Statistical properties of random CO$_2$ flux measurement uncertainty inferred from model residuals

Andrew D. Richardson$^a$,*, Miguel D. Mahecha$^b$, Eva Falge$^c$, Jens Kattge$^b$, Antje M. Moffat$^b$, Dario Papale$^d$, Markus Reichstein$^b$, Vanessa J. Stauch$^e$, Bobby H. Braswell$^a$, Galina Churkin$^b$, Bart Kruijt$^f$, David Y. Hollinger$^g$

$^a$University of New Hampshire, Complex Systems Research Center, Morse Hall, 39 College Road, Durham, NH 03824, USA
$^b$Max Planck Institute for Biogeochemistry, Hans-Knöll-Strasse 10, 07745 Jena, Germany
$^c$Max Planck Institute for Chemistry, Biogeochemistry Department, J.J.v. Becherweg 27, 55128 Mainz, Germany
$^d$DISAFRI, University of Tuscia, via C. de Lellis, 01100 Viterbo, Italy
$^e$Computational Landscape Ecology, Centre for Environmental Research (UFZ), Leipzig, Germany
$^f$Alterra, Wageningen University and Research Centre, 6700 AA Wageningen, The Netherlands
$^g$USDA Forest Service, Northern Research Station, Durham, NH 03824, USA

**Abstract**

Information about the uncertainties associated with eddy covariance measurements of surface-atmosphere CO$_2$ exchange is needed for data assimilation and inverse analyses to estimate model parameters, validation of ecosystem models against flux data, as well as multi-site synthesis activities (e.g., regional to continental integration) and policy decision-making. While model residuals (mismatch between fitted model predictions and measured fluxes) can potentially be analyzed to infer data uncertainties, the resulting uncertainty estimates may be sensitive to the particular model chosen. Here we use 10 site-years of data from the CarboEurope program, and compare the statistical properties of the inferred random flux measurement error calculated first using residuals from five different models, and secondly using paired observations made under similar environmental conditions. Spectral analysis of the model predictions indicated greater persistence (i.e., autocorrelation or “memory”) compared to the measured values. Model residuals exhibited weaker temporal correlation, but were not uncorrelated white noise. Random flux measurement uncertainty, expressed as a standard deviation, was found to vary predictably in relation to the expected magnitude of the flux, in a manner that was nearly identical (for negative, but not positive, fluxes) to that reported previously for forested sites. Uncertainty estimates were generally comparable whether the uncertainty was inferred from model residuals or paired observations, although the latter approach resulted in somewhat smaller estimates. Higher order moments (e.g., skewness and kurtosis) suggested that for fluxes close to zero, the measurement error is commonly skewed and leptokurtic. Skewness could not be evaluated using the paired observation approach, because differencing of paired measurements resulted in a symmetric distribution of the inferred error. Patterns were robust and not especially sensitive to the model used, although more flexible models, which did not impose a particular functional form on relationships between environmental drivers and modeled fluxes, appeared to give the best results. We conclude that evaluation of flux measurement errors from model residuals is a viable alternative to the standard paired observation approach.

© 2007 Elsevier B.V. All rights reserved.
1. Introduction

The eddy covariance method is used at research sites around the world to measure surface-atmosphere exchanges of carbon, water and energy (Baldocchi et al., 2001b). Increasingly, these flux data are being used to constrain large-scale carbon cycle models (e.g., Papale and Valentini, 2003; Friend et al., 2006). In the context of this so-called “model-data synthesis”, Raupach et al. (2005) have argued the importance of accurately characterizing data uncertainties, suggesting that uncertainties are as important as the data values themselves. This is due to the fundamental role of uncertainties in specifying an objective criterion (commonly referred to as the “cost” function; see Raupach et al., 2005), quantifying the mismatch or “distance” between model and data, used as the basis for optimization or synthesis (Trudinger et al., 2007).

If the “true” flux is $F_i$, and the measurement error is a random variable $\delta_i$ (the measured flux is thus $F_i = F_i + \delta_i$), then a convenient way to characterize the uncertainty is $\sigma(\delta_i)$, i.e., as the standard deviation of the random error. Accordingly, $1/\sigma(\delta_i)$ is a quantitative measure of our confidence in the measured data (see Raupach et al., 2005). In conducting model-data syntheses, data values are typically weighted in the cost function by the reciprocal of their uncertainties, i.e., as $1/\sigma(\delta_i)$ (possibly raised to a higher power). An important consequence of this is that estimated model parameters or states depend on both measurements and uncertainties.

There is a range of other applications in which knowledge of flux data uncertainties is also critical. These include, for example, ecosystem model validation against flux data (e.g., Thornton et al., 2002; Churkina et al., 2003), because uncertainty estimates are required to make statistically rigorous comparisons between measurements and model predictions. Additionally, flux data uncertainties are needed for multi-site syntheses (e.g., Law et al., 2002; Churkina et al., 2005) and regional-to-continent integration efforts, especially with regard to carbon accounting and policy decision-making (Wofsy and Harriss, 2002). In both of these latter examples, sites with smaller data uncertainties can be accorded more influence than sites with larger uncertainties.

Flux measurement errors can be broadly classified into random and systematic sources of uncertainty; whereas statistical analyses can be used to estimate the size of random errors, systematic errors can be difficult to detect or quantify in data-based analyses (Moncrieff et al., 1996; but see Lee, 1998). Random flux measurement uncertainties (e.g., Hollinger and Richardson, 2005), which are stochastic, represent the aggregate uncertainty due to surface heterogeneity and a time-varying flux footprint, turbulence sampling errors (especially problematic over tall vegetation), and errors associated with the measurement equipment (e.g., gas analyzer and sonic anemometer). Systematic errors, on the other hand, operate at varying time scales (e.g., fully systematic vs. selectively systematic, see Moncrieff et al., 1996), and can influence measurements in a variety of ways (e.g., fixed offsets vs. relative offsets and positive vs. negative biases). Systematic errors due to the imperfect spectral response of the measurement system, nocturnal biases resulting from inadequate mixing, issues related to advection and non-flat terrain, and problems of energy balance closure, are discussed elsewhere (Goulden et al., 1996; Moncrieff et al., 1996; Lee, 1998; Massman and Lee, 2002; Wilson et al., 2002; Aubinet et al., 2005; Loescher et al., 2006).

While any attempt at quantifying the total flux measurement uncertainty must account for both random and systematic errors, our goal here is to build on previous statistically-based analytical efforts to quantify the random component. For example, Hollinger et al. (2004) differenced paired measurements from two flux towers in an effort to estimate the random flux measurement uncertainty, which was then characterized, as $\sigma(\delta)$, by the standard deviation of the differenced observations (divided by the square root of two). The two towers were located in similar old-growth Picea stands and instrumented identically, but were separated by $\approx 775$ m and had non-overlapping footprints and largely independent turbulence (cf. the more recent study by Rannik et al., 2006, where the $\approx 30$ m separation of two towers resulted in paired measurements that were not fully independent).

Results of the above studies were in close agreement with (but somewhat higher than) predictions of the Lenschow et al. (1994) and Mann and Lenschow (1994) error model based on turbulence statistics.

In a subsequent paper, Hollinger and Richardson (2005) proposed that paired measurements at one tower, separated by 24 h and made under “similar” environmental conditions could also be used to estimate the random flux measurement uncertainty; this method was applied by Richardson et al. (2006a) to data from seven sites (one grassland, an agricultural site, and five forested sites) in the first multi-site synthesis of random measurement uncertainty. The one-tower approach may result in lower uncertainty estimates compared to the two-tower approach because spatial heterogeneity across an ecosystem is likely to be larger than that across the time-varying footprint of a single tower (see Katul et al., 1999; Oren et al., 2006). Conversely, however, imperfect matching (between days) in environmental and turbulent conditions may result in slight over-estimation of uncertainty using the one-tower approach.

We consider here a third method to quantify eddy flux random measurement uncertainty. In the hypothetical instance of a perfect model and no systematic measurement errors, then the concordance between model predictions and measured data is ultimately limited by the random flux measurement uncertainty. Thus an alternative means by which the random flux measurement uncertainty might be evaluated is through posterior analysis of residuals from a fitted model (Stauch et al., in press). However, models are an imperfect representation of reality, and thus a disadvantage of this approach is that the resulting uncertainty estimates will be a combination of both measurement error and model error. With a particularly poor model (inappropriate functional form, mis-representation of processes and driving variables, or suboptimal parameter values), model error could potentially overwhelm the random measurement error. While estimates of measurement error inferred from model residuals will depend somewhat on the model itself, there is some evidence that this may be a reasonable approach to quantifying uncertainties. For example, Richardson and Hollinger (2005) reported that the standard deviation of model residuals for the Lloyd and Taylor (1994) respiration model, fit to nocturnal
data, was only 14% larger than the random flux measurement uncertainty estimated using paired measurements, suggesting that model error is potentially relatively small. In that study, the statistical properties of the residuals were otherwise similar to the uncertainty estimated using paired measurements, i.e., non-Gaussian and leptokurtic (see also Hagen et al., 2006), with comparable patterns in relation to wind speed and flux magnitude.

A number of other studies have also conducted analyses of model residuals for the purpose of evaluating flux measurement uncertainty. Using data from a range of CO2 flux measurement sites, Chevallier et al. (2006) noted that at the daily time step, ORCHIDEE model residuals were non-Gaussian. At least two related studies have purposely avoided making assumptions about the specific probability distribution function for the random flux measurement uncertainty. For example, Hagen et al. (2006) used bootstrap resampling of model residuals to estimate uncertainties in gross productivity at different time scales (from 30 min. to annual), whereas Stauch et al. (in press) used a stochastic model to partition “signal” and “noise” in measured net fluxes, and characterized the noise (i.e., model residuals) using a non-parametric statistical model based on time-varying cubic splines (As noted by both of these authors, annual flux sums are approximately Gaussian, despite the non-Gaussian uncertainties in the 30 min. data; this result follows directly from the Central Limit Theorem).

In the present study, we use a wide array of data and models to investigate whether an approach based on model residuals results in uncertainty estimates that are comparable to those derived from paired observations. We optimize model parameters (on a site-by-site basis) to maximize the agreement between model and measurements, and thus obtain model predictions that are informed by the data, rather than independent of the data. In this manner the models can be used as tools to predict the “true” flux, \( F^\tau \), conditional on the measured fluxes, environmental drivers, and implicit (in the model) relationships between these two.

To evaluate the degree to which model-based estimates of uncertainty are influenced by the particular model chosen (i.e., the degree to which model error confounds random measurement error), we compare results from five different models, ranging from simple nonlinear regressions to more highly parameterized artificial neural networks and a process-based ecosystem model. Some of the models (e.g., neural network and look-up table) contain few, if any, assumptions about the underlying functional form of relationships between CO2 flux and environmental drivers. Parameterization of some of the models (look-up table and nonlinear regression) was time-varying to account for seasonal changes in the flux-driver relationships.

We begin by evaluating, first in the temporal domain and then in the frequency domain, model predictions and model residuals in relation to the measured CO2 flux time series (i.e., the net ecosystem exchange, NEE). By using new tools (improvements over traditional Fourier power analysis) to probe the spectral properties of both the measured fluxes and the inferred random error, we are able to investigate and characterize the correlation structure inherent in each of these time series. This analysis complements the more traditional analysis which follows, where we focus on the moments (standard deviation, skewness and kurtosis) of the probability distribution of the inferred random error. We compare inferred uncertainties calculated from paired observations and model residuals, and look for similarities across sites. Finally, relationships between the standard deviation of the random flux measurement uncertainty and flux magnitude developed here for six forested European sites are compared to those presented previously by Richardson et al. (2006a) for four forested AmeriFlux sites.

### 2. Data and method

The present analysis is based on 10 site-years of data from the CarboEurope database. The sites (see Table 1) span a range of forest types, including deciduous (FR1, DE3), coniferous (FI1), mixed deciduous/coniferous (BE1), and Mediterranean (FR4, IT3). The data were assembled as part of a data set used for evaluation of a standardized data processing algorithm for half-hourly data (Papale et al., 2006), a comprehensive gap filling comparison (Moffat et al., 2007), and an evaluation of different algorithms for partitioning NEE to its component fluxes, gross productivity and ecosystem respiration (Desai et al., unpublished). Data were corrected for canopy storage but not energy balance closure; quality control procedures, \( u^* \) filtering, and gap filling of meteorological data are discussed in the companion paper by Moffat et al. (2007). Measured CO2 fluxes are here denoted \( F_c \), with units of \( \mu \text{mol m}^{-2} \text{s}^{-1} \). The sign convention is that a negative flux indicates carbon uptake.

### Table 1 – Sites and years of data used for the uncertainty analysis. MAT = mean annual temperature

<table>
<thead>
<tr>
<th>Site</th>
<th>Species</th>
<th>Climate (MAT)</th>
<th>Years</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE1 (Vielsalm, Belgium)</td>
<td>Fagus sylvatica, Pseudotsuga menziesii</td>
<td>Temperate-continental (7.5 °C)</td>
<td>2000, 2001</td>
<td>50.30° N</td>
<td>9.58° E</td>
<td>Aubinet et al. (2001)</td>
</tr>
<tr>
<td>FI1 (Hyytiala, Finland)</td>
<td>Pinus sylvestris</td>
<td>Boreal (3.8 °C)</td>
<td>2001, 2002</td>
<td>61.83° N</td>
<td>24.28° E</td>
<td>Suni et al. (2003)</td>
</tr>
<tr>
<td>FR1 (Hesse, France)</td>
<td>Fagus sylvatica</td>
<td>Temperate-suboceanic (9.9 °C)</td>
<td>2001, 2002</td>
<td>48.67° N</td>
<td>7.05° E</td>
<td>Granier et al. (2000)</td>
</tr>
<tr>
<td>FR4 (Puechabon, France)</td>
<td>Quercus ilex</td>
<td>Mediterranean (13.5 °C)</td>
<td>2002</td>
<td>43.73° N</td>
<td>3.58° E</td>
<td>Tedeschi et al. (2004)</td>
</tr>
<tr>
<td>IT3 (Roccasrespampani, Italy)</td>
<td>Quercus cerris</td>
<td>Mediterranean (15.2 °C)</td>
<td>2002</td>
<td>42.40° N</td>
<td>11.92° E</td>
<td>Tedeschi et al. (2006)</td>
</tr>
</tbody>
</table>
by the ecosystem, whereas a positive flux indicates carbon release.

For each site-year, fitted model predictions from five different models selected from among those submitted to the gap filling comparison were used. The models (described fully by Moffat et al., 2007) were as follows: an artificial neural network model (ANN), a non-linear regression model (NLR), a process-based ecosystem model (BETYH), a semi-parametric model (SPM), and a marginal distribution sampling method (MDS). The ANN (Papale and Valentini, 2003) is essentially a multi-stage, empirical, non-linear regression model. An advantage of neural networks is that no a priori decisions about the functional form of relationships between data inputs (e.g., meteorological drivers and time of day/year as a fuzzy variable) and outputs (measured fluxes) are required. The particular NLR used here (Falge et al., 2001) was based on a Michaelis–Menten light response function for gross photosynthesis and a Lloyd and Taylor (1994) temperature function for ecosystem respiration; model parameters were fit twice-monthly. BETHY (Knorr and Katte, 2005) is a terrestrial biosphere model that predicts CO2, water and energy fluxes. In addition to meteorological drivers, required input data include site-specific information about soil and canopy properties (e.g., leaf area index, estimated from remote sensing products). SPM (Stauch and Jarvis, 2006) is analogous to a multi-dimensional look-up table, with an underlying semi-parametric interpolation defined by three-dimensional cubic splines (time-varying functions of PPFD and temperature). For MDS (Reichstein et al., 2005), which is essentially a moving look-up table, fluxes are averaged from similar meteorological conditions (global radiation, air temperature and vapor pressure deficit within pre-specified criteria) across time-windows as small as possible. Thus the models ranged from data based (e.g., ANN) to process based (e.g., BETHY), and the adaptation to measurements ranged from time constant (e.g., BETHY) to time varying parameterization (e.g., MDS). While our goal here is not to identify any one of these as the “best” model, the application of objective model selection criteria (e.g., Akaike’s Information Criterion) to similar models of CO2 exchange data has been described previously (Richardson et al., 2006b).

2.1. Evaluation of spectral properties

The spectral properties (i.e., correlation structure) of a time series can be summarized by calculating the scaling exponent, $\alpha$, of the power spectrum $P(f) = f^\alpha$ (Mandelbrot, 1999). Here $f$ refers to the frequencies contained in a given time series, and $\alpha$ is conventionally estimated as the slope of the log-log transformed relationship between frequency and power (i.e., the Fourier power spectrum). Uncorrelated white noise has $\alpha = 0$ (equal power across all frequencies), whereas a strong autocorrelation structure, such as characterizes Brownian motion, has $\alpha \approx -2$. The power spectrum of this “red” noise is dominated by the lower frequencies, which impart the signal with substantial “memory.” Intermediate between white and red noise is “pink” noise, for which the correlation structure is moderate and $\alpha \approx -1$.

With the goal of evaluating whether the different models were able to reproduce the temporal correlation structure of the measured NEE time series, and to characterize the spectral properties of the model residuals, we used the Lomb-Scargle periodogram (“L-S”, Lomb, 1976; Scargle, 1982) and the multiple segmentation method (“MSM”, Miramontes and Rohani, 2002; Rohani et al., 2004) to estimate $\alpha$ for the measured fluxes, model predictions, and model residuals. The L-S method (following Press et al., 2002) estimates the equivalent of a Fourier power spectrum for unevenly-spaced time series data, and is thus suited to analysis of fragmented time series with up to $\approx 20\%$ missing observations; $\alpha$ is calculated from the resulting L-S power spectrum using the conventional method.

Since a reasonable estimate of the scaling exponent $\alpha$ is only possible for very long time series when standard methods are applied, new methods (e.g., MSM) have been developed to yield statistically robust estimates of $\alpha$ for short time series. The MSM method subdivides a time series into segments, with segment lengths defined by powers of two. The scaling exponent for each segment is calculated as the slope of the power spectrum of the sub-series. An overall estimate of $\alpha$ is obtained by fitting a linear model to the relationship between segment length and the mean of the scaling exponents found at each segment size. In the present study the power spectra of the sub-series in MSM were determined using the L-S method. With this hybrid approach that blends MSM and L-S, we could estimate $\alpha$ not only for short, but also for short and fragmented time series. Thus we were able to estimate the scaling exponent for measured fluxes, without having to resort to gap-filled time series.

2.2. Inferred random flux measurement error

For each site-year of data, the paired observation approach to estimating uncertainty was implemented as described in Richardson et al. (2006a). Briefly, this involved considering as a pair any two flux measurements made exactly 24 h apart, with mean half-hourly photosynthetically active photon flux densities within 75 $\mu$mol m$^{-2}$ s$^{-1}$, air temperatures within 3 $^\circ$C, and vapor pressure deficits within 0.2 kPa. Across sites, between 19% and 27% (overall mean, 23%) of half-hourly measurements met these criteria. For this method, the inferred error, $\delta_i$, equals the difference between the two measurements divided by the square root of two (see also Hollinger and Richardson, 2005).

The model-based approach to estimating uncertainty was implemented using the five models named above. The inferred random error was taken to be the residual, $\epsilon_i$, between the predicted value of the fitted model and the measured flux. With data coverage averaging 69% across all sites, this method provided roughly three-fold as many estimates of the inferred random error, compared to the paired observation approach.

The statistical properties (i.e., moments of the distribution) of the inferred random error were analyzed with respect to flux magnitude. Based on predictions of the ANN model (the standard deviation of the residuals was smallest for this model), observations were grouped into bins of comparable flux magnitude (bin width = 5 $\mu$mol m$^{-2}$ s$^{-1}$) and the mean, standard deviation, skewness, and kurtosis of the inferred random error calculated for each bin; as in the past, we
characterize the random flux measurement uncertainty by the second moment of the inferred random error, i.e., as the standard deviation, \( \sigma(d) \). Moments were only calculated for bins containing at least 10 observations.

### 3. Results

#### 3.1. Time series of model residuals

Model residuals were well-correlated across models, with pairwise linear correlations generally \( \geq 0.80 \), suggesting that the dominant source of variability in residuals is related to flux measurement errors (which would affect residual time series from all models similarly), rather than model error. To use the FR1 2001 data as an example, MDS and BETHY residuals were more weakly correlated (\( r = 0.80 \)) than, for example, ANN and SPM residuals (\( r = 0.94 \)). Visual inspection of the raw residual data (Fig. 1) suggested only modest differences among models, but when the mean bias was calculated over weekly periods, it became apparent that predictions of some models (e.g., BETHY) were consistently biased at certain times of the year, whereas other models (e.g., MDS) exhibited little or no persistent bias.

#### 3.2. Spectral properties of measured fluxes, model predictions, and model residuals

The Lomb-Scargle periodogram suggested that all models did a reasonable job at describing the higher-frequency components of the data, but due to time constant parameterization, some models (e.g., BETHY) were less successful at describing the lower-frequency components than other models (e.g., MDS) that used time-varying parameterization. All residual time series exhibited a spectral peak at frequency of 1/day. This does not necessarily mean that the models fail to adequately capture the diurnal cycle; rather, it just means that there is an underlying structure to the residuals, which could arise from diurnal flux patterns (e.g., day vs. night) and corresponding patterns to the uncertainty.

Similar estimates of the scaling exponent, \( \alpha \), were obtained with both the L-S and hybrid MSM/L-S methods, except that the latter method tended to result in more clearly defined differences in \( \alpha \) among the different models. We focus here on the hybrid MSM/L-S results. For the measured NEE fluxes, \( \alpha \) varied among sites between \( \approx 0.6 \) and \( \approx 1.1 \) (Fig. 2), indicating "pink" noise (an explanation for this variation across sites is given below). On the other hand, the scaling exponent for each model was more typically \( \approx 1.25 \), indicating that the modeled time series have a longer memory effect (i.e., tending towards "red" noise) than the measured NEE time series. The site-to-site variation in \( \alpha \) of the measured fluxes was not captured by any model. The MDS model consistently came closest to reproducing the spectral properties of the measured fluxes, probably because the narrow sliding window at which the algorithm operates minimizes persistent biases that could result from more slowly varying parameters.

The \( \alpha \) of model residuals was generally between \( \approx 0.2 \) and \( \approx 0.8 \) (although for one site, BE1 2000, the values ranged from \( \approx 0.1 \) to \( \approx 0.2 \), rather than zero as would be expected if residuals were uncorrelated white noise (Fig. 2). The fact that the \( \alpha \) of the residuals varied among sites, and in a manner that was similar to that for the NEE series (but not captured by any model), suggests that the residuals maintain a certain temporal correlation structure that is characteristic to each site.

---

**Fig. 1** – Comparison of model residuals for MDS and BETHY models. (A) Half hourly residuals; (B) weekly mean of model residuals. From (B) it is clear that BETHY results in stronger and more persistent biases (e.g., negative bias at days 120 and 240, positive biases at days 150, 180, and 210) than MDS.

**Fig. 2** – \( f' \) scaling exponents estimated from time series of measured (+) and modeled (filled circles) net ecosystem exchange of CO₂ (NEE), and the residual (open circles) between these two. Models are as follows: artificial neural network model (ANN), non-linear regression model (NLR), process-based ecosystem model (BETHY), semi-parametric model (SPM), and a look-up table based on marginal distribution sampling (MDS); sites (y-axis) are as identified in Table 1.
3.3. General properties of the uncertainty

The statistical properties of the aggregated (i.e., calculated across all site-years) data set of inferred random errors for the different methods are shown in Table 2A. In all cases, both the mean model residual, and the mean difference between paired observations, was relatively close to zero. The standard deviation of the inferred random error, $\sigma(\delta)$, was 20% smaller for the paired observation approach (2.19 mmol m$^{-2}$ s$^{-1}$) than for the best of the model-based approaches (2.73–3.10 mmol m$^{-2}$ s$^{-1}$), presumably because uncertainties estimated via the paired observation approach do not incorporate additional model error (conversely, these results also suggest that ANN and MDS have the lowest model error, while BETHY and NLR have the highest model error). The skewness was consistently close to zero, suggesting a relatively symmetric error distribution. However, the strong positive kurtosis indicated by all methods, approximately $\approx 10$, is characteristic of a leptokurtic error distribution, i.e., one characterized by a more strongly defined central peak and heavier tails than a normal distribution.

When a single model was used, and moments compared across sites (results shown for ANN model in Table 2B), the most obvious differences tended to be in terms of the standard deviation of the inferred random error, $\sigma(\delta)$, which was smallest for site-years FI1 2001 and 2002, and largest for site-years BE1 2001 and 2000, and FR1 2001 and 2002. Skewness was not especially prominent at any particular site. All sites had kurtosis of $\approx 3$ or more. For individual sites, skewness and kurtosis estimates were similar for all models.

3.4. Moments in relation to flux magnitude

Results were somewhat different when the data were broken down into bins of different flux magnitude (Fig. 3). Within a site, patterns were quite similar regardless of the model used, and the model-based approach was largely consistent with the paired observation approach, as illustrated by results from FR1 2001 (see Fig. 3A–C; this site-year was chosen for illustration because it had the fewest missing observations, 21%). Thus, the standard deviation of the inferred random error consistently increased with the absolute magnitude of the flux, and there were only small differences among models in the nature of this relationship (Fig. 3A). For this particular site-year, the skewness was generally negligible, i.e., $\leq 1$, and did not show any clear patterns of variation in relation to flux magnitude (Fig. 3B). For the $0 \mu$mol m$^{-2}$ s$^{-1}$ bin, skewness of random error inferred from NLR model residuals was much higher (1.6) than any of the other models, which were otherwise in the range 0.0–0.6 (note that while a similar pattern was seen for the other year of data for this site, FR1 2002, the pattern was not repeated across sites). The kurtosis was most pronounced, i.e., consistently $\geq 5$, for the $0 \mu$mol m$^{-2}$ s$^{-1}$ bin, and appeared to become smaller, i.e., closer to zero, with increasing absolute flux magnitude (Fig. 3C).

Across sites, patterns were consistent when a single model was used, as illustrated for the ANN model (Fig. 3D–F). The rate at which the standard deviation of the flux uncertainty increased with the magnitude of the flux (slope) was quite similar across sites, but the base level of uncertainty (y-axis intercept), i.e., at $F_c = 0$ mmol m$^{-2}$ s$^{-1}$,

### Table 2 - Statistical properties of inferred random flux measurement error (CO$_2$ fluxes measured in μmol m$^{-2}$ s$^{-1}$)

<table>
<thead>
<tr>
<th>Method</th>
<th>Site</th>
<th>N</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) All sites, by method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANN</td>
<td></td>
<td>121217</td>
<td>−0.004</td>
<td>2.73</td>
<td>0.22</td>
<td>11.61</td>
</tr>
<tr>
<td>NLR</td>
<td></td>
<td>121217</td>
<td>0.001</td>
<td>3.10</td>
<td>0.05</td>
<td>9.22</td>
</tr>
<tr>
<td>BETHY</td>
<td></td>
<td>121217</td>
<td>−0.048</td>
<td>3.10</td>
<td>−0.16</td>
<td>8.37</td>
</tr>
<tr>
<td>MDS</td>
<td></td>
<td>121217</td>
<td>0.067</td>
<td>2.80</td>
<td>−0.04</td>
<td>10.23</td>
</tr>
<tr>
<td>SPM</td>
<td></td>
<td>121217</td>
<td>0.000</td>
<td>2.93</td>
<td>0.09</td>
<td>9.66</td>
</tr>
<tr>
<td>Paired</td>
<td></td>
<td>40563</td>
<td>0.029</td>
<td>2.19</td>
<td>0.18</td>
<td>10.52</td>
</tr>
<tr>
<td>(B) ANN method, by site-year</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BE1 2000</td>
<td></td>
<td>12519</td>
<td>−0.043</td>
<td>3.21</td>
<td>0.03</td>
<td>3.55</td>
</tr>
<tr>
<td>BE1 2001</td>
<td></td>
<td>12287</td>
<td>0.000</td>
<td>2.89</td>
<td>0.32</td>
<td>7.49</td>
</tr>
<tr>
<td>DE3 2000</td>
<td></td>
<td>11452</td>
<td>−0.014</td>
<td>2.46</td>
<td>−0.16</td>
<td>7.01</td>
</tr>
<tr>
<td>DE3 2001</td>
<td></td>
<td>11731</td>
<td>−0.005</td>
<td>2.17</td>
<td>0.20</td>
<td>6.29</td>
</tr>
<tr>
<td>FI1 2001</td>
<td></td>
<td>11236</td>
<td>−0.002</td>
<td>1.28</td>
<td>−0.21</td>
<td>7.31</td>
</tr>
<tr>
<td>FI1 2002</td>
<td></td>
<td>11427</td>
<td>−0.015</td>
<td>1.32</td>
<td>0.32</td>
<td>8.46</td>
</tr>
<tr>
<td>FR1 2001</td>
<td></td>
<td>13744</td>
<td>−0.010</td>
<td>3.60</td>
<td>0.20</td>
<td>8.84</td>
</tr>
<tr>
<td>FR1 2002</td>
<td></td>
<td>13651</td>
<td>0.029</td>
<td>4.12</td>
<td>0.26</td>
<td>7.73</td>
</tr>
<tr>
<td>FR4 2002</td>
<td></td>
<td>11190</td>
<td>0.010</td>
<td>1.81</td>
<td>−0.31</td>
<td>2.99</td>
</tr>
<tr>
<td>IT3 2002</td>
<td></td>
<td>11980</td>
<td>0.005</td>
<td>2.29</td>
<td>0.50</td>
<td>6.76</td>
</tr>
</tbody>
</table>
varied among sites (Fig. 3D; note the consistency with patterns reported by Richardson et al., 2006a, indicated by the heavy blue shaded line; see also Table 3). Many sites had skewness ≈ 0 regardless of the magnitude of the flux, but for some site-years (in particular, FI1 2002 and IT3 2002, although these are not specifically identified in the figure), there was pronounced skewness (resulting from long, asymmetric, tails of the inferred random error), especially for the $0 \mu\text{mol m}^{-2}\text{s}^{-1}$ $F_c$ bin (Fig. 3E). The kurtosis pattern reported above for FR1 2001 appeared to be a more general result, as distributions became more mesokurtic (i.e., kurtosis approached zero) as the absolute magnitude of $F_c$ increased, and were in fact only strongly leptokurtic for the $0 \mu\text{mol m}^{-2}\text{s}^{-1}$ $F_c$ bin. The strong kurtosis (as well as the skewness) associated with the $0 \mu\text{mol m}^{-2}\text{s}^{-1}$ $F_c$ bin was due almost entirely to the extreme tails of the inferred random error. Trimming the top and bottom 1% of residuals resulted in a nearly symmetric distribution with less pronounced kurtosis.

These results were generally comparable (compare Fig. 3G and D, or I and F) to those obtained using the paired observation approach. However, although estimates of both the standard deviation and the kurtosis of the random measurement error (calculated for each flux magnitude bin within each site-year) were strongly correlated for model-based (ANN) and paired observation approaches (Fig. 4A and C), the skewness estimates for these two approaches were not at all correlated (Fig. 4B). This discrepancy is attributed to the inability of the paired observation approach to characterize odd-order moments of the distribution, because it is impossible to differentiate between a positive error in one measurement versus a negative error in the other measure-
ment. Thus, the paired observation approach should not be used to evaluate bias or skewness.

Two strong conclusions emerge from the above results. First, flux measurement errors become more Gaussian as fluxes increase in magnitude. For fluxes close to $0 \mu\text{mol m}^{-2} \text{s}^{-1}$, skewness, and especially pronounced kurtosis, are observed. Second, flux measurement errors are clearly heteroscedastic, meaning that the error variance is not constant but rather varies, particularly in relation to flux magnitude. We calculated the relationship between flux magnitude and flux measurement uncertainty for each site using the model-based approach (see Table 3 A for ANN results). Across the 10 site-years, the $y$-axis intercept of each regression was strongly correlated ($r = 0.95$, $P < 0.001$) with the overall standard deviation of the inferred random measurement errors given in Table 2 B, but the differences in slope among sites was not significant ($P = 0.10$). When data were pooled for all site-years, and relationships calculated separately for model-based and paired observation approaches (Table 3 B), the regression lines could not be statistically distinguished ($P > 0.10$). Furthermore, neither of these relationships could be distinguished ($P > 0.10$) from that reported by Richardson et al. (2006a) for $F_c \leq 0$. However, whereas Richardson et al. (2006a) found that the uncertainty increased more rapidly with flux magnitude for positive fluxes (nocturnal release) than negative fluxes (daytime uptake), the present results did not support this, as the difference in slopes for $F_c \leq 0$ and $F_c \geq 0$ was not significant ($P > 0.30$) for either model-based or paired observation approaches (see also Stauch et al., in press). This difference may be a result of more stringent nocturnal rejection criteria ($u^*$ filtering, quality flags and other data cleaning, see Papale et al., 2006) at the European sites compared to the sites used by Richardson et al. (2006a).

### Table 3 – Relationship between CO$_2$ flux magnitude ($F_c$, in $\mu\text{mol m}^{-2} \text{s}^{-1}$) and the standard deviation of the inferred random flux measurement error, $\sigma(i)$

<table>
<thead>
<tr>
<th>Method</th>
<th>Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) By site</td>
<td></td>
</tr>
<tr>
<td>BE1 2000</td>
<td>2.91(0.30) + 0.09(0.03) $\times</td>
</tr>
<tr>
<td>BE1 2001</td>
<td>2.27(0.17) + 0.14(0.02) $\times</td>
</tr>
<tr>
<td>DE3 2000</td>
<td>1.31(0.33) + 0.18(0.02) $\times</td>
</tr>
<tr>
<td>DE3 2001</td>
<td>1.13(0.21) + 0.16(0.01) $\times</td>
</tr>
<tr>
<td>FI1 2001</td>
<td>0.87(0.24) + 0.12(0.03) $\times</td>
</tr>
<tr>
<td>FI1 2002</td>
<td>0.21(0.17) + 0.12(0.02) $\times</td>
</tr>
<tr>
<td>FR1 2001</td>
<td>2.62(0.34) + 0.18(0.02) $\times</td>
</tr>
<tr>
<td>FR1 2002</td>
<td>3.49(0.38) + 0.13(0.02) $\times</td>
</tr>
<tr>
<td>FR4 2002</td>
<td>1.28(0.41) + 0.09(0.05) $\times</td>
</tr>
<tr>
<td>IT3 2002</td>
<td>1.79(0.07) + 0.11(0.01) $\times</td>
</tr>
<tr>
<td>(B) All sites</td>
<td></td>
</tr>
<tr>
<td>Model residuals (ANN)</td>
<td>1.69(0.20) + 0.16(0.02) $\times</td>
</tr>
<tr>
<td>Paired observations</td>
<td>1.47(0.22) + 0.17(0.02) $\times</td>
</tr>
<tr>
<td>(C) Richardson et al. (2006a)</td>
<td></td>
</tr>
<tr>
<td>$F_c \leq 0$</td>
<td>1.42(0.31) + 0.19(0.02) $\times</td>
</tr>
<tr>
<td>$F_c \geq 0$</td>
<td>0.62(0.73) + 0.63(0.09) $\times</td>
</tr>
</tbody>
</table>

Results are based on 10 site-years of data from a range of forested sites, as described in Methods. In (A), results are calculated separately for each site-year, based on ANN model residuals. In (B), results are calculated for residual and paired observation approaches, using data for all sites together. In (C), previously published results for forested sites are shown for comparison. Standard errors on estimated coefficients are given in parentheses.

### Fig. 4 – Comparison of statistical properties (standard deviation, skewness and kurtosis) of inferred flux measurement error calculated using two different approaches: the residual between measured values and a fitted artificial neural network model (“ANN”), and the difference between paired observations made exactly 24 h apart under similar environmental conditions (“Paired”). The 1:1 line is shown in all three panels. Reported correlation coefficients are for all site-years (each site-year shown by a different color) of data.
3.5. Linking flux measurement uncertainty and spectral properties

Additional analysis of the above results suggested some interesting relationships between the scaling exponent \( \alpha \), and the estimated random flux measurement uncertainty as characterized by the overall standard deviation of the inferred random measurement error, \( \sigma(\delta) \). First, the site-specific variation in \( \alpha \) of the measured NEE time series (‘+’ symbols in Fig. 2) was positively correlated with the standard deviation of model residuals for each site. Results are shown in Fig. 5A for the ANN model (\( r = 0.87 \)), but similar results were observed for all other models (\( r \) between 0.80 and 0.86). Thus, it should be feasible to estimate site-specific average uncertainties purely on the basis of the calculated scaling exponent, i.e., as:

\[
\sigma(\delta) = 6.59(0.72) + 4.75(0.82) \times \alpha
\]

(Coefficients were calculated using reduced major axis regression, with standard errors on coefficients given in parentheses). In the same vein, \( \alpha \) was also strongly correlated (\( r = 0.84 \)) with the \( y \)-intercept values reported in Table 3A, and given that results above indicated that the slope of the relationship between flux magnitude and flux uncertainty did not vary among sites, then this also provides a mechanism for developing site-specific estimates of this relationship, purely on the basis of \( \alpha \), i.e., as:

\[
\sigma(\delta) = 5.82(0.79) + 4.61(0.90) \times \alpha + 0.17 \times |F_c|
\]

Second, the variation in \( \alpha \) of model residuals (open circles in Fig. 2) was negatively correlated with the standard deviation of residuals for each model. Results shown in Fig. 5B are based on averages (across all sites) calculated for each model (\( r = -0.89 \), but similar results were obtained for individual sites as well (\( r \) between -0.72 and -0.98)).

4. Discussion

4.1. General characteristics of the measurement error

Random flux measurement uncertainties are characterized by non-stationary probability distributions, in that the statistical properties of the error vary over time and in relation to the magnitude of the flux. From this study we find that the manner in which the uncertainty scales with the magnitude of the flux (Table 3) is similar to that presented previously (e.g., Hollinger and Richardson, 2005; Richardson et al., 2006a; Stauch et al., in press). This provides a solid foundation for implementing a weighted optimization scheme in conducting model-data syntheses, whereby observations measured with greater confidence (lower \( \sigma(\delta) \)) receive more weight during the optimization (i.e., in the cost function) than observations measured with less confidence (higher \( \sigma(\delta) \)). Note that the non-zero intercept in the relationship between \( \sigma(\delta) \) and \( |F_c| \) means that large-magnitude fluxes have a better signal-to-noise ratio, and will still exert a greater influence during the optimization, than small fluxes.

The results presented in Fig. 4 indicate more complex patterns of variation in the statistical properties of the flux uncertainty compared to what has been previously reported (cf. Hollinger and Richardson, 2005). In particular, it would appear that while random measurement errors tend to be approximately Gaussian (symmetric and mesokurtic) for \( |F_c| > 0 \mu \text{mol m}^{-2} \text{s}^{-1} \), they are frequently non-Gaussian for \( |F_c| \approx 0 \mu \text{mol m}^{-2} \text{s}^{-1} \). In this regard, it may be considerably more challenging than has been previously acknowledged to properly implement a maximum likelihood optimizations scheme when inverting ecosystem models against flux data, because the error model needs to account for these differences.

![Fig. 5 – Relationship between scaling exponents and estimated random flux measurement uncertainty, characterized by the standard deviation of model residuals (models are described fully in text). In (A), the scaling exponent of raw NEE time series is plotted against the standard deviation of ANN model residuals, with each data point representing one site-year of data. In (B), the scaling exponent of model residual time series is plotted against the average across all site-years for an individual model. In both panels, the fitted line is based on reduced major axis regression.](image-url)
in the probability distribution function of the measurement errors.

If the random flux measurement uncertainty does not follow a Gaussian distribution, then what distribution does it follow? Hollinger and Richardson (2005) suggested that the distribution of inferred flux measurement errors was better approximated by a double exponential (Laplace) distribution, which has both a prominent central peak and heavy tails, and kurtosis of 3. By comparison, Chevallier et al. (2006) suggested that residuals between the ORCHIDEE model and flux data from a range of sites followed a Cauchy distribution. From the perspective of characterizing uncertainty for the purposes of data-model synthesis, the Cauchy distribution is a poor choice, since its first four moments are undefined. In a recent data-model synthesis Owen et al. (2007) implicitly assumed an error model where a large part of the data is Gaussian and high kurtosis is only introduced by a few high-leverage points (outliers that strongly influence the fourth moment). In the present study, our results indicated pronounced error kurtosis for $F_e \approx 0 \text{ mol m}^{-2} \text{s}^{-1}$, although trimming outliers that accounted for the top and bottom 1% of the error distribution resulted in kurtosis that was intermediate between a Gaussian (kurtosis = 0) and double-exponential (kurtosis = 3) distribution.

Hollinger and Richardson (2005) and Stauch et al. (in press) suggested that heteroscedasticity, combined with the varying frequency of different flux magnitudes, could result in an error distribution that appeared non-Gaussian, even if each error was drawn from a Gaussian distribution. As an example, take 8000 draws from $X_1 \sim N(0,1)$, and 2000 draws from $X_2 \sim N(0,5)$; the distribution of these 10,000 draws is characterized by a strong peak (predominantly because of $X_1$) and heavy tails (predominantly because of $X_2$), and thus appears non-Gaussian despite the normality of both $X_1$ and $X_2$. While this “addition of distributions” no doubt contributes to fact that at each site the overall kurtosis is always $\approx 3$ or larger, it does not explain why when the analysis is restricted to relatively narrow $F_e$ bins, the kurtosis is consistently $> 5$ for $|F_e| \approx 0 \text{ mol m}^{-2} \text{s}^{-1}$. The present analysis gives additional support for the idea that, at least for near-zero fluxes, the uncertainty is better characterized by a double exponential than a Gaussian distribution.

4.2. Patterns across sites

Random flux measurement uncertainties have been shown to vary among sites, especially in relation to vegetation type. For example, Richardson et al. (2006a) reported that uncertainties in both carbon and energy fluxes were smaller at a grassland site compared to forested sites. However, even among forested sites there may be substantial variation in uncertainty; we found, for example, that uncertainties were roughly three-fold larger at FR1 than F11 (Table 3A). The results presented in Table 3A are consistent with those shown previously by Moffat et al. (2007) and Stauch et al. (in press), i.e., for a given $F_e$, there is greater uncertainty at FR1 than either DE3 or FR4. Differences across sites in $\sigma(\delta)$ can be attributed to a range of factors, such as measurement height, surface roughness, wind speed, surface heterogeneity, and measurement system errors. These appear to affect the base level of uncertainty, but not the manner in which $\sigma(\delta)$ scales with $F_e$, as only the intercept, but not the slope, of regressions presented in Table 3A were found to vary significantly among these forested sites.

4.3. Uncertainty and the “color” of flux data time series

Spectral analysis methods enable researchers to view time series in the frequency domain rather than the standard temporal domain; in this manner, the time scales (from minutes to years) at which relevant variation occurs can be identified. Previous frequency domain analyses have made use of orthonormal wavelet transformations (e.g., Katul et al., 2001) and Fourier analysis (e.g., Baldocchi et al., 2001a). Here we used an alternative approach, based on the Lomb-Scargle periodogram (Lomb, 1976; Scargle, 1982) and the multiple segmentation method of Miramontes and Rohani (2002), which is naturally suited to investigating the correlation structure of short time series with missing data (i.e., gaps). This analysis showed that while the $f_\text{c}$ scaling exponent for most sites indicated a moderate correlation structure (“pink” noise), there were some sites that tended more towards “red” noise (e.g., FR4) and other sites that tended towards “white” noise (FR1, BE1). An interesting finding was that there was a robust connection between the estimated random flux measurement uncertainty and the scaling exponent, i.e., a strong correlation between $f_\text{c}$ and $\sigma(\delta)$ (Fig. 5A). A likely explanation for this result is that at sites with larger measurement uncertainties, the respective white noise components of the NEE measurements tends to sufficiently obscure the underlying signal such that persistence is much less pronounced compared to sites with smaller measurement uncertainties. Thus, larger random flux measurement errors lead to “whiter” NEE time series.

Similarly, variation among models in the $\sigma$ of residual time series was found to be correlated with the standard deviation of residuals for each model (Fig. 5B). A possible interpretation for this pattern is that models that fit the NEE time series less well tend to have not only larger residuals (and hence higher $\sigma(\delta)$), but also larger and more long-lived systematic biases (i.e., greater persistence and hence more negative values of $\sigma$). Thus, increasing model error (compare BETHY and MDS) results in “reddened” model residuals.

The most flexible models used here, ANN and MDS, yielded inferred random errors that were closest to white noise and also had statistical properties that were most similar to those of differenced paired measurements. An explanation for this is that predictions of ANN and MDS models for time period effectively represent the mean of all fluxes measured under similar environmental conditions and at a similar time of day and time of the year as those that prevail at time $t$. Residuals from these models therefore can be seen as analogues of the paired measurement approach; the difference being that the model residual is the difference between a measurement ($F_i$) and the mean of a population of measurements ($\mu$), rather than the difference between two pure measurements.

4.4. Systematic errors and biases

Until now, our analysis and discussion has focused on random flux measurement errors. Systematic flux measurement
errors, or biases, are often dealt with by “flagging” according to data quality, after applying pre-defined filters (Foken and Wichura, 1996). Poor quality data often occur when low-frequency turbulent motions or non-stationary conditions occur. This is because there is substantial uncertainty about the extent to which low-frequency covariances contribute to the actual exchange between surface and atmosphere (von Randow et al., 2002). Apart from these, a wide range of error sources contributes to bias, ranging from poor instrument maintenance and uncertain delays in sampling tubes to uncertainty on the height of the zero-plane displacement (which determines the magnitude of several frequency corrections to the data) (Kruj et al., 2004). A number of systematic errors are uncertain even in sign, while for others the corrections to be made are themselves highly uncertain. This suggests that systematic errors (and attempts to correct for them) have a random component to them, although usually the error source varies on a (much) larger time scale (compared to the purely random measurement errors we have quantified here). Kruj et al. (2004) assessed the importance of such errors by assuming ignorance on their magnitude and recalculating fluxes from raw data using a range of configuration settings (note also that Kruj et al. expressed uncertainties in terms of percentages, rather than absolute amounts, which implies that systematic errors are heteroscedastic as well, similar to results reported here for random errors). The resulting standard errors can be used to assess random uncertainty, but, as they act at longer time scales, error will not reduce with the number of data taken as would be expected for independent data points. Rather, for each error source the appropriate number of degrees of freedom in the data should be assessed, by assessing the periodicity of, for example, maintenance, wind directions, and other factors that affect “slow” errors. The methodology to assess such ranges of errors at different time scales still needs to be developed, but in general this suggests that, instead of distinguishing “random” and “systematic” error, there is a range of time scales at which errors should be assessed.

5. Summary and conclusions

We evaluated the statistical properties of random flux measurement errors inferred first from model residuals, and second, using a standard paired measurement approach. The five models used in the present analysis were quite different with respect to structure and parameterization, but all adequately captured the patterns of variation in the flux data time series across a range of temporal scales. However, the power spectra of the model predictions indicated too little power at the highest frequencies (stochastic variation that is due to random measurement errors) and too much power at lower frequencies. There was a tendency for the residuals of some models to exhibit greater persistence than other models, and this memory effect (“reddened” model residuals) was attributed to model error which resulted in larger and more long-lived systematic biases.

Two models in particular, ANN and MDS, made no prior assumptions about the functional form of relationships between environmental drivers and CO$_2$ fluxes, and the resulting uncertainty estimates were in closest agreement with the paired observation approach. Compared to the paired observation approach, an advantage of inferring uncertainties from model residuals is that many more data points are available (an average of three times as many for our data) from which the statistics of the distribution can be estimated; this is especially relevant with short flux time series, where there may be an insufficient number of paired observations meeting the criteria for environmental similarity. While the model residual method does not require two towers, or arbitrary limits for what constitutes “similar” environmental conditions on successive days, care must be taken to ensure that a good model is used, so that model error does not confound uncertainty estimates.

In all cases, the flux measurement uncertainty increased with flux magnitude. The base level of uncertainty differed markedly among sites, and patterns of variation were consistent regardless of whether residuals or paired measurements were used to infer the error, suggesting that both methods yield robust uncertainty estimates that are sensitive to real underlying differences in uncertainty (caused by, among other things, site heterogeneity and measurement system characteristics). We were able to link the site-specific uncertainties to the spectral properties ($f$ scaling) of the original NEE time series: these results indicated the potential for estimating $\sigma(f)$ on the basis of the scaling coefficient $\alpha$.

The results presented here offer a basis by which the random flux measurement errors can be specified for a variety of applications, including data-model fusion efforts, statistically rigorous model evaluation, and regional-to-continental integration and synthesis projects. A logical next step would be to reconcile random and systematic errors under a common framework, so that the total uncertainty in measured fluxes (and annual integrals) might be more accurately specified.

Acknowledgements

A.D.R. was supported by the Office of Science (BER), U.S. Department of Energy, through Interagency Agreement No. DE-A102-07ER64355. D.P. was supported by the CarboEurope-IP research program, funded by the European Commission. The Max Planck Institute for Biogeochemistry provided funding and support for the gap filling comparison workshop held in Jena in September 2006. Site Pls Marc Aubinet (Vielsalm), Werner Kutsch (Hainich), André Granier (Hesse), Serge Rambal (Puechabon), Riccardo Valentini (Roccarespampani) and Timo Vesala (Hytiälä) are thanked for making their data available. Ankur Desai and three anonymous reviewers are all thanked for their constructive criticism.

References


