Improving land surface models with FLUXNET data

M. Williams, A. D. Richardson, M. Reichstein, P. C. Stoy, Philippe Peylin, H. Verbeeck, N. Carvalhais, M. Jung, D. Y. Hollinger, J. Kattge, et al.

HAL Id: bioemco-00450903

https://hal-bioemco.ccsd.cnrs.fr/bioemco-00450903v1

Submitted on 30 May 2020

Biogeosciences

Improving land surface models with FLUXNET data

M. Williams[1], A. D. Richardson[2], M. Reichstein[3], P. C. Stoy[4], P. Peylin[5], H. Verbeeck[6], N. Carvalhais[7], M. Jung[3], D. Y. Hollinger[8], J. Kattge[3], R. Leuning[9], Y. Luo[10], E. Tomelleri[3], C. M. Trudinger[11], and Y. -P. Wang[11]

[1]School of GeoSciences and NERC Centre for Terrestrial Carbon Dynamics, University of Edinburgh, Edinburgh, UK
[2]University of New Hampshire, Complex Systems Research Center, Durham, USA
[3]Max Planck Institute for Biogeochemistry, Jena, Germany
[4]School of GeoSciences, University of Edinburgh, Edinburgh, UK
[5]Laboratoire des sciences du climat et de l'environnement (LSCE), Gif Sur Yvette, France
[6]Laboratory of Plant Ecology, Ghent University, Belgium
[7]Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal
[8]Northern Research Station, USDA Forest Service, Durham, NH, USA
[9]CSIRO Marine and Atmospheric Research Canberra ACT 2601, Australia
[10]Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA
[11]CSIRO Marine and Atmospheric Research, Centre for Australian Weather and Climate Research, Aspendale, Victoria, Australia

Abstract. There is a growing consensus that land surface models (LSMs) that simulate terrestrial biosphere exchanges of matter and energy must be better constrained with data to quantify and address their uncertainties. FLUXNET, an international network of sites that measure the land surface exchanges of carbon, water and energy using the eddy covariance technique, is a prime source of data for model improvement. Here we outline a multi-stage process for "fusing" (i.e. linking) LSMs with FLUXNET data to generate better models with quantifiable uncertainty. First, we describe FLUXNET data availability, and its random and systematic biases. We then introduce methods for assessing LSM model runs against FLUXNET observations in temporal and spatial domains. These assessments are a prelude to more formal model-data fusion (MDF). MDF links model to data, based on error weightings. In theory, MDF produces optimal analyses of the modelled system, but there are practical problems. We first discuss how to set model errors and initial conditions. In both cases incorrect assumptions will affect the outcome of the MDF. We then review the problem of equifinality, whereby multiple combinations of parameters can produce similar model output. Fusing multiple independent and orthogonal data provides a means to limit equifinality. We then show how parameter probability density functions (PDFs) from MDF can be used to interpret model validity, and to propagate errors into model outputs. Posterior parameter distributions are a useful way to assess the success of MDF, combined with a determination of whether model residuals are Gaussian. If the MDF scheme provides evidence for temporal variation in parameters, then that is indicative of a critical missing dynamic process. A comparison of parameter PDFs generated with the same model from multiple FLUXNET sites can provide insights into the concept and validity of plant functional types (PFT) – we would expect similar parameter estimates among sites sharing a single PFT. We conclude by identifying five major model-data fusion challenges for the FLUXNET and LSM communities: (1) to determine appropriate use of current data and to explore the information gained in using longer time series; (2) to avoid confounding effects of missing process representation on parameter estimation; (3) to assimilate more data types, including those from earth observation; (4) to fully quantify uncertainties arising from data bias, model structure, and initial conditions problems; and (5) to carefully test current model concepts (e.g. PFTs) and guide development of new concepts.

Correspondence to: M. Williams
(mat.williams@ed.ac.uk)
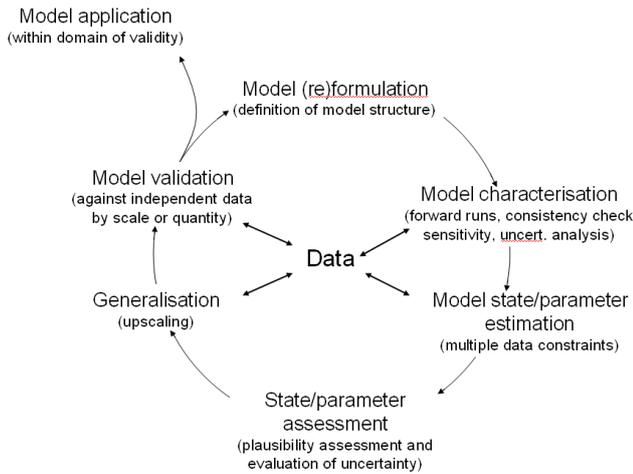
Fig. 1. The multi-stage process for model-data fusion: a conceptual diagram showing the main steps (and the iterative nature of these steps) involved in a comprehensive data-model fusion.

1 Introduction

Land surface models are important tools for understanding and predicting mass and energy exchange between the terrestrial biosphere and atmosphere. A land surface model (LSM) is a typical and critical component of larger domain models, which are aimed at global integration, for example global carbon cycle models and prognostic global climate models. These integrated models are key tools for predicting the likely future states of the Earth system under anthropogenic forcing (IPCC, 2007), and for assessing feedbacks with, and impacts on, the biosphere (MEA, 2005). Land surface models represent the key processes regulating energy and matter exchange – photosynthesis, respiration, evapotranspiration (Bonan, 1995; Foley et al., 1996; Williams et al., 1996; Sellers et al., 1997), and their coupling. These processes are sensitive to environmental drivers on a range of timescales, for example, responding to diurnal changes in insolation, and seasonal shifts in temperature and precipitation. Land surface processes influence the climate system, through their control of energy balance and greenhouse gas exchanges. Forecasts of global terrestrial C dynamics that rely on LSMs show significant variability over decadal timescales (Friedlingstein et al., 2006), especially when coupled to climate, indicating that major uncertainties remain in the representation of critical ecosystem processes and climate feedbacks within global models.

In recent years the widespread use of the eddy covariance (EC) methodology has led to a large increase in data on terrestrial land surface exchanges (Baldocchi et al., 2001). FLUXNET is an international network of EC sites with data processed according to standardized protocols (Papale et al., 2006). The EC time-series data from FLUXNET provide rich insights into exchanges of water, energy and CO_2 across a

range of biomes and timescales. While LSM forward runs are commonly compared with EC data, there is a growing consensus that models must be better constrained with such data to address process uncertainty (Bonan, 2008). A stronger link between models and observations is needed to identify poorly represented or missing processes, and to provide confidence intervals on model parameter estimates and forecasts.

New methods are becoming available to assist data analysis and generate links to models, based on the concept of model-data fusion, MDF (Raupach et al., 2005). MDF encompasses a range of procedures for combining a set or sets of observations and a model, while quantitatively incorporating the uncertainties of both. MDF is used to estimate model states and/or parameters, and their respective uncertainties.

The objective of this paper is to provide guidance to the LSM community on how to make better use of eddy covariance data, particularly via MDF. We first outline the philosophical principles behind model-data fusion for model improvement. We then discuss the structure of typical land surface models and how they are parameterised. Next we detail FLUXNET data availability and quality, specifically in the context of land surface models. We discuss approaches for model and data evaluation, focussing on new techniques using time series and spatial analyses. Finally we discuss formal model-data fusion and highlight the need for multiple constraints in model evaluation and improvement, and effective assessment of model and data errors. We conclude with a set of challenges for the LSM and MDF communities.

2 The philosophy of model-data fusion for model improvement

Model calibration, evaluation, testing, and structural improvement (re-formulation) are all key aspects of model-data fusion; in other words, MDF is not simply tuning model parameters to yield model predictions that match the calibration data. Rather, it is a multi-stage process (Fig. 1). At each of these stages, there is interplay between data, model structure, and modeller. The process details depends somewhat on whether the problem is focussed on state estimation of the system, or on parameter estimation of the model. In both cases a rigorous characterization of the model structure through consistency checks and testing sensitivity to parameters and drivers, in the same way as in classical forward modelling approaches, is still a prerequisite for a meaningful data-model fusion. This model characterization also constitutes the baseline against which any improvements and reductions of uncertainties can be judged. If *state estimation* is the goal, then model states are adjusted to generate closer agreement with the observations. Further analysis can make use of these state adjustments to identify poorly represented processes and their timings. It is important to ensure that state adjustments are consistent with all independent

Table 1. Land surface model components and details, after Liu and Gupta (2007).

Model component	Examples for typical LSMs
System boundary	Lower atmosphere, deep soil/geological parent material
Forcing inputs	Air temperature, short and long-wave radiation, precipitation, wind speed, vapour pressure deficit, atmospheric CO_2 concentration
Initial states	Biogeochemical pools, vegetation and soil temperature and water content
Parameters	Rate constants for chemical processes, physical constants, biological parameters
Model structure	Process definitions and connectivity
Model states	Biogeochemical pools, vegetation and soil temperature and water content
Outputs	Biogeochemical fluxes, dynamics of model states

observations. If so, model deficiencies can be clearly located. If *model parameter estimation* is the goal, then model parameters are adjusted so that the model state(s) come into closer agreement with the observations. Following the optimization of model parameters (defined here loosely to potentially include both parameters *sensu stricto* and state variables), it is vitally important that further analyses be conducted by the modeller to: (1) quantify uncertainties in optimized parameters (and thus identify those parameters, and related processes, that are poorly constrained by the data); (2) evaluate the plausibility and temporal stability of optimized parameter values (reality check); (3) understand when and why the model is failing ("detective work", which may involve additional validation against independent data sets); and (4) identify opportunities for model improvement (re-formulation of structure and process representation). When treated in this manner, MDF has relevance to both basic and applied scientific questions. Thus, not only do identified deficiencies lead to model reformulation (followed by further data-model fusion, possibly with new data sets being brought to bear), but the model can also be applied to answer new science questions. Successful application of the model is, however, contingent on the modeller's understanding (which comes from the posterior analyses) of the domain (in terms of space, time, and prognostic variables), where the model can be applied with precision and confidence.

3 Land surface models

All models consist of seven general components; the system boundary, forcing inputs, initial states, parameters, model structure, model states, and outputs (Liu and Gupta, 2007). The details of these components for typical LSMs are given in Table 1. These are physical, chemical and biological processes, related to fluxes of energy, water and carbon, that are sensitive to changes in environmental forcing on multiple time scales (Fig. 2). Models differ as to whether they also represent processes operating on longer time scales. Some models link CO_2 fluxes to plant traits such as allocation, litter production, phenology, vegetation dynamics and competition, while others rely on prescribed inputs of vegetation
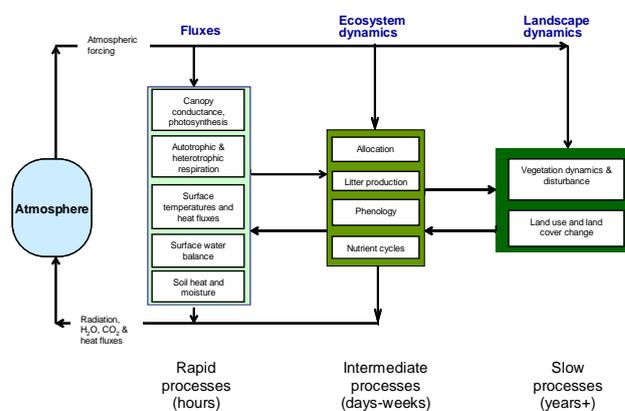
Fig. 2. Schematic of a typical land surface model. LSMs include a range of processes operating on different scales. LSMs can be driven by atmospheric observations or analyses, or directly coupled to atmospheric models. All LSMs tend to represent the hourly processes, but daily and annual processes may not be included.

structure to drive flux predictions. Here, we focus largely on models that couple fluxes to the dynamics of structural C pools, as only these models are capable of prediction over decadal timescales.

The typical approach for LSM parameterisation has been to use the concept of plant functional types (PFT) (Prentice et al., 1992). The terrestrial biosphere is divided into discrete units based on presence or absence of trees, shrubs or grasses; evergreen or deciduous habit; C3 or C4 photosynthetic pathways; and broad or needle leaves. Each of these PFTs has a nominal parameter vector, derived largely from the literature, laboratory experiments, or limited field campaigns focussed on particular ecosystems components, and often supplemented by steady state assumptions. In most models, the distribution of PFTs across the globe is prescribed, but in dynamic global vegetation models (DGVMs) the distribution of PFTs is determined by bioclimatic limits and competition among PFTs within shared bioclimatic space (Sitch et al., 2003). Field observations, however, suggest that intrinsic parameter values vary over space and environmental conditions, even within a PFT, challenging the assumptions of the PFT approach (Wright et al., 2004).

(A)

(B)
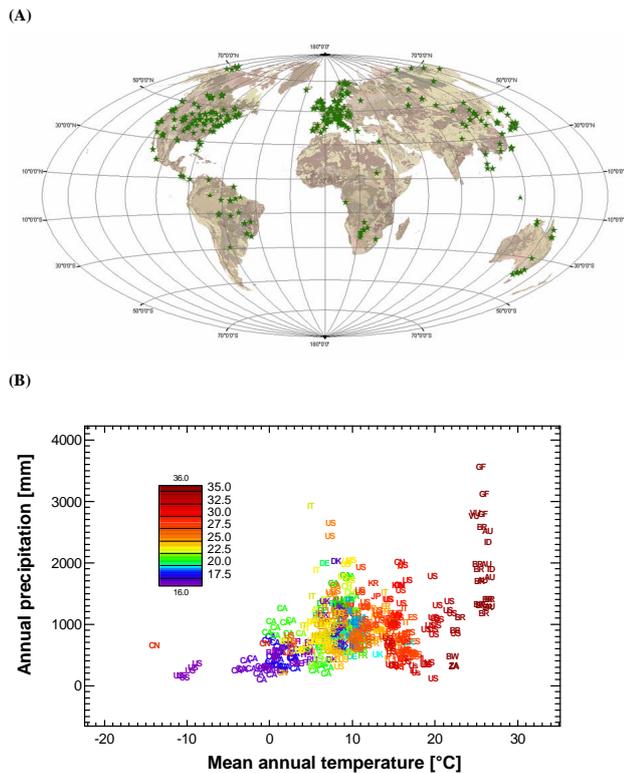
Fig. 3. Distribution of sites that are present in the current La-THuile 2007 FLUXNET database. **(a)** Geographical distribution, **(b)** in climate (annual temperature and precipitation) space. In (b) colours code annual potential shortwave radiation flux density [MJ m^{-2} day^{-1}] according to the legend. Letters are country codes.

Differences among LSMs in structure and parameterisation mean that they have different response functions relating rates of carbon and water fluxes to global change factors (e.g., CO_2 concentration, temperature, and precipitation or soil moisture content). Subtle changes to response functions and parameterizations can yield divergent modelled responses of ecosystems to environmental change, as has been demonstrated by parameter sensitivity studies (e.g. Zaehle et al., 2005; White et al., 2000) and model intercomparisons (e.g. Friedlingstein et al., 2006; Jung et al., 2007). Differences in the way processes interact in the model are also likely candidates for divergent model behaviour, but even harder to detect than differences in process representation (Rastetter, 2003). Rigorous comparison with data provides the basis for better constraining model structure and for generating more defensible parameter estimates with realistic uncertainty estimates.

4 Data availability, limitations and challenges

4.1 Overall overview

Continuous measurements of surface-atmosphere CO_2 exchange using the eddy covariance technique started in 1990s (Verma, 1990; Baldocchi, 2003). Consequently, many EC observation sites have been established, leading to the development of regional networks, such as AmeriFlux and Euroflux, and then the global network FLUXNET, a "network of regional networks" (Baldocchi et al., 2001). The need for harmonized data processing has been recognized in the different networks, and respective processing schemes documented in the literature (Foken and Wichura, 1996; Aubinet et al., 2000; Falge et al., 2001; Reichstein et al., 2005; Papale et al., 2006; Moffat et al., 2007). A first cross-network standardized dataset was established in 2000 (Marconi dataset) and contained data from 38 sites from North America and Europe. In 2007 the first global standardized data set ("La-Thuile FLUXNET dataset", cf. http://www.fluxdata.org) was established.

4.2 Current distribution and representativeness of flux sites

The current data set contains data from >950 years and 250 sites, from all continents except Antarctica. The majority of sites are located within the extra-tropical Northern hemisphere (Fig. 3a) and temperate climates (Table 2), although most areas in the temperature-climate space are covered at least by some sites (Fig. 3b). Similarly, most coarse vegetation types are covered by the current network, however with crop sites underrepresented particularly outside North America and Europe. Due to the history of FLUXNET, long-term data sets with >6 years data exist for only 38 sites (as of August 2008), and most are forests. Even within North America or Europe, where sites are reasonably well distributed among the major biomes (Hargrove et al., 2003), archived flux data on their own cannot possibly provide the information needed for a full spatio-temporal understanding of terrestrial carbon processes. Processes with time constants longer than the data record usually cannot be reliably identified and parameterised, and spatial representativeness may still be limited at the regional scale. For example, Fig. 4 shows temporal development of MODIS FPAR in a 100×100 km area surrounding one eddy-flux tower. The figure demonstrates that FPAR at the tower site exceeds average FPAR for the broader region, and indicates that the tower represents the more productive portion of the landscape, assuming FPAR is a proxy for productivity. This limitation can be overcome by integration of data across multiple temporal and spatial scales – i.e. data on carbon stocks and pools, and how these pools change over time combined with spatially extensive observations from remote sensing (Quaife et al., 2008) or national

Table 2. Distribution of sites in the current FLUXNET La-Thuile data set with respect to climate and vegetation classes. Climate is defined according to aggregated Kppen-Geiger classification, cf. www.fluxdata.org. Vegetation classes (top line) are from IGBP definitions: Croplands, closed shrublands, deciduous broadleaf forest, evergreen broadleaf forest, evergreen needleleaf forest, grassland, mixed forest, open shrublands, savanna, wetlands, woody savanna.

Number of Sites available	CRO	CSH	DBF	EBF	ENF	GRA	MF	OSH	SAV	WET	WSA	Total
Tropical	1	0	0	10	0	1	0	1	1	0	2	16
Dry:	0	0	0	1	1	3	0	1	1	0	3	10
Subtropical/mediterranean:	5	3	11	5	17	11	2	3	2	0	4	63
Temperate	17	0	8	2	12	18	4	0	0	4	0	65
Temperate continental	7	1	9	1	17	7	8	3	0	0	0	53
Boreal	0	0	2	0	22	4	2	4	0	4	0	38
Arctic	0	0	2	0	0	1	0	0	0	3	0	6
Total	30	4	32	19	69	45	16	12	4	11	9	251

forest inventories, e.g. US Forest Inventory and Analysis Program (Gillespie et al., 1999).

4.3 Data characteristics and limitations

The global network of eddy covariance towers not only provides flux observations themselves but also the infrastructure necessary to study ecosystem processes, and relationships between ecosystem processes and environmental forcing, across a range of spatial and temporal scales. The time averaging of the measurements (typically either 30 or 60 min) is adequate to resolve both diurnal cycles and fast ecosystem responses to changes in weather. In addition to the measured fluxes (typically CO_2 as well as latent and sensible heat) and enviro-meteorological data, many sites maintain a programme of comprehensive ecological measurements, including detailed stand inventories, litterfall collections, soil respiration and leaf-level photosynthesis, soil and foliar chemistry, phenology and leaf area index.

However, a careful consideration of the limitations of the data is mandatory to avoid overinterpretation of model-data mismatches in evaluation or assimilation schemes. Flux data are noisy and potentially biased, containing the "true" flux, plus both systematic and random errors and uncertainties (Table 3). There is a growing recognition within the EC community on the need to quantify the random errors inherent in half-hourly flux measurements (Hollinger and Richardson, 2005; Richardson et al., 2006), but also uncertainties in the corrections of systematic errors. Because data uncertainties enter directly into model-data fusion (see below), misspecification of data uncertainties affects parameter estimates and propagates into the model predictions, leading Raupach et al. (2005) to suggest that "data uncertainties are as important as data values themselves".

Recent research, using data from sites in both North America (Richardson et al., 2006) and Europe (Richardson et al., 2008), indicates that the random error scales with the magnitude of the flux, thus violating one key assumption (homoscedasticity) underlying ordinary least squares optimiza-
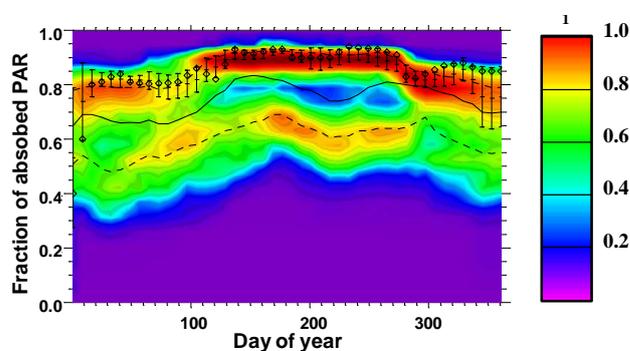
Fig. 4. Temporal development of MODIS fPAR-estimates during 2002 at the French Puchabon FLUXNET site, an evergreen broadleaf forest (Holm Oak). Diamonds indicate the time series at the tower-pixel, vertical lines denote between-pixel standard deviation of 3×3 pixels around the tower. Colours code the frequency of pixels within 100×100 km around the tower having the respective fPAR at the day of the year indicated on the x-axis. Solid and dashed lines indicate median, lower and upper quartile fPAR over all pixels 100×100 km around the tower. It can be seen that the tower represents the more productive portion of the landscape (assuming FPAR is a proxy for productivity).

tion. On the other hand, Lasslop et al. (2008) analysed further properties of the random error component and found little temporal autocorrelation and cross-correlation of the H_2O and CO_2 flux errors, implying that the weighted least squares criterion yields maximum likelihood estimates of the model parameters. The assumptions about the statistical distribution of the flux error influence the weighting on data in parameter and state estimation procedures. For instance, a least squares weighting is required if flux error is Gaussian, while mean absolute error is required if flux error is Laplacian (Richardson and Hollinger, 2005). Random flux measurement errors are relatively large at the half-hourly time step ($\sim$20% of the flux), but – owing to the large number of measurements – diminish for annual sums when propagated

Table 3. Characteristics of systematic and random uncertainties in eddy covariance measurements of surface-atmosphere exchanges of CO_2, H_2O and energy.

	Systematic	Random
Examples	– inadequate nocturnal turbulence (low u^*)	– uncertainties due to surface heterogeneity and time varying footprint
	– attenuation and imperfect spectral response; calibration errors	– turbulence sampling errors
	– incomplete energy balance closure	– measurement equipment glitches (IRGA and sonic anemometer "spikes")
	– advection and non-flat terrain	
Time scale	– operate at varying time scales: fully systematic vs. selectively systematic	– affect all measurements; assume no autocorrelation
	– accumulate linearly over time, i.e. as $\sum d_i$	– accumulate in quadrature over time, i.e. as $\sqrt{\sum (\varepsilon_i)^2}$
Nature	– deterministic, with variety of effects: e.g. fixed offset vs. relative offset	– stochastic: probability distribution function with mean zero and standard deviation s
Potential for identification	– cannot be identified through statistical analyses	– characteristics of pdf can be estimated via statistical analyses (but may be time-varying)
Potential for correction	– can correct for systematic errors (but corrections themselves are uncertain)	– cannot correct for random errors
Effect on data-model fusion	– uncorrected systematic errors will bias data-model fusion analyses	– random errors limit agreement between measurements and models, but should not bias results

as independent errors, typically resulting in an uncertainty $<50\,\mathrm{g\,C\,m^{-2}\,yr^{-1}}$ (Richardson et al., 2006; Lasslop et al., 2009). However, the characterization of the statistical properties of flux measurement error remains currently under debate and probably varies between sites and environmental conditions (Richardson et al., 2006; Lasslop et al., 2008).

Systematic flux biases are less easily evaluated, but are certainly non-trivial (Goulden et al., 1996; Moncrieff et al., 1996; Baldocchi, 2003). One major source of bias is incorrect measurement of net ecosystem exchange under low-turbulence conditions, especially at night. Attempts to minimize this bias use so-called u^*-filtering (Papale et al., 2006), in which measurements are rejected when made under conditions of low turbulence (i.e. below a threshold friction velocity). However, the threshold value for u^* is often uncertain. Preliminary analysis of the La Thuile data set indicated that the uncertainty introduced varies between sites, and can be $>150\,\mathrm{gC\,m^{-2}\,yr^{-1}}$ in some cases (Reichstein et al., unpublished data); hence a careful analysis of this effect is warranted.

Energy balance closure at most sites is poor: the sum of sensible, latent and soil heat fluxes is generally $\sim$20% lower than the total available radiative energy, which may indicate systematic errors in measured CO_2 fluxes as well (Wilson et al., 2002). Advection is believed to occur at many sites, particularly those with tall vegetation, and is thought to bias

annual estimates of net sequestration upwards, i.e. more negative NEE (Lee, 1998). While these (and other) systematic errors are known to affect flux measurements, much less is known about how to adequately quantify and correct for them; typically the time scale at which the correction needs to be applied, and in some cases even the direction of the appropriate correction, is highly uncertain (see discussion in Richardson et al., 2008).

Although attempts are made to make continuous flux measurements, in reality there are gaps in the flux record. These gaps most often occur when conditions are unsuitable for making measurements (e.g., periods of rain or low turbulence), or as a result of instrument failure. A variety of methods have been proposed and evaluated for filling these gaps and producing continuous time series; the best methods appear to approach the limits imposed by the random flux measurement errors described above (e.g., Moffat et al., 2007). In general, MDF will want to use only measured, and not gap-filled, data, because gap-filling involves the use of a model and so can contaminate the MDF. However, in some cases (if, for example, the model time step is different from that of the data, e.g. daily) it may be necessary to resort to using gap-filled data, in which case it is important that the distinction between "measured" and "filled" data be understood.

Another significant challenge to obtaining process-level insights from measured fluxes is that tower-based

measurements reflect the net ecosystem exchange (NEE) of CO_2, whereas the process-related quantities of interest are typically the component fluxes, gross photosynthesis (GPP) and ecosystem respiration (R_E; also the above- and below-ground, and autotrophic and heterotrophic, components of ecosystem respiration), which, unlike NEE, represent distinct sets of physiological activity. In a model-data fusion context, the net flux (particularly when integrated over a full day or even month) does not constrain the overall dynamics as well as the component fluxes because the net flux could be mistakenly modelled by gross fluxes having large compensating errors; any combination of GPP and R_E magnitudes can result in a given NEE; the problem is mathematically unconstrained.

4.4 Links to other datasets

While flux data themselves are valuable for evaluating and constraining models, meteorological and ecosystem structural data are needed as drivers and additional constraints. While there are also uncertainties associated with ancillary meteorological data, these are primarily due to spatial heterogeneity rather than instrument accuracy or precision. Similar and often larger problems of measurement accuracy, sampling uncertainty and representativeness occur with biometric measurements (e.g. fine root NPP) and chamber based flux estimates (e.g. soil and stem respiration) (Savage et al., 2008; Hoover, 2008).

Long-term continuous flux observations are much more valuable when coupled with a full suite of ancillary ecological measurements. Having measurements of both pools and fluxes means that both can be used to constrain models. With multiple constraints the biases in any single set of measurements becomes less critical and data-set internal inconsistencies – that can never be resolved by a model obeying conservation of mass and energy – can be identified (Williams et al., 2005). Thus flux sites most appropriate for model data fusion under current conditions have multiple years of continuous flux and meteorological data, with well-quantified uncertainties, along with reliable biometric data on vegetation and soil pools and primary productivity, as well as temporal estimate of leaf area index, soil respiration and sapflux. Many global change experiments that manipulate temperature, CO_2, precipitation, nitrogen and other factors often provide extensive measurements of many carbon, water, and energy processes. Data sets from manipulation experiments are not only valuable for constraining parameters in land surface models, but also useful for examining how environmental factors influence key model parameters (Luo et al., 2003; Xu et al., 2006).

5 Techniques for model and data evaluation

There exist a range of metrics for evaluating models against observations which are an important component of the model-data fusion process (Fig. 1). Typical techniques include calculations of root-mean-square error (RMSE), residual plots, and calculation of statistics like R^2 to determine the amount of measurement variability explained by the model. These approaches are discussed elsewhere in detail (Taylor, 2001). Here, instead, we discuss novel model evaluation techniques, using eddy flux data to evaluate models in the time and frequency domains, and in physical and climate space, before considering model-data fusion itself.

5.1 Evaluation in time: measurement patterns to assess model performance

The evaluation of LSMs using traditional methods may not be optimal because of large, non-random errors and bias in instantaneous or aggregated fluxes. We suggest that model evaluation may be improved by quantifying patterns of C exchange at different frequencies (Baldocchi et al., 2001; Braswell et al., 2005), recognizing especially the tendency for models to fail at lower frequencies, i.e. annual or interannual time scales (Siqueira et al., 2006).

We demonstrate such an approach for model improvement and evaluation by comparing eight years of measured latent heat exchange (LE) from the Tumbarumba flux monitoring site (Leuning et al., 2005) with corresponding predictions from the CABLE model (Kowalczyk et al., 2006), and a CABLE improvement that explicitly accounts for soil and litter layer evaporation (CABLE$_{SL}$ – Haverd et al., 2009). CABLE$_{SL}$ represents an improvement in overall model fit compared to CABLE (Fig. 5a and b), but conventional scatterplots and associated goodness-of-fit statistics cannot determine the times or frequencies when the model has been improved, or how the model can be further improved. (We also note that the daily flux sums are not statistically independent). The wavelet coherence, similar in mathematical formulation to Pearson's correlation coefficient, can be used to identify the times and time scales at which two time series are statistically related (Grinsted et al., 2004), and we note that these time series can represent those of measurements and models (Fig. 5c and d). Specifically, Grinsted et al. noted that wavelet coherence values above 0.7 (yellow-to-red colours in Fig. 5c and d) represent a significant relationship at the 95% confidence limit.

In the Tumbarumba example, variability in CABLE$_{SL}$ is significantly related to measurements across time at seasonal time scales of approximately 4 months (102.1 days); note the band at this time scale in Fig. 5c and its near disappearance in Fig. 5d. At the same time, swaths of poor model/measurement coherence exist at shorter time scales, particularly during the years 2003 and 2005–2006. Unique events that are responsible for these oscillations in
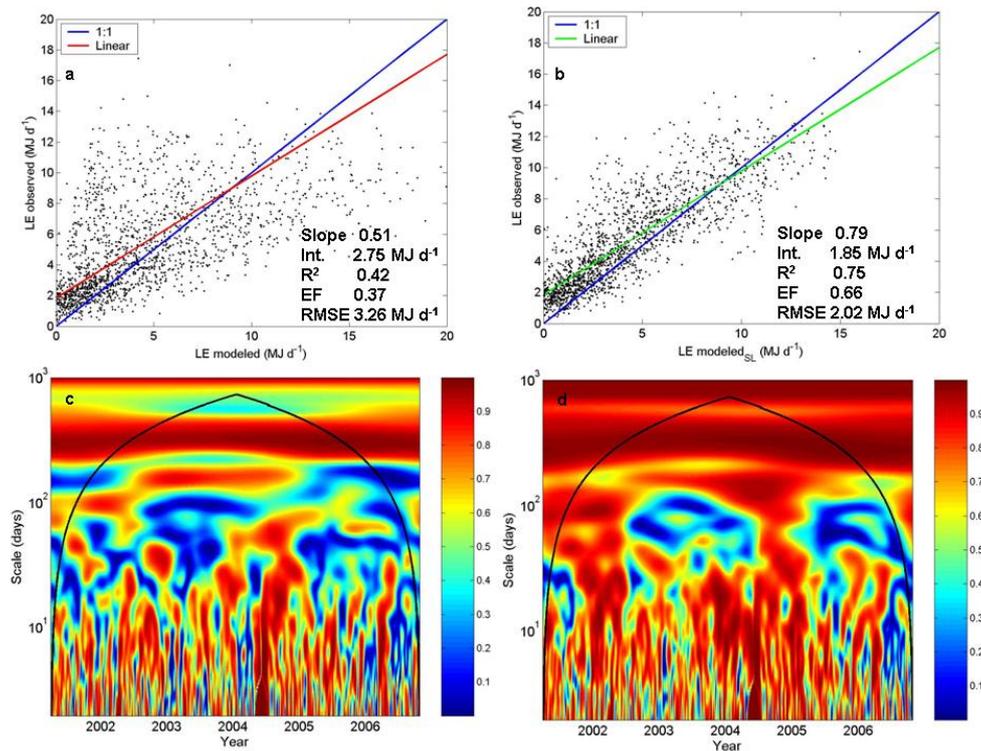
Fig. 5. A comprehensive model evaluation framework using measured latent energy flux (LE) from the Tumbarumba flux monitoring site and corresponding predictions from the CABLE model and model improvement by adding a soil and litter-layer evaporation scheme (CABLESL). Instantaneous (half-hourly) observed versus CABLE **(a)** and CABLESL **(b)** at the Tumbarumba FLUXNET site for 2001–2006. Some common evaluation metrics are also displayed. Int=intercept, RMSE=root-mean-square-error. EF=model efficiency (Meyer and Butler, 1993). The wavelet coherence between observations and the model CABLE **(c)** and CABLESL **(d)** are shown as the full wavelet half-plane. Regions at different time scales (the ordinate) corresponding to different periods of time (the abscissa) in which the model and measurement are significantly related at the 95% confidence interval are those that exceed a coherence value of 0.7.

the measurement record can be identified, and the CABLE model further improved with such a complete "picture", via the display of wavelet coherence in the wavelet half-plane (i.e. Fig. 5c and d).

5.2 Towards a spatially explicit continental and global scale evaluation of LSMs

LSMs are often evaluated at the site level using EC time series, which can provide detailed information on the performance of the temporal dynamics of the model. However, models developed for continental to global applications should be evaluated on the appropriate scale in both time and space. Regional comparisons have largely been neglected partially due to the strong site-by-site focus of flux evaluation to date and partially due to difficulties in "upscaling". Spatially and temporally explicit maps of carbon fluxes can be derived from empirical models that use FLUXNET measurements in conjunction with remotely sensed and/or meteorological data. Different upscaling approaches exist, ranging from simple regression models, to complex machine learning algorithms (Papale and Valentini, 2003; Yang et al., 2007),

The results of such data-oriented models can be used to evaluate process-oriented LSMs operating at regional scales. However, the success of the evaluation depends on whether there is sufficient confidence in the empirical upscaling results, or in other words, if the error of the empirical upscaling is substantially smaller than the error of the LSM simulations.

We compared regional GPP predictions from several data-oriented models against several process-based LSMs. We undertook a comparison of mean annual GPP estimates for 36 major watersheds of Europe between four different data-oriented models (generally using remote sensing data and FLUXNET calibrations) and three process-oriented models (i.e. LSMs, see Table 4 for model details). The comparison revealed a much closer agreement regarding the spatial pattern of GPP among the data-oriented models ($R^2 > 0.58$) than among the process-oriented models. A graphical intercomparison of the models is shown in Fig. 6, which indicates the correlation among model predictions. It is obvious that two of the process-oriented models had substantially poorer fit to the data-oriented estimates than the third, LPJmL. The

Table 4. A description of the data-oriented and process-oriented models used to generate catchment scale predictions of GPP across Europe, which are compared in Fig. 6. Simulations of ORC, LPJmL, BGC, MOD17+, and ANN are from Vetter et al., 2008. MOD17+, ANN, FPA+LC, WBA are calibrated with FLUXNET data.

Acronym	Model name	Model type	Time step	Model drivers	References
ORC	Orchidee	Process-oriented	Sub-daily	Meteorology, Land cover, soil water holding capacity	(Krinner et al., 2005)
LPJ	Lund Potsdam Jena managed Land (LPJmL)	Process-oriented	Daily	Meteorology, Land cover, soil water holding capacity	(Sitch et al., 2003; Bondeau et al., 2007)
BGC	Biome-BGC	Process-oriented	Daily	Meteorology, Land cover, soil water holding capacity	(Thornton, 1998)
MOD17+		Radiation Use efficiency model	Daily	Remote sensing based FAPAR (MODIS), Meteorology, Land cover	(Running et al., 2004)
ANN	Artificial Neural Network	Data-oriented	Daily	Remote sensing based FAPAR (Modis), Meteorology, Land cover	(Papale and Valentini, 2003) updated Vetter et al. (2008)
FPA+LC	FAPAR based Productivity Assessment+Land Cover	Vegetation type specific regression models between FAPAR and GPP	Annual	Remote sensing based FAPAR (SeaWiFS), Land cover	Jung et al., 2009
WBA	Water Balance Approach	Upscaling of water use efficiency	Multi-annual	Remote sensing based LAI (MODIS, CYCLOPES), Meteorology, Land cover, river runoff data	(Beer et al., 2007), updated

Lund Potsdam Jena managed Land model (LPJmL) showed a reasonable agreement (i.e. high R^2) with the different data-oriented results because it has a realistic implementation of crops, in contrast to Biome-BGC and Orchidee, which simulate crops simply as productive grasslands. This treatment appears to be inadequate, as active management and nutrient manipulation differ between a natural grassland and a managed agricultural area. Confidence in the data-oriented models is gained by the agreement among independent upscaling approaches.

6 Model-data fusion

Rather than compare model outputs with a particular dataset in order to test or calibrate the model, model-data fusion techniques combines the two sources of information, model and data, into a single product. We define the term "model-data fusion", sometime called "data assimilation", to include both state estimation and parameter estimation. State estimate involves updating the model state vector PDFs according to observations, using techniques such as the Kalman filter (Williams et al., 2005), or variational approaches typical with weather forecasting (Courtier et al., 1994). Parameter estimation involves determination of parameter PDFs using observations. Model-data fusion allows for integrating multiple and different types of data, including the associated uncertainties, and for including prior knowledge on model parameters and/or initial state variables. However, the timescales of model processes and observations must over-
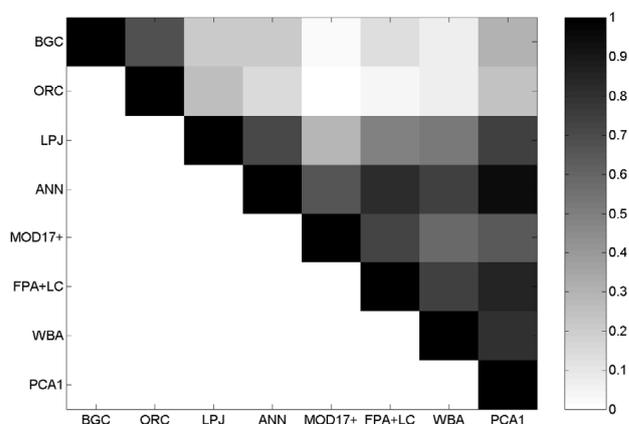
Fig. 6. Model estimates of catchment mean annual GPP for 36 major watersheds in Europe are compared among different process- and data oriented models. The figure shows the R2 value for each paired comparison among the 8 models used (see scale bar on right). PCA1 is the first principal component and can be regarded as the common information content of all models. See Table 4 for abbreviations.

lap, otherwise useful information cannot be conveyed to the model from those data by the assimilation scheme.

The primary objective of model-data fusion is to improve a model's performances by either optimizing/refining the value of the unknown parameters and initial state variables, or by correcting the model's trajectory (state variables) according to a given data set. In the context of the philosophy for

model-data fusion outlined above, the residuals between data and model output after assimilation and the residuals between prior and posterior state and/or parameter values are important information for subsequent steps of model structural development. These residuals help identify model capabilities, particularly those processes which are best constrained by data. MDF techniques also allow for the propagation of uncertainties between the parameter space and the observation space, and thus estimation of model prediction uncertainties, e.g. modelled carbon fluxes (Verbeeck et al., 2006). Below, we first give a brief overview of different MDF techniques with examples to illustrate application and interpretation. We then discuss various problems associated with MDF and suggest solutions.

6.1 Technical implementation

All MDF techniques require certain steps in common. First, the model must be specified, and uncertainties on the model initial states and parameters must be determined a priori. For some algorithms the model uncertainty itself must be estimated (e.g. Kalman filter). One approach for setting model errors in Kalman Filter schemes is to ensure that they are set within bounds that are small enough to avoid tracking daily noise in observations, and large enough to shift over weekly-seasonal timescales in response to process signals. Second, data and their uncertainties must be specified. Third, an observation operator must be defined, which links the state vector to the observations, accounting for intermediate processes (e.g. transport processes), requisite scaling and aggregation (e.g. NEE=R_E−GPP). Fourth, the optimality condition must be defined, whether as some sort of cost function to be minimised, or a joint model-data likelihood. Fifth, the optimal solution is determined. Finally, the posterior probability density functions of the parameters and the state vector must be analysed and interpreted.

The search for an optimal solution and the estimation of uncertainties can be achieved with three different classes of algorithms:

1. "Global search" algorithms, often based on a random generator (Braswell et al., 2005; Sacks et al., 2006). These methods are well adapted for highly non-linear models with multiple minima, but are usually associated with high computational cost because many model evaluations are required. Several algorithms exist, e.g. Metropolis or Metropolis Hastings, genetic algorithms.

2. "Gradient-descent" algorithms, following identified directions in the parameter space. These methods are highly efficient, but are not best suited for highly dimensional nonlinear models as they may end up finding local rather than global minima. To determine posterior uncertainties, they require the calculation of model

output sensitivity to model parameters (Santaren et al., 2007).

3. "Sequential" algorithms, such as the Kalman filter (Williams et al., 2005; Gove and Hollinger, 2006). These methods process data sequentially, in contrast to the first two classes that treat all observations at once (i.e. batch methods).

Combinations of methods can be successful (Vrugt et al., 2005), such as a global search method to get close to the global minimum, followed by a gradient-descent method to pinpoint the minimum more accurately.

The optimization process for both global search and gradient-descent generally requires the calculation of the mismatch between model outputs (M) and observations (O) and the mismatch between initial or "prior" (X_p) parameter values and those later determined to be optimal (X) in a Bayesian framework (Eq. 1). The essence of the Bayesian approach is the multiplication of probabilities; in practice the log of the data-model mismatch (log likelihood) is summed with the log of the prior probabilities. Both mismatches are weighted according to confidence on observations (**R** matrix) and prior knowledge about parameter error covariance (**P** matrix). The **R** matrix may contain different kinds of data, which are matched by their error covariances. The **P** matrix is crucial because it allows the consistent combination of different kinds of data into one formulation of the cost function. H is the observation operator, which relates the model state vector to the observations. The data-model and parameter mismatches define the objective or cost function, J, used in many batch optimisation methods (Lorenc, 1995):

$$J = 1/2[(HM-O)^T \mathbf{R}^{-1}(HM-O)+(X-X_p)^T \mathbf{P}^{-1}(X-X_p)] \quad (1)$$

There are similarities between J and the Kalman gain (K) used in Kalman filters. K is used to adjust the model forecast (M^f) to generate an optimal analysis (M^a) based on the differences between model and observations, and the confidence in model (**P** matrix) and observations (**R** matrix):

$$K = \mathbf{P}^f H^T (H P^f H^T + \mathbf{R})^{-1} \quad (2)$$

$$M^a = M^f + K(O - HM^f) \quad (3)$$

The model error covariance matrix (**P**) is updated dynamically in the Kalman filter as observations are assimilated.

The observation term can take several forms depending on the information that we want to retrieve: either the diurnal cycle with half hourly fluxes, or the seasonal cycle with daily or monthly means. Temporal error correlations (systematic biases) are likely to be severe between sub-daily fluxes, for models and observations. To handle this data redundancy, one can include error correlations in **R** (a rather difficult task), sub-sample the whole data set, or use a mean diurnal cycle over a relatively long period (Santaren et al., 2007).

The cost function defined above (Eq. 1) is commonly used when the error is known or assumed to be Gaussian, but other options are possible. For example, the sum of absolute deviations rather than squared deviations between data and model form the likelihood when data errors are exponential, reducing the influence of outliers in the cost function compared to Eq. (1). The choice of cost function is an important one; an optimization intercomparison by Trudinger et al. (2007) found significantly more variation in estimated parameters due to definition of the cost function than choice of optimization technique. Another important choice in the optimisation is whether to include all parameters at once at the risk that interactions among parameters yield big uncertainties and a poor convergence of the algorithm. Optimizing for just a small set of pre-identified parameters may not include key processes (Wang et al., 2001; Reichstein et al., 2003), and in fact is mathematically equivalent to holding constant those parameters not optimized.

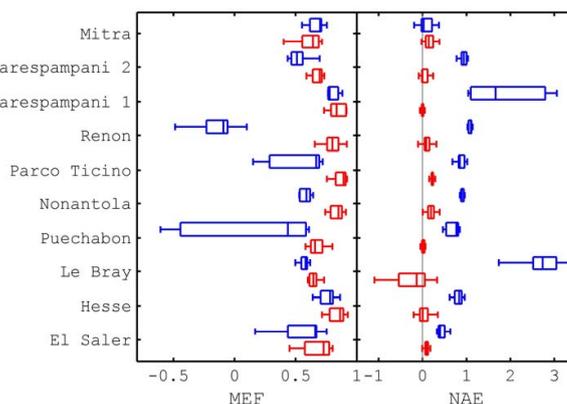### 6.2 Error estimations: random and systematic error

There are two important differences between random and systematic measurement errors. First, whereas statistical analyses can be used to estimate the size of random errors (Lasslop et al., 2008), systematic errors are difficult to detect or quantify in data-based analyses (Moncrieff et al., 1996); but see Lee (1998). Second, random errors in data will increase the uncertainty in data-model fusion parameter estimates and model predictions, but should not bias the posterior estimates. Systematic errors in data, on the other hand, are more insidious in that, if not corrected first, they will directly result in biased posteriors, see Lasslop et al. (2008) for an example.

However, as noted above, systematic errors will be identified by the multiple-constraints approach via model-data mismatch. On the other hand, large random errors in observations, while increasing uncertainties of the estimated parameters, decrease the usefulness of the model-data fusion approach because the analysis errors are not much reduced from the original background (prior) errors.

Estimates of prior values and errors on model parameters is usually undertaken using literature values, global databases (http://www.try-db.org/, Wright et al., 2004) and expert estimation. This critical task partly conditions the realism of the inverse estimates. The role of the user in setting parameter and model errors requires careful inspection of posterior distributions to determine if MDF has succeeded (see Sect. 6.6).
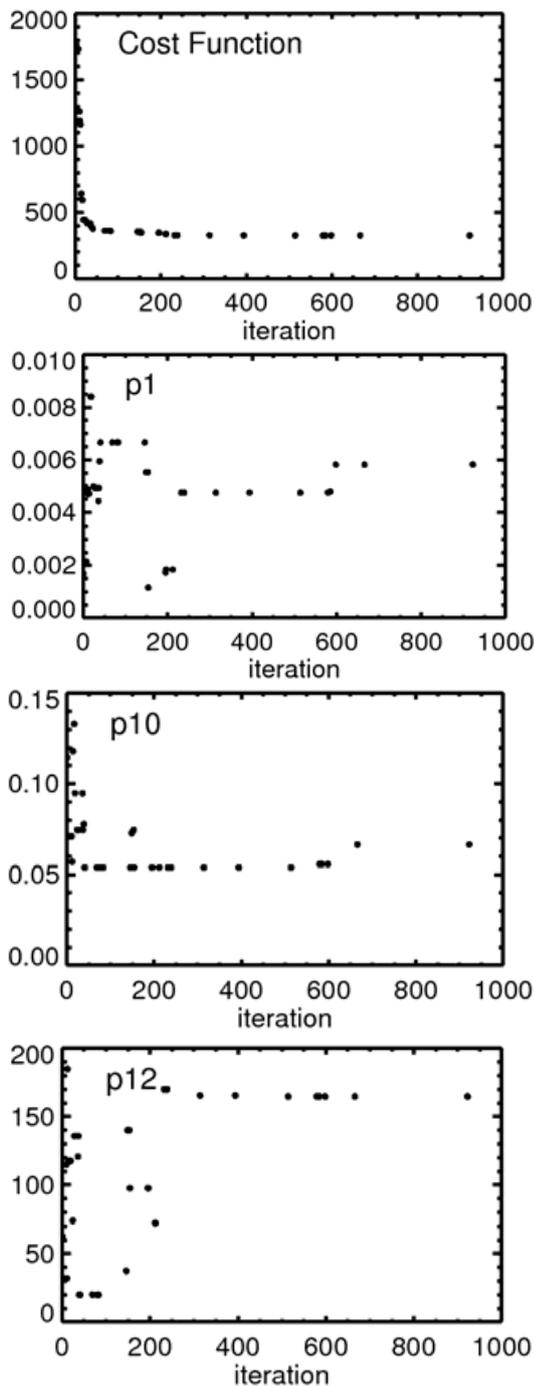
6.3 The initial condition problem

Correct estimates of both vegetation and soil carbon pools constitute an initial condition problem of significant relevance in those LSMs with dynamic biogeochemical (BGC) modelling. The current best practice in LSMs applied globally is to use spin-up routines until the C cycle is in equilibrium with pre-industrial climate, followed by transient industrial climate, atmospheric CO_2 and N deposition runs (Churkina et al., 2003; Thornton et al., 2002; Morales et al., 2005), and then impose historical land-use change and management. The initialization process is often prescribed until ecosystem atmosphere fluxes are in equilibrium, which may lead to overestimation of both soil (Pietsch and Hasenauer, 2006; Wutzler and Reichstein, 2007) and vegetation pools, consequently increasing total respiration (R_E). The equilibrium assumption may be criticised as not being fully applicable to natural ecosystems. In BGC-model-data fusion approaches, both empirical as well as process-based methodologies have been proposed to address these issues (Santaren et al., 2007; Carvalhais et al., 2008; Luo et al., 2001; Pietsch and Hasenauer, 2006). In ecosystems far from equilibrium, model parameter uncertainties and biases can be avoided by relaxing the steady state assumption, i.e. optimizing the initial conditions of C pools (Carvalhais et al., 2008), which reduces model estimates uncertainties in upscaling exercises. As an example, in Fig. 7 we show CASA model performance statistics for estimated net ecosystem production compared against data from ten EC sites. We compared performance with spun-up initial conditions (i.e. steady state) against an approach that relaxed the steady state assumption. Performance after parameter optimisation was higher using the relaxed approach rather than one that considered the system in equilibrium.

The integration of ancillary information on soils and standing biomass pools allows the initial condition problem to be addressed independently from other model evaluation perspectives. Modelled soil pools should be consistent with measured pools. Further, information concerning tree age or DBH distribution in forested sites can be a robust proxy

Fig. 7. CASA model performance statistics for NEP at ten eddy covariance sites is higher in relaxed steady state approaches (red) than in cases considering ecosystems in equilibrium for model optimization (blue) (MEF – model efficiency; and NAE – normalized average error) (Carvalhais et al., 2008).

Fig. 8. Results from a genetic algorithm (GA) applied to parameter estimation in the DALEC model as part of the REFLEX project (Fox et al., 2009). Synthetic, noisy and sparse data were supplied to the GA. As the calculations proceeded, the current best estimate of parameters and the corresponding cost function were saved, and values of the cost function and selected parameters (p1, p10, p12) are shown here.

for woody biomass constraints, for which forest inventories as well as synthesis activities are key contributors (Luyssaert et al., 2007). Disturbance and management history (Thornton et al., 2002) are additional factors that control the initial condition of any particular simulation. There are clearly considerable challenges related to generating the historical data and including historical and current management regimes.

6.4 Equifinality

An important issue in parameter estimation is equifinality, where different model representations, through parameters or model structures, yield similar effects on model outputs, and so can be difficult to distinguish (Medlyn et al., 2005; Beven and Freer, 2001). Such correlations imply that quite different combinations of model parameters or representations can give a similar match of model outputs to observations. This problem is particularly relevant to observations of net fluxes, which on their own may not provide enough information to constrain parameters associated with the component gross fluxes. Further, the associated uncertainties in a posteriori regional upscaling are significant and variable in space and time (Tang and Zhang, 2008) Equifinality in parameter estimation is illustrated in Fig. 8, which shows the results of a genetic algorithm applied to the DALEC model (Williams et al., 2005), as part of the REFLEX project (Fox et al., 2009). The cost function decreased quickly at first, but after about 100 iterations of the genetic algorithm there was only gradual further improvement, yet the values of some parameters varied significantly after this point (Fig. 8). Addressing equifinality requires identification of covariances between parameter estimates, the use of multiple, orthogonal data sets, and testing a variety of cost functions to constrain unidentifiable parameters. Thus, it is critical to assess whether the available observations can adequately constrain the model parameters or whether more information is required.

6.5 Parameter validation and uncertainty propagation

Estimated parameters should always be interpreted with great care. Posterior values are not always directly interpretable (i.e. in a biological sense) due to equifinality (see above) or to compensating mechanisms related to model deficiencies/biases. In addition, simplified model structures require aggregated parameters (e.g. a $V_{c\,max}$ – maximum rate of carboxylation – value of a "big leaf") that will always be different from observed parameter values (e.g. determined from measurements of $V_{c\,max}$ on an individual leaf). Nevertheless, comparison with independent real measured parameter values is an important part of the analysis and validation process. For example, Santaren et al. (2007) successfully compared their estimated $V_{c\,max}$ with estimates from leaf-level measurements (their Fig. 7).

A benefit of most model-data fusion approaches is to estimate parameter uncertainties and the corresponding model

output uncertainties. The outcome of the MDF is a set of parameter PDFs which can then be used to generate an ensemble of model runs, using time series of climate forcing data. As an example, we generated synthetic, noisy and sparse data from the DALEC model using nominal parameters, and supplied these to an Ensemble Kalman filter (Williams et al., 2005) for parameter estimation. The outputs of the EnKF were parameter frequency distributions that indicated the statistics of the retrieved parameters generated by inverting the observations. In Fig. 9 we show the retrieved PDFs for the turnover rates for four C pools, which can be compared to the true parameter values (vertical lines), and the prior parameter values, which are represented by the width of the x-axes. The process of estimating confidence intervals of model states and predictions is highly dependent on the prior estimates of parameter (and model) error, as shown in the REFLEX experiment (Fox et al., 2009). We are still learning how to properly quantify confidence intervals on parameter and state analyses.

6.6 How to assess a model-data fusion scheme

A first step to testing if MDF approaches are effective is to use synthetic data, where the underlying "true" state of the system and model parameters are known. A synthetic truth is generated by running the model with given parameters, noise is added, and data are thinned. These data are provided to the MDF scheme to test estimation of parameters and retrieval of C fluxes, all of which are known (Fig. 9). A second step is to examine posterior parameter distributions relative to priors. Have parameters been constrained? Is there evidence that parameter priors were correct? For example, in the synthetic case shown in Fig. 9 it is clear that turnover rates of foliage and soil organic matter were better constrained by NEE data, with PDFs concentrated around the true parameter values used to generate the synthetic data. Posterior turnover rates for fine roots and wood are barely different from the priors, with broad distributions spanning the prior range. A third step is to check model residuals on NEE (or any other observations), to see if they are Gaussian and not autocorrelated, a typical hypothesis of Bayesian approaches (see above). If multiple data time series are used, are they all consistent with the model, or do they reveal potential biases in data or model (Williams et al., 2005)? It is useful to iterate the optimization process from a different starting point (initial conditions) to see if the posteriors are similar. Testing different assumptions in the MDF is also useful, for instance, uniform versus Gaussian priors, altered model error estimates in KF schemes etc.

6.7 Spatial and temporal parameter variability

With increasing FLUXNET data availability over a wide range of ecosystems, model-data fusion approaches can improve our understanding and process representation ability
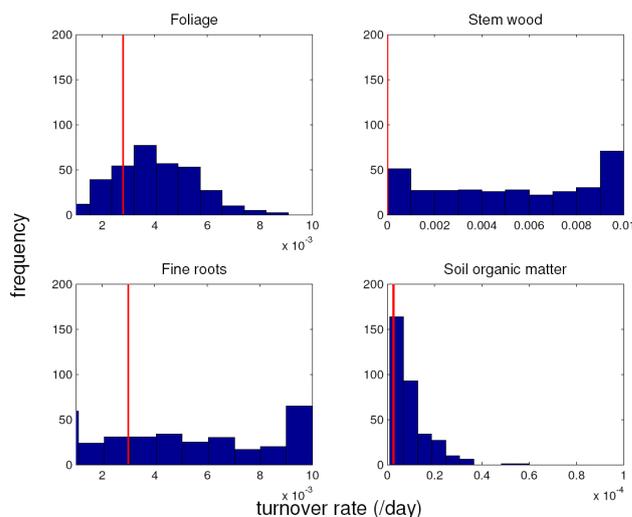

Fig. 9. Parameter retrieval from a synthetic experiment using the DALEC model (Williams et al., 2005). Synthetic, noisy and sparse data generated from DALEC were supplied to an Ensemble Kalman filter (Williams et al., 2005) for parameter estimation. "True" parameter values for turnover rates of foliage, stem wood, fine roots and soil organic matter are shown by the red lines. The frequency distributions indicate the statistics of the retrieved parameters generated by inverting the observations. The x-axis limits indicate the spread of the prior estimates for each parameter.

at regional and global scales. So far most biogeochemical models use the concept of "plant functional type" to represent the ecosystem diversity, with typically 6 to 15 classes. Whether the associated parameters or process representations are generic enough to apply across climate regimes, species differences, as well as large temporal domains or whether they should be refined are critical questions that FLUXNET data can help resolve.

6.7.1 Temporal variability

For a prognostic model to be useful, dynamic processes must be represented within the model, and parameters must be constant in time. If there is evidence that parameters vary over time (Hui et al., 2003; Richardson et al., 2007), it means the parameters are sensitive and that some component of model structure is missing. The relevant structure must be included in the model for it to have prognostic value. Likewise, if there is no evidence that parameters vary, then they are shown to be insensitive. The estimation of process parameters in variable conditions over time is likely to give new insights in processes that are poorly understood (Santaren et al., 2007). Building on the work of Santaren et al. (2007), we optimised the ORCHIDEE biogeochemical model against eddy covariance data to retrieve the year-to-year variability of maximum photosynthetic capacity ($V_{c\,max}$). Significant interannual variations in $V_{c\,max}$ for the Aberfeldy and
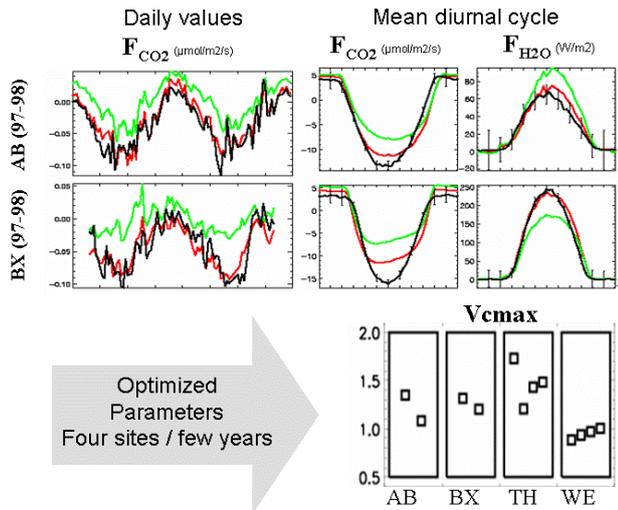
Fig. 10. Some results from the optimisation of 18 parameters of the ORCHIDEE model at 4 temperate needleleaf forest sites for several years: 2 years at le Bray (BX) and Aberfeldy (AB) and 4 years at Tharandt (TH) and Weilden (WE). Green is prior ORCHIDEE model, red is optimised model, black are the data. One parameter set is optimized each year using NEE, latent, sensible and net radiation fluxes (using half hourly values, see Santaren et al., 2007). Upper panels show the model-data fit (daily values and mean diurnal cycle during the growing season) while the lower panel shows the estimated Vcmax parameter values for each sites and each year: values a reported with respect to the prior value.

Tharandt sites suggest there area additional processes related to productivity not captured by ORCHIDEE, and possibly linked to the N cycle (Fig. 10). Similarly, Reichstein et al. (2003) studied drought events by estimating the seasonal time course of parameters of the PROXEL model using eddy covariance carbon and water fluxes at Mediterranean ecosystems.

6.8 Spatial variability

The concept of plant functional type (PFT) simplifies the representation of ecosystem functioning and groups model parameter values estimated at site level. However, inter-site parameter variability for analogous PFTs show limitations of such a regionalization approach (Ibrom et al., 2006; van Dijk et al., 2005). Parameters that show large between-site but low temporal variability should be used to help define more realistic PFTs than are currently applied. The spatial variation in parameters should be explicable on the basis of external drivers such as climatic, topographic or geological variability. Model parameter regionalization approaches usually build on remotely sensed variables as model drivers which are providing a basis for the extrapolation of spatial patterns, but the quantification of the FLUXNET representativeness and heterogeneity is fundamental to assess the upscaling potential of both model parameters and observed processes.

6.9 Evaluation of model deficiencies and benchmarking

One of the major interests of assimilation techniques for modellers is in the potential to highlight model deficiencies. For this objective, being able to include all sources of uncertainty within a rigorous statistical framework represents a critical advantage. Abramowitz et al. (2005, 2008, 2007) used artificial neural networks (ANN) linking meteorological forcing and measured fluxes to provide benchmarks against which to assess performance of several LSMs. Performance of the ANNs was generally superior to that of the LSMs, which produced systematic errors in the fluxes of sensible heat, water vapour and CO_2 at five test FLUXNET sites. These errors could not be eliminated through parameter optimization and it appears that structural improvements in the LSMs are needed before their performance matches that of the benchmark ANNs. Predictions using LSMs may be satisfactory in some parts of bioclimatic space and not in others. Abramowitz et al. (2008) used self-organizing maps to divide climate space into a set distinct regions to help identify conditions under which model bias is greatest, thereby helping with the "detective work" of model improvement.

With conventional model-data evaluation (e.g. RMSE analyses) there is always the risk of over-interpretation of a particular model-data mismatch, that only reflects poor parameter calibrations. The details of these outcomes are probably only relevant to a given model structure and thus cannot be extrapolated to other models. As an example, Fig. 10 illustrates the difficulty of one model to properly capture the amplitude of both the seasonal cycle and the summer diurnal cycle of NEE (and not for the latent heat flux). Note that the issue of temporal and spatial variability, discussed above, is also relevant to this objective. However, a caution with any parameter optimization process is not to correct for model biases with unrealistic parameter values, by carefully evaluating the posterior parameter estimates. Model failure after optimisation identifies inadequate model structure, which ultimately leads to improvements in process representation and increased confidence in the model's projections.

7 Conclusions

There has been significant progress in LSM development, in generating flux data with error assessments from key biomes, and in coupling models and data for optimal analyses. We have shown how the FLUXNET database can be used to improve forecasts of global biogeochemical and climate models. MDF approaches provide a means to identify which components of the carbon, water and energy balance are effectively constrained by current FLUXNET data and how uncertainty grows with prognostic runs. There are now some clear opportunities for LSM and FLUXNET communities to develop this research.

Our vision for the future includes: (1) identifying and funding critical new locations for FLUXNET towers in poorly studied biomes, for instance in tropical croplands; (2) nesting EC towers within the regional footprints of tall towers that sample the CO_2 concentration of the planetary boundary layer (Helliker et al., 2004) to assist upscaling; (3) linking EC towers effectively with the atmospheric column CO_2 measurements generated by the satellites OCO and GOSAT, expected from late 2009 (Feng et al., 2008); (4) continued linkage to optical products from remote sensing, as these provide critical information on canopy structure and phenology (Demarty et al., 2007); (5) use of plant functional traits database to identify more realistic priors and to validate posteriors in the data assimilation process (Luyssaert et al., 2007). Full exploitation of this vision is dependent on model-data fusion. Atmospheric transport models are critical for linking land surface exchanges to CO_2 concentration measurements from (2) and (3). Radiative transfer models are vital for properly assimilating the raw optical information from remote sensing (Quaife et al., 2008).

We have identified five major model-data fusion challenges for the FLUXNET and LSM communities to tackle, for improved assessment of current and future land surface exchanges of matter and energy:

- To determine appropriate use of current data and to explore the information gained in using longer time series regarding future prediction. What can be learned from assimilating 10+ years of EC data?

- To avoid confounding effects of missing processes in model representation on parameter estimation.

- To assimilate more data types (e.g. pools/stocks of carbon, earth observation data) and to define improved observation operators. It would be valuable to determine which biometric time series would best complement EC data at FLUXNET sites.

- To fully quantify uncertainties arising from data bias, model structure, and estimates of initial conditions. We believe that MDF with multiple independent data time series will best identify bias, and recommend a multi-site test of this idea.

- To carefully test current model concepts (e.g. PFTs) and guide development of new concepts. FLUXNET data can be used to determine whether parameters need to vary continuously in space in response to some remotely sensed trait, following approaches applied, for instance, by Williams et al. (2006). Thus, FLUXNET data can be used to challenge and enrich the PFT approach, possibly leading to a redefinition of PFTs.

References

Abramowitz, G.: Towards a benchmark for land surface models, Geophys. Res. Lett., 32, L22702, doi:10.1029/2005GL024419, 2005.

Abramowitz, G. and Pitman, A.: Systematic bias in land surface models, J. Hydrometeorol., 8, 989–1001, 2007.

Abramowitz, G. and Gupta, H.: Toward a model space and model independence metric, Geophys. Res. Lett., 35, L05705, 10.1029/2007GL032834, 2008.

Aubinet, M., Grelle, A., Ibrom, A., Rannik, U., Moncrieff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grunwald, T., Morgenstern, K., Pilegaard, K., and Rebmann: Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology, Adv. Ecol. Res., 30, 113–175, 2000.

Baldocchi, D., Falge, E., Gu, L. H., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X. H., Malhi, Y., Meyers, T., Munger, W., Oechel, W., U, K. T. P., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, B. Am. Meteor. Soc., 82, 2415–2434, 2001.

Baldocchi, D. D.: Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: past, present and future, Glob. Ch. Biol., 9, 479–492, 2003.

Beer, C., Reichstein, M., Ciais, P., Farquhar, G. D., and Papale, D.: Mean annual GPP of Europe derived from its water balance, Geophys. Res. Lett., 34, L05401, doi: 10.1029/GL029006, 2007.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249, 11–29, 2001.

Bonan, G. B.: Land atmosphere CO_2 exchange simulated by a land-surface process model coupled to an atmospheric general-circulation model, J. Geophys. Res.-Atmos., 100, 2817–2831, 1995.

Bonan, G. B.: Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests, Science, 320, 1444–1449, doi:10.1126/science.1155121, 2008.

Bondeau, A., Smith, P. C., Zaehle, S., Schaphoff, S., Lucht, W., Cramer, W., and Gerten, D.: Modelling the role of agriculture for the 20th century global terrestrial carbon balance, Glob. Change Biol., 13, 679–706, 2007.

Braswell, B. H., Sacks, W. J., Linder, E., and Schimel, D. S.: Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, Glob. Change Biol., 11, 335–355, 2005.

Carvalhais, N., Reichstein, M., J., S., G.J., C., Santos Pereira, J., Berbigier, P., Carrara, A., Granier, A., Montagnani, L., Papale, D., Rambal, S., Sanz, M. J., and Valentini, R.: Implications of Carbon Cycle Steady State Assumptions for Biogeochemical Modeling Performance and Inverse Parameter Retrieval, Global Biogeochem. Cy., 22, GB2007, doi:10.1029/2007GB003033, 2008.

Churkina, G., Tenhunen, J., Thornton, P., Falge, E. M., Elbers, J. A., Erhard, M., Grunwald, T., Kowalski, A. S., Rannik, U., and Sprinz, D.: Analyzing the ecosystem carbon dynamics of four European coniferous forests using a biogeochemistry model, Ecosystems, 6, 68–184, 2003.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-VAR, using an incremental approach, Q. J. Roy. Meteor. Soc., 120, 1367–1387, 1994.

Demarty, J., Chevallier, F., Friend, A. D., Viovy, N., Piao, S., and Ciais, P.: Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model, Geophys. Res. Lett., 34, L15402, doi:10.1029/2007GL030014, 2007.

Falge, E., Baldocchi, D., and Olson, R. J.: Gap Filling Strategies for Defensible Annual Sums of Net Ecosystem Exchange, Ag. For. Meteorol., 107, 43–69, 2001.

Feng, L., Palmer, P. I., Bösch, H., and Dance, S.: Estimating surface CO_2 fluxes from space-borne CO_2 dry air mole fraction observations using an ensemble Kalman Filter, Atmos. Chem. Phys., 9, 2619–2633, 2009,
http://www.atmos-chem-phys.net/9/2619/2009/.

Foken, T. and Wichura, B.: Tools for quality assessment of surface-based flux measurements, Ag. For. Meteorol., 78, 83–105, 1996.

Foley, J. A., Prentice, I. C., Ramunkutty, N., Levis, S., Pollard, D., Sitch, S., and Haxeltine, A.: An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics, Global Biogeochem. Cy., 10, 603–628, 1996.

Fox, A. M., Williams, M., Richardson, A. D., Cameron, D., Gove, J., Reichstein, M., Quaife, T., Ricciuto, D., Tomelleri, E., Trudinger, C. M., and Van Wijk, M. T.: The REFLEX project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data, Ag. For. Meteorol., in press, 2009.

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., Yoshikawa, C., and Zeng, N.: Climate-carbon cycle feedback analysis: Results from the (CMIP)-M-4 model intercomparison, J. Climate, 19, 3337–3353, 2006.

Gillespie, A. J. R.: Rationale for a National Annual Forest Inventory Program, J. Forestry, 97, 16–20, 1999.

Goulden, M. L., Munger, J. W., Fan, S.-M., Daube, B. C., and Wofsy, S. C.: Measurements of carbon storage by long-term eddy correlation: Methods and a critical evaluation of accuracy, Global Change Biol., 2, 169–182, 1996.

Gove, J. H. and Hollinger, D. Y.: Application of a dual unscented Kalman filter for simultaneous state and parameter estimation in problems of surface-atmosphere exchange, J. Geophys. Res., 111, D08S07, doi:10.1029/2005JD00621, 2006.

Grinsted, A., Moore, J. C., and Jevrejeva, S.: Application of the cross wavelet transform and wavelet coherence to geophysical time series, Nonlin. Processes Geophys., 11, 561–566, 2004, http://www.nonlin-processes-geophys.net/11/561/2004/.

Hargrove, W., Hoffman, F., and Law, B.: New analysis reveals representativeness of the AmeriFlux network, Eos Trans. AGU, 82, p. 529, p. 525, 2003.

Haverd, V., Leuning, R., Griffith , D. W., van Gorsel, E., and Cuntz, M.: The Lagrangian time scale for turbulent transport in forest canopies, determined from measured fluxes and concentrations and modelled source distributions, Bound.-Lay. Meteorol., 130, 209–228, 2009.

Helliker, B. R., Berry, J. A., Betts, A. K., Bakwin, P. S., Davis, K. J., Denning, A. S., Ehleringer, J. R., Miller, J. B., Butler, M. P., and Ricciuto, D. M.: Estimates of net CO2 flux by application of equilibrium boundary layer concepts to CO_2 and water vapor measurements from a tall tower, J. Geophys. Res.-Atmos., 109, D20106, doi:10.1029/2004JD004532, 2004.

Hollinger, D. Y. and Richardson, A. D.: Uncertainty in eddy covariance measurements and its application to physiological models, Tree Phys., 25, 873–885, 2005.

Hui, D., Luo, Y., and Katul, G.: Partitioning interannual variability in net ecosystem exchange between climatic variability and function changes, Tree Phys., 23, 433–442, 2003.

Ibrom, A., Jarvis, P. G., Clement, R., Morgenstern, K., Oltchev, A., Medlyn, B. E., Wang, Y. P., Wingate, L., Moncrieff, J. B., and Gravenhorst, G.: A comparative analysis of simulated and observed photosynthetic CO2 uptake in two coniferous forest canopies, Tree Phys., 26, 845–864, 2006.

IPCC: Climate Change 2007 Synthesis Report, Cambridge University Press, 1009 pp., 2007.

Jung, M., Vetter, M., Herold, M., Churkina, G., Reichstein, M., Zaehle, S., Ciais, P., Viovy, N., Bondeau, A., Chen, Y., Trusilova, K., Feser, F., and Heimann, M.: Uncertainties of modeling gross primary productivity over Europe: A systematic study on the effects of using different drivers and terrestrial biosphere models, Global Biogeochem. Cy., 21, GB4021, doi:10.1029/2006GB002915, 2007.

Kowalczyk, E. A., Wang, Y. P., Davies, R. M. L. L., McGregor, J. L., and Abramowitz, G.: The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model, CSIRO Mar. Atmos. Res., 13, 13 pp., 2006.

Krinner, G., Viovy, N., de Noblet-Ducoudre, N., Ogee, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cy., 19, GB1015, doi:10.1029/2003GB002199, 2005.

Lasslop, G., Reichstein, M., Kattge, J., and Papale, D.: Influences of observation errors in eddy flux data on inverse model parameter estimation, Biogeosciences, 5, 1311–1324, 2008,
http://www.biogeosciences.net/5/1311/2008/.

Lee, X.: On micrometeorological observations of surface-air exchange over tall vegetation, Agr. Forest Meteorol., 91, 39–49, 1998.

Leuning, R., Cleugh, H. A., Zegelin, S. J., and Hughes, D.: Carbon and water fluxes over a temperate Eucalyptus forest and a tropical wet/dry savanna in Australia: measurements and comparison with MODIS remote sensing estimates, Ag. Forest Meteorol., 129, 151–173, 2005.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water. Resour.

Res., 43, W07401, doi:10.1029/2006WR005756, 2007.

Lorenc, A. C.: Atmospheric Data Assimilation, UK Meteorological Office, BracknellScientific Paper No. 34, 15 pp., 1995.

Luo, Y. Q., Wu, L. H., Andrews, J. A., White, L., Matamala, R., Schafer, K. V. R., and Schlesinger, W. H.: Elevated CO_2 differentiates ecosystem carbon processes: Deconvolution analysis of Duke Forest FACE data, Ecol. Monogr., 71, 357–376, 2001.

Luo, Y. Q., White, L. W., Canadell, J. G., DeLucia, E. H., Ellsworth, D. S., Finzi, A., Lichter, J., and Schlesinger, W. H.: Sustainability of terrestrial carbon sequestration: A case study in Duke Forest with inversion approach, Global Biogeochem. Cy., 17, 1021, doi:10.1029/2002GB001923, 2003.

Luyssaert, S., Inglima, I., Jung, M., Richardson, A. D., Reichsteins, M., Papale, D., Piao, S. L., Schulzes, E. D., Wingate, L., Matteucci, G., Aragao, L., Aubinet, M., Beers, C., Bernhoffer, C., Black, K. G., Bonal, D., Bonnefond, J. M., Chambers, J., Ciais, P., Cook, B., Davis, K. J., Dolman, A. J., Gielen, B., Goulden, M., Grace, J., Granier, A., Grelle, A., Griffis, T., Grunwald, T., Guidolotti, G., Hanson, P. J., Harding, R., Hollinger, D. Y., Hutyra, L. R., Kolar, P., Kruijt, B., Kutsch, W., Lagergren, F., Laurila, T., Law, B. E., Le Maire, G., Lindroth, A., Loustau, D., Malhi, Y., Mateus, J., Migliavacca, M., Misson, L., Montagnani, L., Moncrieff, J., Moors, E., Munger, J. W., Nikinmaa, E., Ollinger, S. V., Pita, G., Rebmann, C., Roupsard, O., Saigusa, N., Sanz, M. J., Seufert, G., Sierra, C., Smith, M. L., Tang, J., Valentini, R., Vesala, T., and Janssens, I. A.: CO_2 balance of boreal, temperate, and tropical forests derived from a global database, Global Change Biol., 13, 2509–2537, 2007.

MEA: Ecosystems and Human Well-Being, Millennium Ecosystem Assessment, Island Press, 948 pp., 2005.

Medlyn, B. E., Robinson, A. P., Clement, R., and McMurtrie, R. E.: On the validation of models of forest CO_2 exchange using eddy covariance data: some perils and pitfalls, Tree Physiol., 25, 839–857, 2005.

Meyer, D. G. and Butler, D. G.: Statistical validation, Ecol. Mod., 68, 21–32, 1993.

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

Moncrieff, J., Malhi, Y., and Leuning, R.: The propagation of errors in long-term measurements of land-atmosphere fluxes of carbon and water, Global Change Biol., 2, 231–240, 1996.

Morales, P., Sykes, M. T., Prentice, I. C., Smith, P., Smith, B., Bugmann, H., Zierl, B., Friedlingstein, P., Viovy, N., Sabate, S., Sanchez, A., Pla, E., Gracia, C. A., Sitch, S., Arneth, A., and Ogee, J.: Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes, Global Change Biol., 11, 2211–2233, 2005.

Papale, D. and Valentini, A.: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization, Global Change Biol., 9, 525–535, 2003.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and Yakir, D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algo-

rithms and uncertainty estimation, Biogeosciences, 3, 571–583, 2006,
http://www.biogeosciences.net/3/571/2006/.

Pietsch, S. A. and Hasenauer, H.: Evaluating the self-initialization procedure for large-scale ecosystem models, Glob. Ch. Biol., 12, 1658–1669, 2006.

Prentice, I. C., Cramer, W., Harrison, S. P., Leemans, R., Monserude, R. A., and Solomon, A. M.: A global biome model based on plant physiology and dominance, soil properties and climate, J. Biogeogr., 19, 117–134, 1992.

Quaife, T., Lewis, P., De Kauwe, M., Williams, M., Law, B. E., Disney, M. D., and Bowyer, P.: Assimilating Canopy Reflectance data into an Ecosystem Model with an Ensemble Kalman Filter, Remote Sens. Environ., 111, 1347–1364, 2008.

Rastetter, E. B.: The collision of hypotheses: What can be learned from comparisons of ecosystem models?, in: Models in Ecosystem Science, edited by: Canham, C. D., Cole, J. J., and Lauenroth, W. K., Princeton University Press., Princeton (NJ), 211–224, 2003.

Raupach, M. R., Rayner, P. J., Barrett, D. J., DeFries, R. S., Heimann, M., Ojima, D. S., Quegan, S., and Schmullius, C. C.: Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications, Global Change Biol., 11, 378–397, 2005.

Reichstein, M., Tenhunen, J., Roupsard, O., Ourcival, J.-M., Rambal, S., Miglietta, F., Peressotti, A., Pecchiari, M., Tirone, G., and Valentini, R.: Inverse modeling of seasonal drought effects on canopy CO_2//H_2O exchange in three Mediterranean ecosystems, J. Geophys. Res., 108(D23), 4726, doi:10.1029/2003JD003430,, 2003.

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

Richardson, A. D. and Hollinger, D. Y.: Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models, Ag. Forest Meteorol., 131, 191–208, 2005.

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

Richardson, A. D., Hollinger, D. Y., Aber, J. D., Ollinger, S. V., and Braswell, B. H.: Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange, Global Change Biol., 13, 788–803, 2007.

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

Sacks, W. J., Schimel, D. S., Monson, R. K., and Braswell, R. H.:

Model-data synthesis of diurnal and seasonal CO_2 fluxes at Niwot Ridge, Colorado, Global Change Biol., 12, 240–259, 2006.

Santaren, D., Peylin, P., Viovy, N., and Ciais, P.: Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France, Global Biogeochem. Cy., 21, GB2013, doi:10.1029/2006GB002834, 2007.

Savage, K. E., Davidson, E. A., and Richardson, A. D.: A conceptual and practical approach to data quality and analysis procedures for high frequency soil respiration measurements, Funct. Ecol., 22, 1000–1007, 2008.

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., Collatz, G. J., Denning, A. S., Mooney, H. A., Nobre, C. A., Sato, N., Field, C. B., and Henderson-Sellers, A.: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere., Science, 275, 502–509, 1997.

Siqueira, M. B. S., Katul, G. G., Sampson, D. A., Stoy, P. C., Juang, J.-Y., McCarthy, H. R., and Oren, R.: Multi-scale model intercomparisons of CO_2 and H_2O exchange in a maturing southeastern US pine forest, Global Change Biol., 12, 1189–1207 2006.

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Global Change Biol., 9, 161–185, 2003.

Tang, J. and Zhang, Q.: Equifinality in parameterization of process-based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics, J. Geophys. Res., 113, G04010, doi:10.1029/2008JG000757, 2008.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, 2001.

Thornton, P. E.: Description of a numerical simulation model for predicting the dynamics of energy, water, carbon, and nitrogen in a terrestrial ecosystem, University of Montana, Missoula, MT, 280 pp., 1998.

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

Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M., Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y. P., Briggs, P., Barrett, D., and Nikolova, S.: OptIC project: An intercomparison of optimization techniques for parameter estimation in terrestrial biogeochemical models, J. Geophys. Res.-Biogeosciences, 112, G02027, doi:10.1029/2006JG000367, 2007.

van Dijk, A. I. J. M., Dolman, A. J., and Schulze, E.-D.: Radiation, temperature, and leaf area explain ecosystem carbon fluxes in boreal and temperate European forests, Global Biogeochem. Cy., 19, GB2029, doi:2010.1029/2004GB002417, 2005.

Verbeeck, H., Samson, R., Verdonck, F., and Lemeur, R.: Parameter sensitivity and uncertainty of the forest carbon flux model FORUG: a Monte Carlo analysis, Tree Phys., 26, 807–817, 2006.

Verma, S. B.: Micrometeorological methods for measuring surface fluxes of mass and energy, Remote sensing reviews, 5, 99-115, 1990.

Vetter, M., Churkina, G., Jung, M., Reichstein, M., Zaehle, S., Bondeau, A., Chen, Y., Ciais, P., Feser, F., Freibauer, A., Geyer, R., Jones, C., Papale, D., Tenhunen, J., Tomelleri, E., Trusilova, K., Viovy, N., and Heimann, M.: Analyzing the causes and spatial pattern of the European 2003 carbon flux anomaly using seven models, Biogeosciences, 5, 561–583, 2008, http://www.biogeosciences.net/5/561/2008/.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water. Resour. Res., 41, W01017, doi:10.1029/2004WR003059, 2005.

Wang, Y. P., Leuning, R., Cleugh, H. A., and Coppin, P. A.: Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful?, Global Change Biol., 7, 495–510, 2001.

White, M. A., Thornton, P. E., Running, S. W., and Nemani, R. R.: Parameterization and sensitivity analysis of the BIOME-BGC terrestrial ecosystem model: net primary production controls, Earth Interactions, 4, 1–85, 2000.

Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Wofsy, S. C., Shaver, G. R., Melillo, J. M., Munger, J. W., Fan, S.-M., and Nadelhoffer, K. J.: Modelling the soil-plant-atmosphere continuum in a *Quercus-Acer* stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties, Plant Cell Environ., 19, 911–927, 1996.

Williams, M., Schwarz, P., Law, B. E., Irvine, J., and Kurpius, M. R.: An improved analysis of forest carbon dynamics using data assimilation, Global Change Biol., 11, 89–105, 2005.

Williams, M., Street, L., Wijk, M. T. V., and Shaver, G. R.: Identifying differences in carbon exchange among arctic ecosystem types, Ecosystems, 9, 288–304, 2006.

Wilson, K., Goldstein, A., Falge, E., Aubinet, M., Baldocchi, D., Berbigier, P., Bernhofer, C., Ceulemans, R., Dolman, H., Field, C., Grelle, A., Ibrom, A., Law, B. E., Kowalski, A., Meyers, T., Moncrieff, J., Monson, R., Oechel, W., Tenhunen, J., Valentini, R., and Verma, S.: Energy balance closure at FLUXNET sites, Ag. Forest Meteorol., 113, 223–243, 2002.

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., Midgley, J. J., Navas, M.-L., Niinemets, U., Oleksyn, J., Osada, N., Poorter, H., Poot, P., Prior, L., Pyankov, V. I., Roumet, C., Thomas, S. C., Tjoelker, M. G., Veneklaas, E. J., and Villar, R.: The worldwide leaf economics spectrum, Nature, 428, 821–827, 2004.

Wutzler, T. and Reichstein, M.: Soils apart from equilibrium – consequences for soil carbon balance modelling, Biogeosciences, 4, 125–136, 2007, http://www.biogeosciences.net/4/125/2007/.

Xu, T., White, L., Hui, D., and Luo, Y.: Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction, Global Biogeochem. Cy., 20, GB2007, doi:10.1029/2005GB002468, 2006.

Yang, F., Ichii, K., White, M. A., Hashimoto, H., Michaelis, A. R., Votava, P., Zhu, A. X., Huete, A., Running, S. W., and Nemani,

R. R.: Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through Support Vector Machine approach, Remote Sens. Environ., 110, 109–122, 2007.

Zaehle, S., Sitch, S., Smith, B., and Hatterman, F.: Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics, Global Biogeochem. Cy., 19, GB3020, doi:10.1029/2004GB002395, 2005.