reference
CA-Ca1	British Columbia – Campbell River – Mature Forest Site	49.870	−125.340	1 – Evergreen Needleleaf Forest	Humphreys et al. (2006)
CA-Ca2	British Columbia – Campbell River – Clearcut Site	49.870	−125.290	1 – Evergreen Needleleaf Forest	Humphreys et al. (2006)
CA-Ca3	British Columbia – Campbell River – Young Plantation Site	49.520	−124.900	1 – Evergreen Needleleaf Forest	Humphreys et al. (2006)
CA-Gro	Ontario – Groundhog River-Mature Boreal Mixed Wood	48.220	−82.160	5 – Mixed Forest	McCaughey et al. (2006)
CA-Let	Lethbridge	49.710	−112.940	10 – Grasslands	Flanagan et al. (2002)
CA-Mer	Eastern Peatland – Mer Bleue	45.410	−75.520	11 – Permanent Wetlands	Lafleur et al. (2003)
CA-NS2	UCI-1930 burn site	55.910	−98.520	1 – Evergreen Needleleaf Forest	Goulden et al. (2006)
CA-NS3	UCI-1964 burn site	55.910	−98.380	1 – Evergreen Needleleaf Forest	Goulden et al. (2006)
CA-NS4	UCI-1964 burn site wet	55.910	−98.380	1 – Evergreen Needleleaf Forest	Goulden et al. (2006)
CA-NS5	UCI-1981 burn site	55.860	−98.490	1 – Evergreen Needleleaf Forest	Goulden et al. (2006)
CA-NS6	UCI-1989 burn site	55.920	−98.960	1 – Evergreen Needleleaf Forest	Goulden et al. (2006)
CA-NS7	UCI-1998 burn site	56.640	−99.950	1 – Evergreen Needleleaf Forest	Goulden et al. (2006)
CA-Oas	Sask – SSA Old Aspen	53.630	−106.200	4 – Deciduous Broadleaf Forest	Black et al. (2000)
CA-Obs	Sask – SSA Old Black Spruce	53.990	−105.120	1 – Evergreen Needleleaf Forest	Bergeron et al. (2007)
CA-Ojp	Sask – SSA Old Jack Pine	53.920	−104.690	1 – Evergreen Needleleaf Forest	Howard et al. (2004)
CA-Qcu	Quebec Boreal Cutover Site	49.270	−74.040	7 – Open Shrublands	Giasson et al. (2006)
CA-Qfo	Quebec Mature Boreal Forest Site	49.690	−74.340	1 – Evergreen Needleleaf Forest	Bergeron et al. (2007)
CA-SF2	Sask – Fire 1989	54.250	−105.880	6 – Closed Shrublands	Mkhabela et al. (2009)
CA-SF3	Sask – Fire 1998	54.090	−106.010	6 – Closed Shrublands	Mkhabela et al. (2009)
CA-SJ1	Sask – 1994 Harv. Jack Pine	53.910	−104.660	1 – Evergreen Needleleaf Forest	Zha et al. (2009)
CA-SJ2	Sask – 2002 Harvested Jack Pine	53.950	−104.650	1 – Evergreen Needleleaf Forest	Zha et al. (2009)
CA-WP1	Western Peatland – LaBiche-Black Spruce/Larch Fen	54.960	−112.460	11 – Permanent Wetlands	Syed et al. (2006)
US-ARM	ARM Southern Great Plains site – Lamont – Oklahoma	36.610	−97.490	12 – Croplands	Fischer et al. (2007)
US-Atq	Atqasuk – Alaska	70.470	−157.410	11 – Permanent Wetlands	Oechel et al. (2000)
US-Aud	Audubon Research Ranch – Arizona	31.590	−110.510	10 – Grasslands	Wilson and Meyers (2007)
US-Blo	Blodgett Forest – California	38.900	−120.630	1 – Evergreen Needleleaf Forest	Goldstein et al. (2000)
US-Bn1	Delta Junction 1920 Control site	63.920	−145.370	1 – Evergreen Needleleaf Forest	Liu et al. (2005)
US-Bn2	Delta Junction 1987 Burn site	63.920	−145.370	4 – Deciduous Broadleaf Forest	Liu et al. (2005)
US-Bn3	Delta Junction 1999 Burn site	63.920	−145.740	7 – Open Shrublands	Liu et al. (2005)
US-Bo1	Bondville – Illinois	40.010	−88.290	12 – Croplands	Meyers and Hollinger (2004)
US-Bo2	Bondville – Illinois (companion site)	40.010	−88.290	12 – Croplands	Meyers and Hollinger (2004)
US-Brw	Barrow – Alaska	71.320	−156.630	11 – Permanent Wetlands	Harazono et al. (2003)
US-CaV	Canaan Valley – West Virginia	39.060	−79.420	10 – Grasslands	Wilson and Meyers (2007)
US-Dk1	Duke Forest-open field – North Carolina	35.970	−79.090	10 – Grasslands	Stoy et al. (2006)
US-Dk2	Duke Forest-hardwoods – North Carolina	35.970	−79.100	4 – Deciduous Broadleaf Forest	Stoy et al. (2006)
US-Dk3	Duke Forest – loblolly pine – North Carolina	35.980	−79.090	1 – Evergreen Needleleaf Forest	Stoy et al. (2006)
US-FPe	Fort Peck – Montana	48.310	−105.100	10 – Grasslands	Wilson and Meyers (2007)
US-Goo	Goodwin Creek – Mississippi	34.250	−89.970	10 – Grasslands	Wilson and Meyers (2007)
US-Ha1	Harvard Forest EMS Tower – Massachusetts (HFR1)	42.540	−72.170	4 – Deciduous Broadleaf Forest	Urbanski et al. (2007)
US-Ha2	Harvard Forest Hemlock Site – Massachusetts	42.540	−72.170	1 – Evergreen Needleleaf Forest	Hadley and Schedlbauer (2002)

Table 1. Continued.

Site Code	Site Name	Latitude	Longitude	Land Cover	Reference
US-Ho1	Howland Forest (main tower) – Maine	45.200	−68.740	1 – Evergreen Needleleaf Forest	Hollinger et al. (1999)
US-Ho2	Howland Forest (west tower) – Maine	45.210	−68.750	1 – Evergreen Needleleaf Forest	Hollinger et al. (2004)
US-KS1	Florida-Kennedy Space Center (slash pine)	28.460	−80.670	1 – Evergreen Needleleaf Forest	Bracho et al. (2008)
US-KS2	Florida-Kennedy Space Center (scrub oak)	28.610	−80.670	6 – Closed Shrublands	Powell et al. (2006)
US-Los	Lost Creek – Wisconsin	46.080	−89.980	6 – Closed Shrublands	Sulman et al. (2009)
US-Me2	Metolius-intermediate aged ponderosa pine – Oregon	44.450	−121.560	1 – Evergreen Needleleaf Forest	Thomas et al. (2009)
US-Me4	Metolius-old aged ponderosa pine – Oregon	44.500	−121.620	1 – Evergreen Needleleaf Forest	Anthoni et al. (2002)
US-MMS	Morgan Monroe State Forest – Indiana	39.320	−86.410	4 – Deciduous Broadleaf Forest	Schmid et al. (2000)
US-MOz	Missouri Ozark Site	38.740	−92.200	4 – Deciduous Broadleaf Forest	Gu et al. (2006)
US-Ne1	Mead – irrigated continuous maize site – Nebraska	41.100	−96.290	12 – Croplands	Verma et al. (2005)
US-Ne2	Mead – irrigated maize-soybean rotation site – Nebraska	41.100	−96.280	12 – Croplands	Verma et al. (2005)
US-Ne3	Mead – rainfed maize-soybean rotation site – Nebraska	41.180	−96.440	12 – Croplands	Verma et al. (2005)
US-NR1	Niwot Ridge Forest – Colorado (LTER NWT1)	40.030	−105.550	1 – Evergreen Needleleaf Forest	Monson et al. (2002)
US-PFa	Park Falls/WLEF – Wisconsin	45.950	−90.270	5 – Mixed Forest	Davis et al. (2003)
US-SO2	Sky Oaks- Old Stand – California	33.370	−116.620	6 – Closed Shrublands	Luo et al. (2007)
US-SO3	Sky Oaks- Young Stand – California	33.380	−116.620	6 – Closed Shrublands	Luo et al. (2007)
US-SO4	Sky Oaks- California	33.370	−116.620	6 – Closed Shrublands	Luo et al. (2007)
US-SP1	Slashpine-Austin Cary – 65yrs nat regen-FL	29.740	−82.220	1 – Evergreen Needleleaf Forest	Powell et al. (2008)
US-SP2	Slashpine-Mize-clearcut – 3yr-regen-FL	29.760	−82.240	1 – Evergreen Needleleaf Forest	Bracho et al. (2012)
US-SP3	Slashpine-Donaldson-mid-rot – 12yrs-FL	29.750	−82.160	1 – Evergreen Needleleaf Forest	Bracho et al. (2012)
US-Syv	Sylvania Wilderness Area – Michigan	46.240	−89.350	5 – Mixed Forest	Desai et al. (2005)
US-Ton	Tonzi Ranch – California	38.430	−120.970	8 – Woody Savannas	Ma et al. (2007)
US-UMB	Univ. of Mich. Biological Station – Michigan	45.560	−84.710	4 – Deciduous Broadleaf Forest	Gough et al. (2008)
US-Var	Vaira Ranch – Ione – California	38.410	−120.950	10 – Grasslands	Ma et al. (2007)
US-WCr	Willow Creek – Wisconsin	45.810	−90.080	4 – Deciduous Broadleaf Forest	Cook et al. (2004)

Table 2. Total number of parameters resulting from the nine different schemes used to group observation sites for VPRM parameter estimation.

	site groupings in time		
	monthly intervals	annual intervals	all available data, 2000–2006
site groupings in space			
individual sites (65)	21 840	1820	260
PFTs (9)	3360	280	40
all sites together	336	28	4

Table 3. Range parameter values (km) for VPRM flux error best-fit parametric variogram models. VPRM parameterizations are described in Table 2. Where a value is present, Akaike's information criterion (AIC) analysis concludes the exponential variogram model fit more parsimoniously than the pure nugget model. Where the range is blank the pure nugget model fit most parsimoniously, indicating no spatial correlation is present.

season	VPRM Parameterization		Year						
	space	time	2000	2001	2002	2003	2004	2005	2006
DJF	site	monthly		1640	3051	4	1518		605
	site	annual	1299		368	772	3787		
	site	all data			403	2382	115		
	PFT	monthly							
	PFT	annual				249	301		48
	PFT	all data				258	460		76
	all sites	monthly							509
	all sites	annual				1			450
	all sites	all data				1			461
MAM	site	monthly				285			
	site	annual				207			
	site	all data							
	PFT	monthly							
	PFT	annual				748			
	PFT	all data				769			
	all sites	monthly				4			
	all sites	annual				5			
	all sites	all data				6			
JJA	site	monthly					4130	1617	24
	site	annual		267	631		1354		
	site	all data	34	405	1	368	316		
	PFT	monthly							
	PFT	annual	30	461		323			
	PFT	all data	734	407		287			8
	all sites	monthly		401	537	289			
	all sites	annual		401	534	292			
	all sites	all data		404	534	296			
SON	site	monthly				29	54		2059
	site	annual	227	314	840		4600		129
	site	all data			1004		2		
	PFT	monthly		2689		710			56
	PFT	annual		406	1046	1466	70		77
	PFT	all data			1664	1183	89		35
	all sites	monthly		787		2199			20
	all sites	annual		746	1032	2288	0		16
	all sites	all data		649	1041	2086	0		19

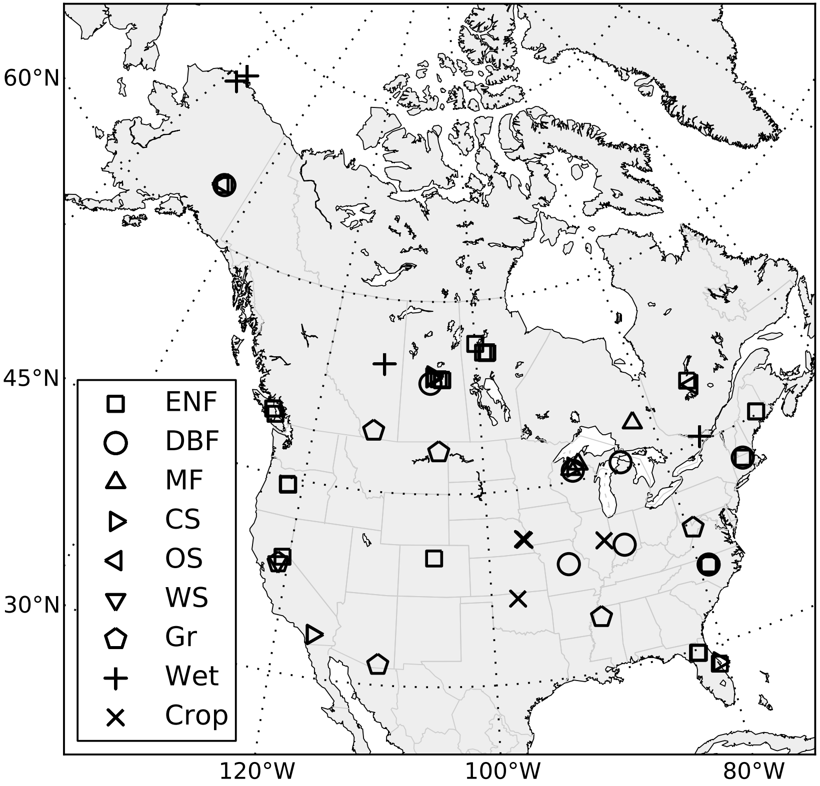

Fig. 1. The 65 eddy covariance flux tower sites from the Fluxnet network (http://www.fluxdata.org) that provide observations for VPRM parametrization and VPRM flux residual calculation. ENF: evergreen needleleaf forest, DBF: deciduous broadleaf forest, MF: mixed forest, CS: closed shrubland, OS: open shrubland, WS: woody savanna, Gr: grassland, Wet: permanent wetland, Crop: cropland.

BGD

9, 7073–7116, 2012

Spatial structure in land surface model residuals

T. W. Hilton et al.

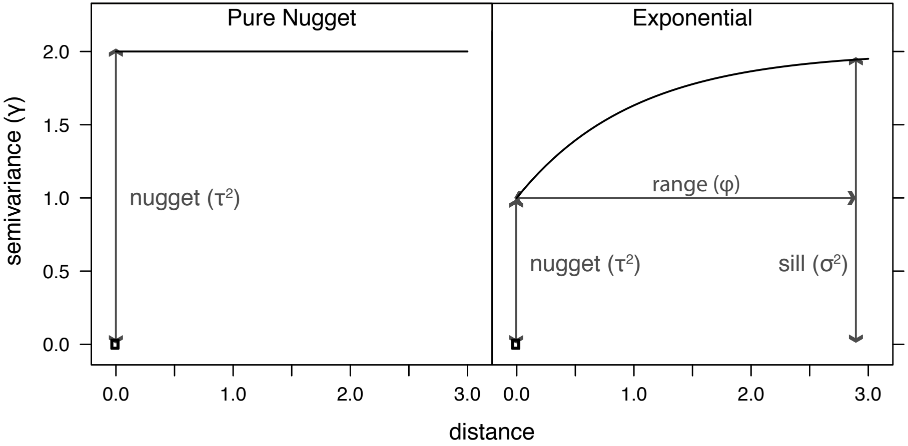

Fig. 2. Two examples of generic parametric variogram models. The parameter symbols correspond to Sect. 2.4 and Eq. (8). Because these are purely illustrative, units for semivariance and distance are irrelevant.

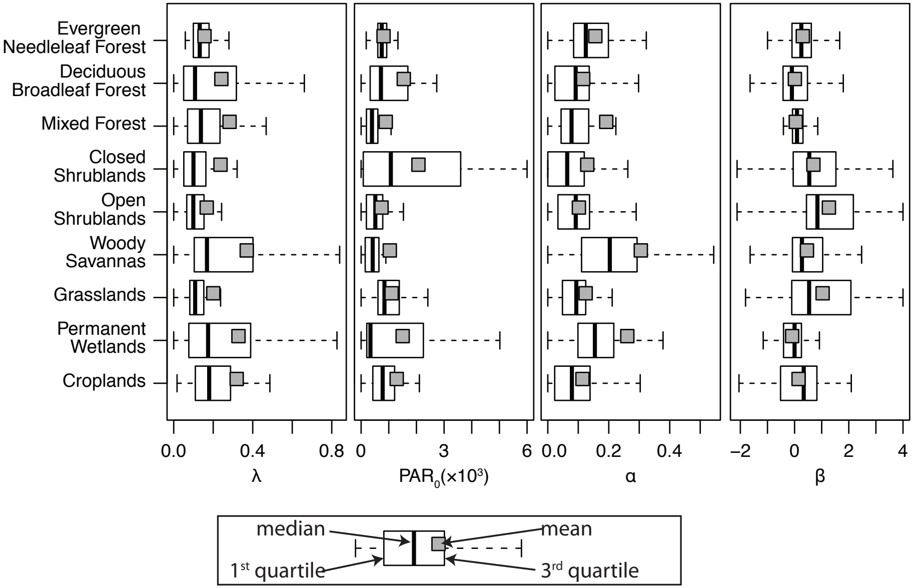

Fig. 3. Box and whisker plots for values of VPRM parameters, estimated monthly by plant functional type (PFT). Whiskers show 1.5 times the interquartile range. Units for parameters are: λ: μmol CO$_2$ m^{-2} s^{-1}/μmol PAR m^{-2} s^{-1}; α: μmol CO$_2$ m^{-2} s^{-1}°C^{-1}; β: μmol CO$_2$ m^{-2} s^{-1}; PAR$_0$: μmol PAR m^{-2} s^{-1}.

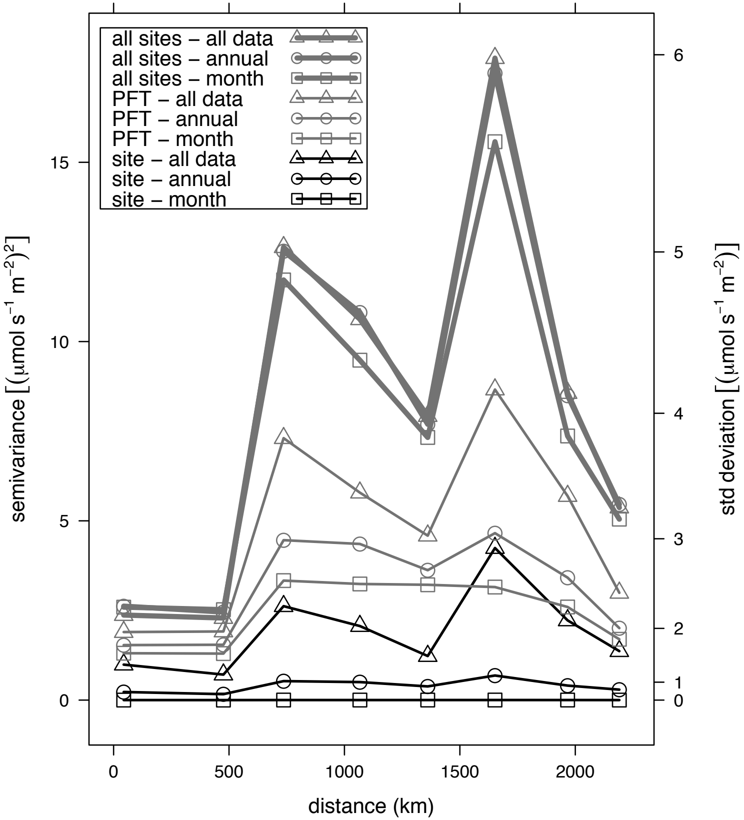

Fig. 4. June-July-August mean VPRM NEE residual empirical semivariograms. Each point represents the mean semivariance and mean separation distance from grouping pairs of towers into 300 km bins. VPRM parametrizations are described in Table 2. The left vertical axis shows units of semivariance ($\hat{\gamma}$), and the right vertical axis shows units of standard deviation (σ). σ is related to $\hat{\gamma}$ by $\sigma = (2\hat{\gamma})^{1/2}$.

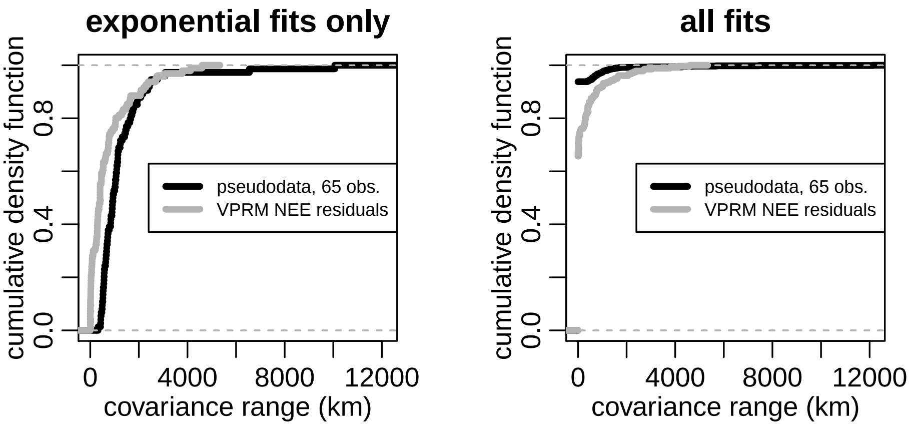

Fig. 5. Cumulative density functions for covariance range parameter, estimated by maximum likelihood estimation (MLE). Each panel shows range values for seasonal VPRM NEE residuals (values in Table 3) along with values from 1000 Gaussian random fields (GRFs) with similar (and known) covariance structure. The left panel shows the range distribution for only the pseudodata range estimations where the exponential covariance model fit more optimally than the pure nugget; the right panel shows the distribution for all 1000 GRFs.

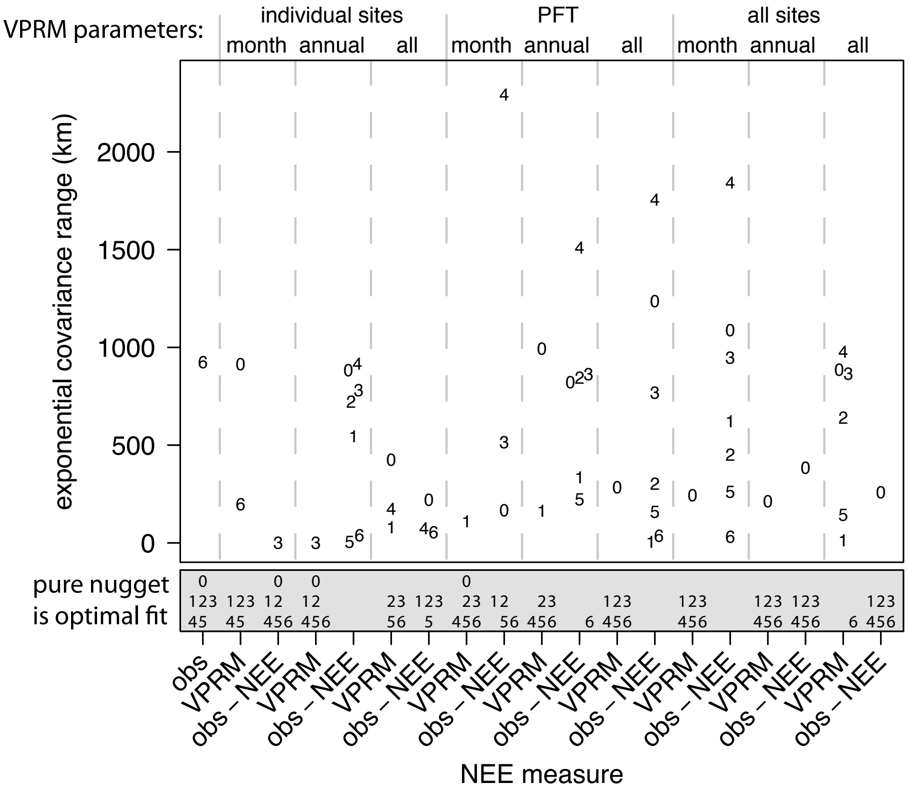

Fig. 6. Best-fit range values (km) for cumulative annual anomalies of observed NEE, VPRM NEE, and VPRM NEE residuals. Best-fit values were determined by AIC as described in Sect. 2.4. The number y plotted denotes the year 200y. Years where the pure nugget covariance function fit more optimally than the exponential are shown in the shaded box. Anomalies were calculated as the departure from the mean value of 2000 to 2006 annual mean cumulative observed values.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| | |▶
◀ | ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

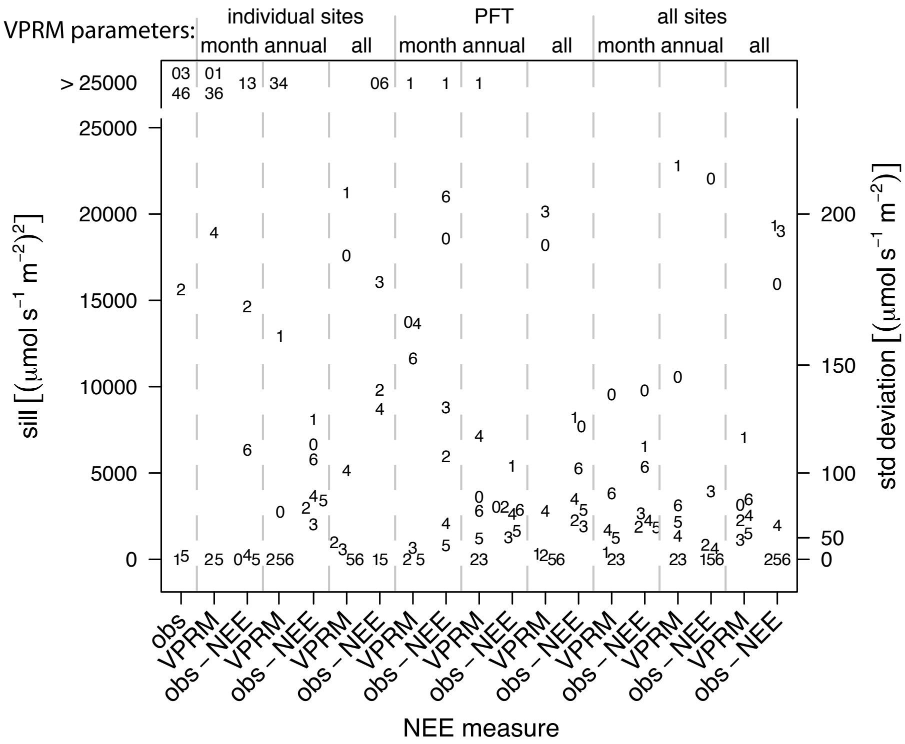

Fig. 7. As Fig. 6, but displaying semivariogram sill values.