



How errors on meteorological variables impact simulated ecosystem fluxes: a case study for six French sites

Y. Zhao¹, P. Ciais¹, P. Peylin², N. Viovy¹, B. Longdoz³, J. M. Bonnefond⁴, S. Rambal⁵, K. Klump⁶, A. Olioso^{7,8}, P. Cellier⁹, F. Maignan¹, T. Eglin¹, and J. C. Calvet¹⁰

¹Laboratoire des Sciences du Climat et de l'Environnement, UMR8212, CEA-CNRS-UVSQ, 91191 Gif-sur-Yvette, France

²Laboratoire de Biogéochimie Isotopique, LBI, Bâtiment EGER, 78026 Thiverval-Grignon, France

³INRA, Centre INRA de Nancy, UMR1137, Ecologie et Ecophysiologie Forestières, 54280 Champenoux, France

⁴INRA, UMR1263, EPHYSE, 33140, Villenave d'Ormon, France

⁵Dream CEFE-CNRS, 1919 route de Mende, 34293 Montpellier, France

⁶INRA, Grassland Ecosystem Research, 234 Avenue du Brézat, Clermont-Ferrand, 63039, France

⁷INRA, UMR1114, EMMAH, Domaine Saint-Paul, 84914 Avignon Cedex 9, France

⁸UAPV, UMR1114, EMMAH, 33, Rue Lois pasteur, 84000, Avignon, France

⁹INRA, UMR1091, Environnement et Grandes Cultures, 78850 Thiverval-Grignon, France

¹⁰CNRM/GAME, Météo-France, CNRS, 42 avenue Coriolis, 31057 Toulouse Cedex 1, France

Correspondence to: Y. Zhao (yan.zhao@lscce.ipsl.fr)

Received: 29 November 2010 – Published in Biogeosciences Discuss.: 9 March 2011

Revised: 29 April 2012 – Accepted: 7 May 2012 – Published: 11 July 2012

Abstract. We analyze how biases of meteorological drivers impact the calculation of ecosystem CO₂, water and energy fluxes by models. To do so, we drive the same ecosystem model by meteorology from gridded products and by meteorology from local observation at eddy-covariance flux sites. The study is focused on six flux tower sites in France spanning across a climate gradient of 7–14 °C annual mean surface air temperature and 600–1040 mm mean annual rainfall, with forest, grassland and cropland ecosystems. We evaluate the results of the ORCHIDEE process-based model driven by meteorology from four different analysis data sets against the same model driven by site-observed meteorology. The evaluation is decomposed into characteristic time scales. The main result is that there are significant differences in meteorology between analysis data sets and local observation. The phase of seasonal cycle of air temperature, humidity and shortwave downward radiation is reproduced correctly by all meteorological models (average $R^2 = 0.90$). At sites located in altitude, the misfit of meteorological drivers from analysis data sets and tower meteorology is the largest. We show that day-to-day variations in weather are not completely well reproduced by meteorological models, with R^2 between analysis data sets and measured local meteorology going from 0.35

to 0.70. The bias of meteorological driver impacts the flux simulation by ORCHIDEE, and thus would have an effect on regional and global budgets. The forcing error, defined by the simulated flux difference resulting from prescribing modeled instead of observed local meteorology drivers to ORCHIDEE, is quantified for the six studied sites at different time scales. The magnitude of this forcing error is compared to that of the model error defined as the modeled-minus-observed flux, thus containing uncertain parameterizations, parameter values, and initialization. The forcing error is on average smaller than but still comparable to model error, with the ratio of forcing error to model error being the largest on daily time scale (86 %) and annual time scales (80 %). The forcing error incurred from using a gridded meteorological data set to drive vegetation models is therefore an important component of the uncertainty budget of regional CO₂, water and energy fluxes simulations, and should be taken into consideration in up-scaling studies.

1 Introduction

The terrestrial biosphere is a key component of the global carbon cycle that receives large attention in terms of climate change mitigation because of its current carbon sink (Prentice et al., 2001; Schimel et al., 2001) and because of potential positive feedbacks with future climate change (Friedlingstein et al., 2006). Process oriented Terrestrial Biosphere Models (TBMs) are useful tools to quantify and understand carbon flux and pool variability at a range of spatial scales, and to predict the response of ecosystems in response to climate and environmental changes. Global or regional meteorological fields on a grid, generated by numerical weather prediction models such as the European Center for Medium-Range Weather Forecasts (ECMWF) or the National Center for Environmental Prediction (NCEP), or by optimal data interpolation schemes (Mitchell et al., 2009), are commonly used to drive TBMs for regional and global applications. Weather is the main driver of variations in CO₂, H₂O and heat fluxes on short time scales going from days to months (Mahecha et al., 2007). Climate plays a key role in interaction with biotic drivers by controlling fluxes on a seasonal to interannual time scale (Knorr et al., 2005; Peylin et al., 2005; Ciais et al., 2005; Richardson et al., 2007). One of the first studies acknowledging that bias in meteorological drivers affected the estimation of photosynthesis (GPP) by models at regional scale is the one of Jung et al. (2007). They found GPP differences over Europe of 34 % on a seasonal time scale and of 40–60 % on an interannual time-scale, given different drivers. However, how bias in meteorology translates into uncertainty on Net Ecosystem Exchange (NEE), Latent Heat (LH) and Sensible Heat (SH) fluxes has rarely been investigated in a systematic approach (Ricciuto et al., 2009; Sczcypta et al., 2011).

To tackle this problem, we use continuous measurements of CO₂, H₂O and heat fluxes made by an eddy-covariance technique at six flux tower sites in France. The six sites cover three forests, two croplands and one grassland site, growing under contrasted climate conditions. The choice of France as a case study to analyze the effects of meteorological driver biases can be justified because a high-resolution meteorological forcing, SAFRAN (Durand et al., 1993, 1999), is available over the country from the French meteorological service Météo-France on a 8 km by 8 km grid. The SAFRAN regional high resolution forcing can be compared with other products from coarser resolution global weather analyses commonly prescribed as input to TBMs.

The TBM used in this study is ORCHIDEE, a process-oriented model that simulates ecosystem processes and resulting carbon, water and energy fluxes at the time step of a half-hour. This allows explicit calculation of the diurnal variation of ecosystem fluxes and consistency with the 30 min acquisition time step of flux data (Baldocchi et al., 2001; Reichstein et al., 2005). At each eddy-covariance site, meteorological parameters are measured together with CO₂, H₂O,

and heat fluxes. Site observed meteorology will serve as a basis against which meteorological analysis data sets can be compared. For applications limited to sites, local meteorology is the best possible driver for TBMs, although it contains uncertainty (Sect. 2.3) and scale mismatch with the footprint of eddy-covariance measurements (Baldocchi et al., 2001). For modeling regional carbon budgets, meteorological forcing is needed on a grid; hence, modelers cannot avoid using imperfect model data sets. Although there is a scale issue between local observations and gridded data from meteorological models, the comparison at site scale is crucial to assess model performance at regional scale.

The goal of this study is to investigate how the errors on meteorological variables impact simulated ecosystem fluxes. In specific, we address the following questions:

1. How different is meteorology at flux tower locations between local observation and gridded data-products: change to data sets?
2. Is the error of modeled meteorology random or systematic?
3. What is the uncertainty of simulated ecosystem CO₂, water vapour and heat fluxes (here using the ORCHIDEE model) induced by errors in meteorological drivers relative to model errors?
4. What is the sensitivity of ecosystem fluxes simulated by ORCHIDEE to each specific meteorological driver?

These questions are investigated for Net Ecosystem Exchange (NEE), photosynthesis (GPP), Total Ecosystem Respiration (TER), latent (LH) and sensible heat fluxes (SH). The time scales investigated go from hour to year, yet focus on the growing season period.

2 Material and methods

2.1 Eddy-covariance data from six flux towers

The six sites (Table 1) cover a deciduous broad-leaved beech forest (Hesse), a temperate needle-leaved maritime pine forest (Le Bray), a Mediterranean green oak forest (Puéchabon), an extensively managed grassland (Laqueuille), and two intensive cropland sites, one in the Paris region (Grignon) and one in the south of France (Avignon). The sites' climate space distribution over France is shown in Fig. 1. The entire data set represents a total of 42 site x years (Table 1).

- Hesse (HES) is a 40-yr old beech forest in the north-east of France. The growing season spans from 1 May to mid-October. The forest is thinned each 4–5 yr, and approximately 20 % of the basal area is removed each time. The measured stand was thinned in the winters of 1995/1996, 1999/2000 and 2004/2005.

Table 1. Summary of eddy flux observation sites used in this study.

Site name	Hesse (HES)	Le Bray (LBR)	Puéchabon (PUE)	Laqueuille (LQE)	Avignon (AVI)	Grignon (GRI)
Vegetation Class	90 % DBF, 10 % grass	80 % ENF, 20 % grass	90 % EBF, 10 % soil	C3 grass	crop	crop
Dominant Species (Age)	Beech (40 yr)	maritime pine (40 yr)	Mediterranean green oak (70 yr)	Extensively grazed grassland	Rotation, wheat-peas-sorghum	Rotation maize-wheat-barley
Mean annual Biomass (gC m ²)	7000	7850	16 600	300	400	225
Location	7.06° E, 48.67° N	−0.77° E, 44.72° N	3.6° E, 43.74° N	2.75° E, 45.64° N	4.88° E, 43.92° N	1.95° E, 48.84° N
Elevation (m)	300	61	270	1040	32	125
Mean annual temperature(°C)	14,2	13,2	13,2	7,4	14,2	11,1
Annual precipitation (mm)	975	972	900	1081	480	600
Observation Period	1997–2007	1996–2007	2000–2007	2004–2007	2004–2007	2005–2007
Annual NEE (gC m ²)	335	365	235	245	155	235
References	Longdoz et al. (2008)	Delzon and Loustau (2005) Jarosz et al. (2008)	Rambal et al. (2003)	Soussana et al. (2004)	Oliosio et al. (2005)	Cellier et al. (2002)

- Le Bray (LBR) is an even-aged maritime pine forest seeded in 1970, part of the Les Landes forest near the Atlantic ocean. The growing season is almost all year round. The site is managed according to a standard local management strategy and was thinned in 1991, 1996 and 2004.
- Puéchabon (PUE) is a 70-yr old Holm oak forest, typical of Mediterranean southeastern France. It has a Mediterranean climate type. Rainfall mainly occurs during fall and winter with about 80 % between September and April; the summer is very dry. The growing season goes mostly from March to mid-August.
- Laqueuille (LQE) is an extensively managed grassland located in Massif Central (central France). The growing season goes from the end of April until October. During that period the grassland is lightly grazed continuously and no fertilizer is applied.
- Avignon (AVI) is a long established agricultural site located in southeastern France. Durum wheat, peas and durum wheat are the rotation crops grown during 2003/2004, 2004/2005 and 2005/2006, respectively. All

are winter crops and their harvest date is at the end of June. In 2007 sorghum, a C4 summer crop, was grown and harvested in the middle of October. Irrigation is applied in particular to sorghum and peas. In the following, we focus the comparison of AVI fluxes with ORCHIDEE simulations (winter C3 crop type) for the periods of winter crop cultivation.

- Grignon (GRI) is an intensive cropland site situated in the Paris area. The rotation was maize-wheat-barley in 2005, 2006 and 2007, with mustard as an intercrop between barley and maize. Maize is seeded by early May and harvested at the end of September. Wheat and barley are seeded in the middle of October and harvested around early-to-mid July. The site is managed with superficial tillage and a slurry application every three years at mustard sowing.

The growing season (GS) and peak growing season (PGS) are site- and definition- dependent. We define GS as the period going from 1 May to 30 September for all sites, except for PUE where it is from 1 March to 31 August. The PGS, the two-month period after GPP reaches its peak, spans from

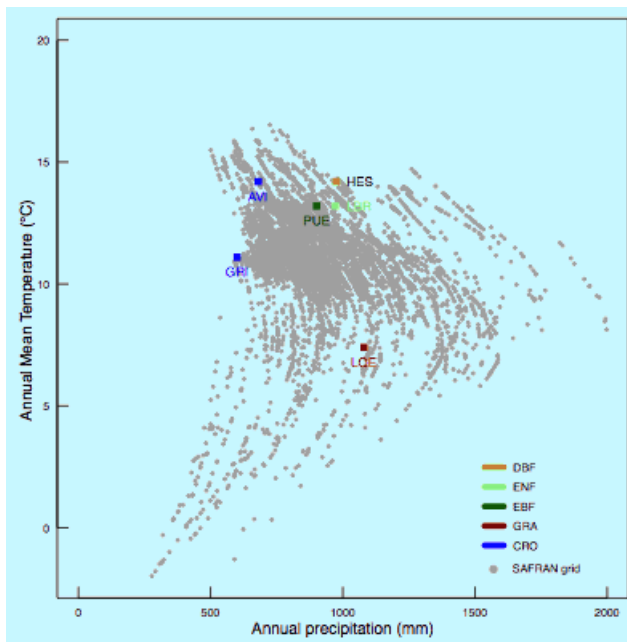


Fig. 1. The ecosystem types (DBF=deciduous broad-leaved forest, EBF=evergreen broad-leaved forest, ENF=evergreen needle-leaved forest, CRO=cropland, GRA=grassland) of six selected flux sites over France in the climatic space: mean temperature versus annual precipitation, which are calculated as the mean of SAFRAN results over 1994 to 2007. The observed mean temperature and annual precipitation at each site is indicated. The six sites are: Hesse (HES), Puéchabon (PUE), La Bray (LBR), Laqueuille (LQE), Avignon (AVI) and Grignon (GRI).

1 July to 31 August for HES, LBR and LQE, and 1 May to 30 June for PUE, AVI and GRI.

All sites are equipped to measure NEE, SH and LH by the covariance technique at every 30 min time step (Baldocchi et al., 2001). Meteorological data were continuously measured and averaged every half-hour. Because flux measurements are affected by (1) both random and systematic (bias) errors, which arise from limitations of the measurement technique, (2) the stochastic nature of turbulence, and (3) site-specific differences in data processing protocols (Moncrieff et al., 1996; Papale et al., 2006; Richardson et al., 2006, 2008), quality checks of the data were done according to CarboEurope-IP guidelines (Aubinet et al., 2000). Gap-filling was performed according to the marginal distribution sampling (MDS) method (Reichstein et al., 2005), for which uncertainties were quantified in gap-filling by Moffat et al. (2007). The MDS technique showed a consistently good gap-filling performance and low annual sum bias. According to the study of Moffat et al. (2007), the gap-filling techniques are already at or very close to the noise limit of the measurements and the effects of gapfilling on the annual sums of NEE are modest ($\sim 50 \text{ g C m}^{-2} \text{ yr}^{-1}$). The flux tower data were downloaded from the CarboEurope-IP database

(<http://gaia.agraria.unitus.it>) for CO_2 fluxes, SH, LH and all meteorological parameters except longwave radiation, which was complementarily provided by each site manager. Level-4 data sets are used for the study, in which flux separation techniques for splitting the observed net carbon fluxes into assimilation and respiration have been employed (Reichstein et al., 2005).

2.2 ORCHIDEE model

2.2.1 Model description

The ORCHIDEE terrestrial biosphere model describes the carbon, energy and water fluxes, (Krinner et al., 2005) and ecosystem carbon and water dynamics. It contains three sub-modules: a land surface energy and water balance model, SECHIBA (de Rosnay and Polcher, 1998; Ducoudré et al., 1993); a land carbon cycle model, STOMATE; and a model of long-term vegetation dynamics that includes competition and disturbances, adapted from Sitch et al. (2003). In this study, prescribed vegetation is used at each site (5 plant functional types (PFT) – see Table 1 – being relevant in this study).

The half-hourly energy and water balance of vegetated and non-vegetated surfaces, as well as canopy-level photosynthesis, is modeled by using coupled leaf-level photosynthesis and stomatal conductance equations (Ball et al., 1987; Farquhar et al., 1980). Stomatal conductance is reduced by the soil water stress (McMurtrie et al., 1990) function of soil moisture and root profiles. Two soil water reservoirs are considered, a surface reservoir which refills in response to rain events and which is brought to zero during dry periods, and a deeper bucket reservoir of 2 m depth updated from evaporation, root uptake, percolation and runoff on a daily time scale.

Autotrophic respiration is modeled at a half-hourly time step, and plant growth, mortality, soil carbon decomposition and phenology at a daily time step. Leaf onset is calculated as a function of growing degree-days and chilling days requirements, or as a function of soil moisture changes specific to each PFT (Botta et al., 2000). Assimilated carbon can be allocated to stems, leaves, fruits, carbohydrate reserves, fine and coarse roots. Allocation is controlled by phenology, and by light availability, temperature and soil water (Friedlingstein et al., 1999). Autotrophic respiration is the sum of temperature-dependent maintenance respiration processes and GPP-dependent growth respiration processes. Litter and soil organic matter decomposition are calculated on a daily time step using first-order kinetics for decay of 5 C pools based upon the CENTURY model equations (Parton et al., 1988).

2.2.2 Model set-up for site simulations

ORCHIDEE is run at each site, driven by meteorological data (see Sect. 2.3.1 for the required meteorological variables).

The fractional coverage of PFT at each site is prescribed according to site species data (Table 1). Each simulation includes an equilibrium spin-up run followed by a transient run. In the spin-up run, observed meteorology for the period of observation is used cyclically to drive the model for 1500 yr until equilibration of carbon (and water) pools is attained, with $<0.05\%$ yearly increment. Biomass pools reach equilibrium within typically 200 yr. The atmospheric CO_2 concentration is set to its present-day value of 370 ppm during the entire spin-up. After spin up, the model is run during the period of observation with yearly atmospheric CO_2 concentration (GLOBALVIEW- CO_2 , 2011). Given this approach, the long-term modeled carbon balance (NEE) is by construction equal to zero on a multi-year basis, unlike as observed at each flux tower. Therefore, ORCHIDEE NEE is always biased because all the sites have a net CO_2 uptake, $300 \text{ gC m}^{-2} \text{ yr}^{-1}$ on average for the forest sites and $200 \text{ gC m}^{-2} \text{ yr}^{-1}$ on average for the grass and crop sites (Table 1). At the cropland sites, where harvested biomass is exported away from ecosystem, and thus not respired, the flux-tower observed NEE is a permanent atmospheric CO_2 sink compared to the model estimate of zero. This bias of long-term NEE at each site can be corrected by scaling the disequilibrium of soil C pools by an empirical factor (Sect. 2.5).

2.3 Local flux towers meteorology and gridded model products

2.3.1 Flux tower meteorology

Meteorological half-hourly forcing data required to drive ORCHIDEE are surface air temperature (T_{air}), surface air specific humidity (Q_{air}), precipitation (Rainfall), downwelling shortwave (SW_{down}) and longwave (LW_{down}) radiation, surface pressure, wind speed and annual atmospheric CO_2 concentration. The uncertainty of the last three drivers has little impact on the model output, and are thus discarded in the following.

Observed meteorology (LOCAL) is recorded continuously on top of each tower, averaged every half-hour and quality checked (Aubinet et al., 2000; Reichstein et al., 2005). However, in the Level-4 data set, there are still some missing values due to either data quality control or instrumental failure, which account for about 1 % of total number of time steps for T_{air} , Q_{air} , Rainfall and SW_{down} . Because running ORCHIDEE requires continuous forcing, these gaps are filled with the SAFRAN model products, using a linear regression between tower and SAFRAN daily data in order to keep consistency on a daily time resolution. Daily regression results are further disaggregated into half-hourly values. Half-hourly T_{air} for gap-filling is generated from SAFRAN minimum and maximum T_{air} using a sine wave, assuming that maximum temperature occurs at 14:00 UTC + 1 and minimum temperature at 05:00 local time (Campbell and Norman, 1998). Using the same formula, instead of air temperature, daily Q_{air} is

disaggregated into half-hourly values. Daily Rainfall for gap-filling is simply given an even distribution throughout the day. Missing hourly SW_{down} data are gap-filled from daily values and solar zenithal angle at each site. Note that LW_{down} is not measured at LQE and HES, and covers a subset of the entire flux record at PUE (2004–2007). In the case of missing measurements, LW_{down} is gap-filled by the empirical relationship:

$$\text{LW}_{\text{down}} = (\text{cloudy} + (1 - \text{cloudy}) \cdot (1670 \cdot Q_{\text{air}})^{0.08}) \cdot \text{Stefans} \cdot T_{\text{air}}^4 \quad (1)$$

Where cloudy (cloud cover) is the ratio of SW_{down} to maximum SW_{down} under clear sky conditions, based on solar zenithal angle; Q_{air} is the specific humidity (kg kg^{-2}); Stefans is the Stefan-Boltzmann constant ($5.67 \times 10^{-8} \text{ Wm}^{-2} \text{ K}^{-4}$); and T_{air} is the air temperature in K. Tests during the period with observations show that this Eq. (1) overestimates mean annual LW_{down} by 5–15 Wm^{-2} . Comparison of ORCHIDEE fluxes driven by LOCAL or by modeled meteorology is focused on the growing season of period 2004–2007, during which gaps in LOCAL are negligible (less than 0.01 %). In other words, the comparison between LOCAL and SAFRAN in this study can be thought of as quasi-independent.

2.3.2 Gridded meteorology

Four gridded meteorology products are studied: SAFRAN, EC-OPERA, ERA-I, and REMO (Table 2). SAFRAN is an optimal data interpolation product based upon French synoptic weather station measurements and model results (Durand et al., 1993, 1999). The three other products are generated by Numerical Weather Prediction models that assimilate synoptic in situ and satellite data. The EC-OPERA data set is from the European Centre for Medium-range Weather Forecast (ECMWF). Frequent updates in the ECMWF operational assimilation package and atmospheric model cause discontinuities in the analyzed products. The ERA-I reanalysis (Berrisford et al., 2009) is a consistent production of meteorology and climate generated with the same model. The REMO product is a regional meteorological data set obtained by driving the REgionalMOdel (Jacob and Podzun, 1997) over Europe, with initial and boundary conditions prescribed from ECMWF global fields (Kalnay et al., 1996). Each gridded data-product has a different horizontal and vertical resolution. For comparison with tower meteorology, we selected the corresponding point in each model grid. A detailed description of each gridded meteorology product is given in Appendix A.

2.4 Model–data comparison rationale and statistics

Ecosystem-atmosphere fluxes are shaped by a variety of fluctuations on different scales of characteristic variability, ranging from hourly, diurnal, synoptic, seasonal-annual, to

Table 2. Summary of atmospheric analyses.

Atmospheric analyses	Period	Temporal resolution	Spatial resolution	References
SAFRAN	1994–2007	hourly	8 km	Durand et al. (1993, 1999), Quintana-Segui, et al. (2008)
EC-OPERA	2001–2006	6-h	40 km	www.ecmwf.int
ERA-I	1989–2008	6-h	79 km	Berrisford et al. (2009)
REMO	1988–2007	daily	50 km	Jacob and Podzun (1997) Kalnay et al. (1996)

decadal periodicities (Kattal et al., 2001; Stoy et al., 2005, 2009; Mahecha et al., 2007). Previous studies show that model–data agreement is a matter of frequency (Baldocchi et al., 2001; Stoy et al., 2005). Despite that spectral analysis has the advantage of identifying the model–data disagreement in a characteristic scale, it gives no information on the mean values and thus is unsuitable for bias estimation, one of the interests of this study. Therefore, in this study we adopt the conventional time series decomposition practice by varying aggregation levels to construct the half-hourly time series into four characteristic time series: hourly, daily, monthly and annual, referred to as X_i ($i = 1, 4$), respectively (see Eq. 2 in below). However, the resulting hourly (X_1) and daily (X_2) time series obviously bear strong seasonality. We thus need to go one step further to remove the seasonal cycle. We build the new hourly (Y_1) and daily (Y_2) time series by subtracting the daily mean (X_2) and monthly mean (X_3) from the aggregated hourly (X_1) and daily (X_2) time series, respectively. Consistently, a new monthly time series (Y_3) is constructed from the monthly time series (X_3) by removing its annual mean (X_4). We keep annual time series X_4 unchanged; that is Y_4 is equal to X_4 because the long term tendency is not observed within the period of less than 10-yr. Thus, the newly constructed time series Y_i ($i = 1, 3$) are essentially series of “anomaly”. The above described process is summarized as the following equation:

$$Y_i = X_i - X_{i+1} \quad i = 1, 4; \text{ if } i = 4 \text{ then } X_{i+1} = 0 \quad (2)$$

where X_i is the conventional aggregated hourly ($i = 1$), daily ($i = 2$), monthly ($i = 3$) and annual ($i = 4$) time series. Y_i ($i = 1, 4$) is the respective time series of “anomaly”. We note that in constructing daily time series Y_2 , the monthly series X_3 is technically replaced by a series of 31-point running means of daily aggregation X_2 . The resulting series Y_2 is not essentially different from the one made with true monthly mean X_3 , but it avoids the series jumping abruptly at the node of two successive months. In the following, the statistics are based on time series Y_i ($i = 1, 4$) if it is not specified.

Statistical criteria applied to the differences between measured and modeled meteorology (Sect. 3) and to modeled and observed ecosystem fluxes (Sect. 4) are (1) the Mean Abso-

lute bias Error (MAE), (2) the squared correlation coefficient (R^2), and (3) the Root Mean Squared Error (RMSE).

2.5 Optimization of TER

The annual mean simulated TER is expected to be overestimated by ORCHIDEE, given the equilibrium assumption for model set up (Sect. 2.2) where soil carbon pools are at their maximum value for the given climate. These overestimates of soil C pools and mean TER also have an impact on modeled variability as studied by Carvahais et al. (2010). In order to improve the simulation of TER, we scaled the simulated annual respiration to match the observed mean, according to the following steps:

Step 1: optimize the simulated maintenance autotrophic respiration ($RA_{M \text{ model}}$) by multiplication by a “biomass disequilibrium factor” defined as the ratio of measured to simulated average total biomass ($\langle B_{\text{obs}} \rangle / \langle B_{\text{model}} \rangle$), and assume that GPP-dependent growth respiration ($RA_{G \text{ model}}$) is perfectly simulated;

Step 2: estimate the mean average heterotrophic respiration (RH_{obs}) from the difference between observed TER and the sum of $RA_{M \text{ model new}} + RA_{G \text{ model}}$, and calculate an optimized simulated heterotrophic respiration at each time step by multiplying the modeled RH by an average “soil C disequilibrium ratio” defined as the ratio of measured to simulated average RH.

This gives:

$$RA_{M \text{ model opt}} = RA_{M \text{ model}} \cdot \langle B_{\text{obs}} \rangle / \langle B_{\text{model}} \rangle \quad (3)$$

$$RH_{\text{model opt}} = RH_{\text{model}} \cdot \langle TER_{\text{obs}} - RA_{M \text{ model opt}} - RA_{G \text{ model}} \rangle / \langle RH_{\text{model}} \rangle \quad (4)$$

$$TER_{\text{model opt}} = RH_{\text{model opt}} + RA_{G \text{ model}} + RA_{M \text{ model opt}} \quad (5)$$

We apply this optimization procedure to all sites except PUE. The observed total biomass provided by each site manager is

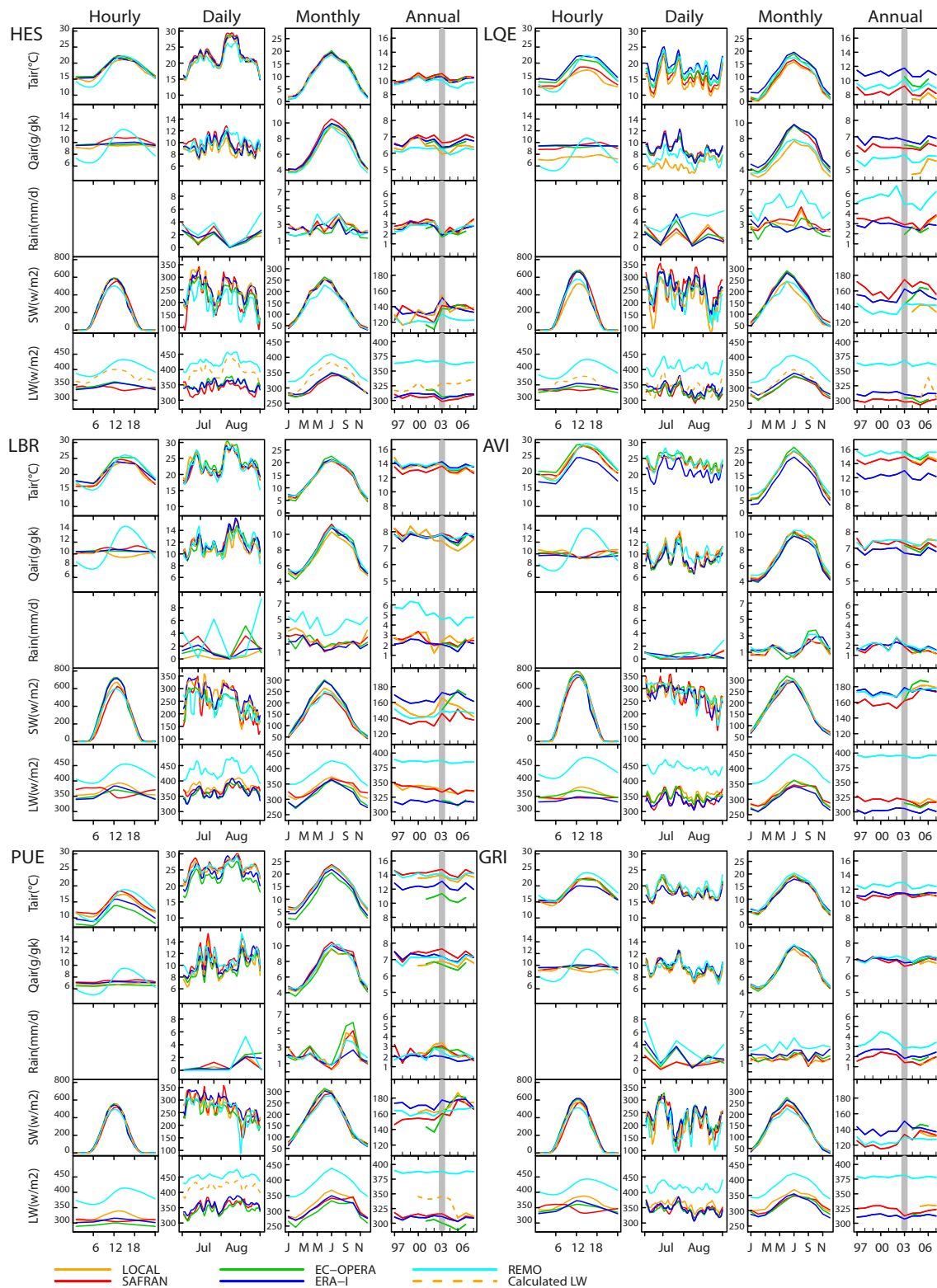


Fig. 2. Meteorological drivers in in situ and in gridded data sets at six sites. First column: hourly mean diurnal cycle over peak growing season (PGS). Second column: daily mean with a running mean of 3 days for July–August of 2003 at HES, LBR and PUE and of 2005 for LQE, AVI and GRI. Rainfall is calculated as 5-day aggregated values: third column: monthly mean seasonal cycle; fourth column: annual mean. The hourly mean diurnal cycle and monthly mean seasonal cycle correspond to 2004–2007 except for EC-OPERA (2003 to 2006). In the case of site-year without measured LW_{down} , calculated LW_{down} is plotted in dash lines. See text Sect. 2.1 for the definition of PGS.

7000, 7850, 300, 400 and 225 $\text{gC m}^{-2} \text{yr}^{-1}$ at HES, LBR, LQE, AVI and GRI, respectively. The biomass disequilibrium factors are respectively 0.35, 0.55, 0.45, 0.58 and 0.27, and the soil disequilibrium factors are respectively 0.60, 0.77, 0.85, 0.67 and 0.30 at these five sites. In other words, heterotrophic respirations are overestimated by ORCHIDEE from 18 % at LQE to 220 % at GRI.

This optimization procedure for TER is not applied to the Mediterranean forest site PUE because the overestimated TER at this site is caused by both discrepancies in carbon allocation between root and aboveground reservoirs and the equilibrium assumption: calculation from the above procedure would give negative $\text{RH}_{\text{model opt}}$, which is not realistic. We thus simply optimized the simulated TER with the average observed TER. The ratio of averaged observed to simulated TER is about 0.67, indicating that TER is overestimated by ORCHIDEE by about 49 % at PUE.

Thus NEE is optimized for all six sites according to

$$\text{NEE}_{\text{model opt}} = \text{TER}_{\text{model opt}} - \text{GPP}_{\text{model}}. \quad (6)$$

In the study, the optimized TER and resulting NEE are presented and discussed without specification.

3 Comparing gridded meteorology forcing against flux tower data

Figure 2 shows a comparison between observed (LOCAL) and modeled meteorology from hourly to interannual time scales based on aggregated time series X_i ($i = 1, 4$). Due to the short length of records, some sites were excluded: GRI, where the first year of observations is 2005 and whereas EC-OPERA forcing data is only available till 2006. Figure 3 shows the R^2 of LOCAL vs. gridded meteorology on different time scales (Eq. 2). Figure 4 shows the MAE; hourly and daily statistics are calculated only during the PGS, and monthly statistics during the GS.

3.1 Hourly time scale

Strictly speaking, the analyses-observation comparison on an hourly time scale is only applicable to SAFRAN set because the time resolutions in the other three gridded sets are lower than hourly (Table 2). However, ORCHIDEE includes an interpolator to convert 6-h or less into half-hours; that is, the coarse time step of SW_{down} is converted into half-hourly time steps as a function of solar zenithal and site location. The other meteorological variables are linearly interpolated between two measured times. Daily time series (in the case of REMO) are disaggregated into a half-hourly scale following the procedure described in Sect. 2.3.1 for gap-filling, when downscaling daily to half-hourly is required. Thus, analyses-observation comparison on an hourly scale provides an opportunity for insight on how good the conversion procedure from coarse time scale to half-hourly is, and whether the spurious behavior has any impact on the simulated carbon and

water fluxes, because conversion of daily to hourly forcing is common practice for TBMs.

The average diurnal cycles of T_{air} and SW_{down} are well simulated by all gridded products, with R^2 from 0.51 to 0.97 ($p < 0.01$) (Fig. 3). The daytime values of T_{air} between 06:00 and 19:00 UMT appear, however, to be overestimated at the forest sites, but within 2 °C of local observations at the crop sites. This may reflect local evaporative cooling over forests (Zaitchik et al., 2006; Teuling et al., 2010). The pronounced diurnal cycle of Q_{air} presented in REMO is found neither in LOCAL, nor in any other gridded data set. This spurious diurnal amplitude is most likely caused by our conversion of daily to hourly values, rather than a structural bias of the REMO physics (Campbell and Norman, 1998). The observed LW_{down} diurnal amplitude ($\approx 40 \text{ W m}^{-2}$) at LBR, PUE, AVI and GRI is underestimated by SAFRAN, EC-OPERA and ERA-I, which give values of 22, 26 and 34 W m^{-2} , respectively, while it is overestimated by Eq. (1) when applied to daily REMO output (60 W m^{-2}). At LBR, HES and GRI, the diurnal cycle of LW_{down} in SAFRAN is opposite to that of other models.

3.2 Daily time scale

Comparison between tower data and gridded model products is focused on the summer 2003 heat-wave (July–August) period at HES, LBR and PUE, and on summer 2005 at LQE, AVI and GRI (Fig. 2). The main result is that the synoptic-scale variability of daily T_{air} , Q_{air} and SW_{down} is well captured by all gridded data sets when compared to LOCAL. For daily variability of T_{air} during July–August 2003 or 2005, the mean R^2 of LOCAL and models is 0.87 (from 0.79 in REMO to 0.94 in SAFRAN, $p < 0.01$). The synoptic variability of T_{air} is best captured at HES, where the mean R^2 of the four models is 0.96. The July–August mean T_{air} at LQE is overestimated by all models, from 1.1 °C in SAFRAN to 4.2 °C in ERA-I. This summer bias must be compared to the annual mean T_{air} bias of 0.8 °C in SAFRAN and 3.3 °C in ERA-I, due to unresolved topography.

Concerning Rainfall, SAFRAN is in good agreement with LOCAL for daily values during July–August at all sites, except for LBR. At LBR, SAFRAN produces a mean Rainfall of 101 mm against 23 mm only in LOCAL, but the rain gauge data quality was poor during 2002 to 2006 due to instrumental failure (Loustau, personal communication, 2010). The mean summer 2003 Rainfall is 71 mm in SAFRAN and 75 mm in LOCAL across the six sites, excluding LBR. REMO overestimates Rainfall by 80 mm in summer 2003, which would cause problems for simulating the response of plants to drought during the dry 2003 summer. The daily variability of Q_{air} is characterized by overall mean R^2 values between models and LOCAL of 0.72, REMO having the lowest correlation (0.44) and SAFRAN the highest (0.86). The AVI site has the highest R^2 between observed and modeled Q_{air} during July–August 2005 ($R^2 = 0.90$, $p < 0.01$). The LQE

R2: Meteor. Analysis vs. Local Obs.

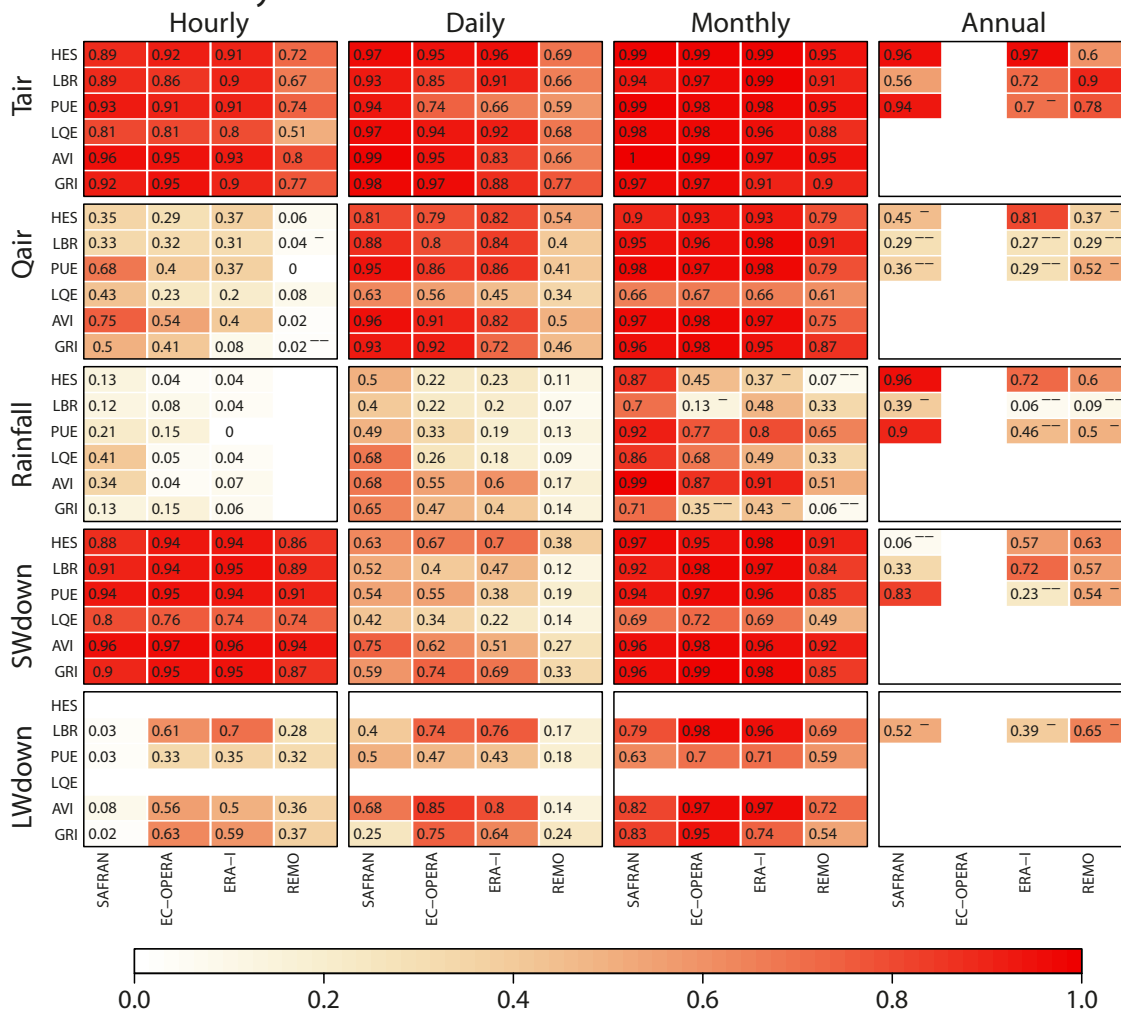


Fig. 3. Squared correlation (R^2) between meteorological gridded data and in situ data over 2004 to 2007, except for EC-OPERA which covers 2004–2006. Panels of R^2 from left to right are for hourly, daily, monthly and annual time scales, respectively. Time series used to calculate R^2 correspond to growing season (GS). See text Sect. 2.1 for the definition of GS. The default statistical confidence level of R^2 is $p < 0.01$. Otherwise, the signal “-” at the upright indicates a confidence level of $p < 0.05$ and “--” for $p > 0.05$.

site has the smallest R^2 (0.20, $p < 0.01$), with low Q_{air} observed during July, and early August being captured by none of the models. The daily variability of SW_{down} has a mean R^2 across the six sites of 0.49, with a range going from 0.27 in REMO to 0.68 in EC-OPERA. The HES forest has the highest R^2 for SW_{down} (0.63) and the PUE Mediterranean forest the lowest (0.32). But the value of R^2 between LOCAL and analysis data is lower for SW_{down} than for T_{air} or Q_{air} , and thus errors in SW_{down} will be a concern in driving TBM models like ORCHIDEE by gridded products (see Sect. 5.4).

The LW_{down} daily variability is well represented by SAFRAN, EC-OPERA and ERA-I, with R^2 values going from 0.55 ($p < 0.01$) in SAFRAN to 0.78 ($p < 0.01$) in EC-OPERA across the sites at which LW_{down} measure-

ments were collected during summer 2003; REMO gives poor performances ($R^2 = 0.25$, $p < 0.01$). Observed LW_{down} (excluding gap-filling values) in summer 2003 and 2005 is about 365 W m^{-2} , which is about 5% ($p < 0.01$) underestimated by SAFRAN, EC-OPERA and ERA-I, but 18% overestimated by REMO ($p < 0.01$).

3.3 Monthly time scale

The mean seasonal cycle of T_{air} , Q_{air} , SW_{down} and LW_{down} is well captured by all gridded products (Fig. 2), with mean R^2 above 0.95 ($df = 10$, $p \ll 0.01$). The seasonality of Rainfall is well represented by gridded data sets at PUE and AVI where the Mediterranean summer is very dry, and most Rainfall is delivered in autumn ($R^2 = 0.76 \pm 0.12$, $p \ll 0.01$). The agreement between observed and modeled seasonal

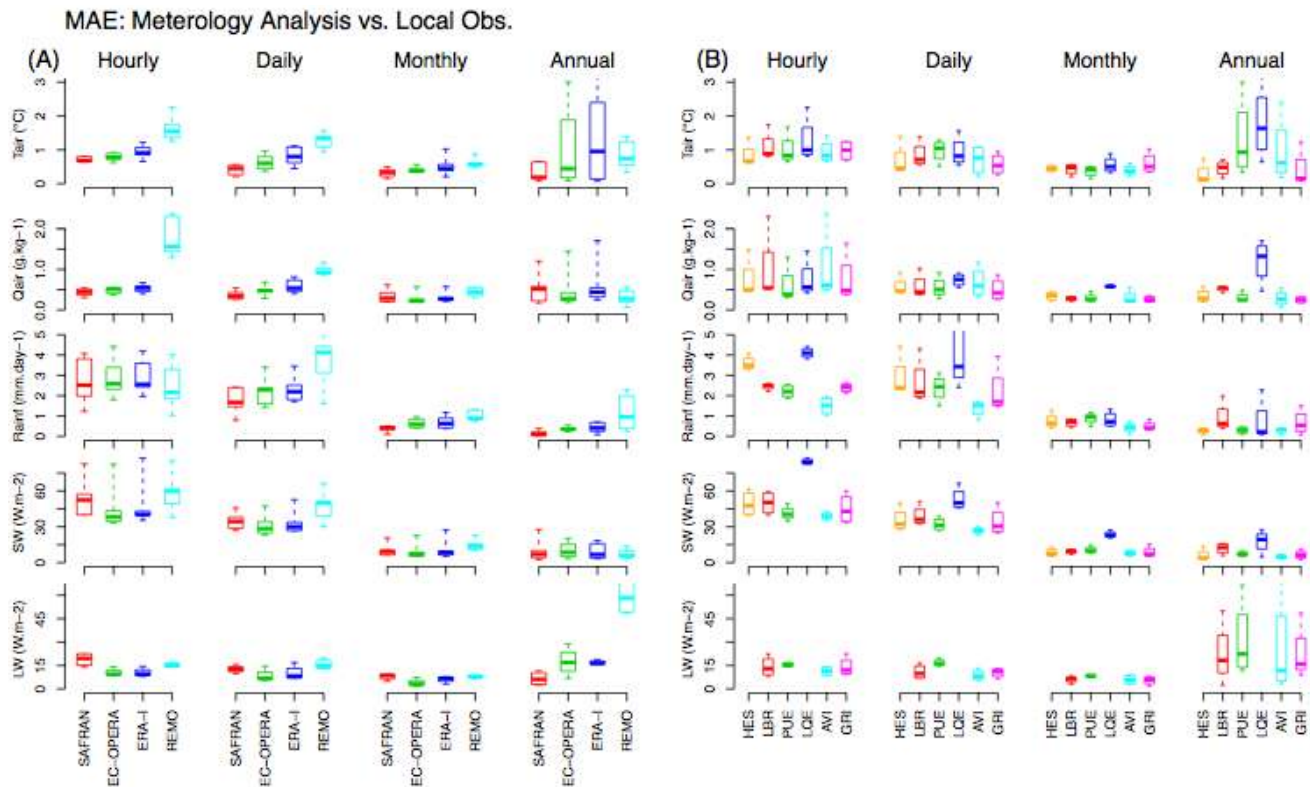


Fig. 4. Mean absolute error (MAE) between meteorological gridded data and in situ data (LOCAL) over 2004–2007, except for EC-OPERA which covers only 2004–2006. Time series used to calculate MAE correspond to growing season except for at the annual time scale, which covers continuously from 1997 to 2007, whenever data is available.

Rainfall is better in SAFRAN ($R^2=0.80$, $p < 0.01$) than in other analysis data sets ($R^2 \approx 0.50$, $p < 0.05$). The seasonal cycle of LW_{down} in all gridded data sets is in agreement with LOCAL, but this variable is overestimated by REMO (by 60 W m^{-2}) and slightly underestimated by the other gridded data sets (5 W m^{-2} in SAFRAN, and 15 W m^{-2} in EC-OPERA and ERA-I). The seasonal phase of LW_{down} in REMO correlates well with observations ($R^2=0.86$, $p < 0.01$); the positive bias of LW_{down} being correlated with the wet bias of this model.

3.4 Annual time scale

At the 6 sites, SAFRAN, EC-OPERA and ERA-I do not show systematic bias compared to LOCAL, while REMO shows an obvious wet bias: Rainfall is overestimated about 2 mm day^{-1} and LW_{down} is about 18 % higher (60 W m^{-2}) than LOCAL. At the mountain grassland LQE there is a positive bias of atmospheric T_{air} in gridded products going from 0.8°C in SAFRAN (high resolution) to 4.3°C in ERA-I (coarse resolution). This bias reflects the coarse topography of weather models (Sczcypta et al., 2011). A further examine shows that the altitude at the corresponding grid where meteorology is extracted is always lower than the real altitude

(1040 m), leading to higher T_{air} , which further impacts the calculation of Q_{air} and SW_{down} .

Considering HES, LBR and PUE have records longer than 6 yr, and excluding years with long gaps (a week-long gap during growing season or 2 week-long gaps during non-growing season) in measurements, the average R^2 values between annual modeled and observed meteorology are 0.80 ± 0.15 ($p < 0.01$), 0.40 ± 0.17 ($p < 0.05$), 0.51 ± 0.32 ($p > 0.05$) and 0.50 ± 0.24 ($p > 0.05$) for T_{air} , Q_{air} , Rainfall and SW_{down} , respectively (Fig. 3). A conclusion can not be made for LW_{down} as there is only one site (LBR) with long observed records, but R^2 tends to be low (0.40–0.65). The main result is that the weather models are only able to reproduce correctly, i.e. R^2 and $p < 0.01$, the interannual variability of T_{air} , but the variability of other drivers is poorly captured. In particular, the interannual variability of SW_{down} is not well reproduced. The interannual variability of Rainfall is faithfully reproduced at HES ($R^2=0.96$, $p < 0.01$) and PUE ($R^2=0.90$, $p < 0.01$) by SAFRAN. This gives higher confidence in SAFRAN meteorology to drive carbon flux on a year-to-year basis. At the LBR site, this conclusion can not be made because of rain gauge disfunction in some years (Loustau, personal communication, 2010). The mismatch of

Rainfall between the measurement and analysis data at LBR was also identified by Chen et al. (2007).

3.5 Summary of gridded data sets performance

3.5.1 Correlations between modeled and observed variability

In general, gridded data sets compare better with local observations on the monthly scale than other time scales (Fig. 3). The overall mean monthly R^2 across 5 variables, 6 sites and 4 modeled data sets is 0.82 ± 0.21 ($p \ll 0.01$), significantly better than 0.53 ± 0.35 ($p \ll 0.01$) for the hourly scale, 0.58 ± 0.27 ($p < 0.01$) for the daily scale and 0.55 ± 0.26 ($p > 0.05$) for the annual scale. The diurnal cycles of T_{air} and SW_{down} are realistic in all gridded data sets with overall mean R^2 of 0.85 ± 0.11 ($p \ll 0.01$) and 0.90 ± 0.07 ($p \ll 0.01$), respectively, which indicate that the conversions of daily values to hourly are reasonable for these two variables. By contrast, the diurnal cycle of Rainfall is not well captured by any of the gridded data sets (0.12 ± 0.11 , $p < 0.01$), not even by SAFRAN because sub-daily Rainfall was not included in the analysis of SAFRAN (Quintana-Segui et al., 2007). The diurnal cycle of LW_{down} in SAFRAN is not reproduced or even opposite of the local observation and other gridded data sets. No skill in presenting diurnal cycle of Q_{air} in REMO is attributed to the conversion of daily values to hourly. The daily variabilities of T_{air} (0.85 ± 0.13 , $p \ll 0.01$) and Q_{air} (0.72 ± 0.20 , $p \ll 0.01$) are better captured than SW_{down} (0.46 ± 0.19 , $p \ll 0.01$) and LW_{down} (0.50 ± 0.25 , $p \ll 0.01$). It is still difficult for weather models to capture the daily variability of Rainfall (0.33 ± 0.20 , $p < 0.01$). Although measured daily Rainfall from climatological networks has been included to produce the SAFRAN analysis, the skill is modest ($R^2=0.57 \pm 0.12$, $p < 0.01$), indicating the problem of spatial heterogeneity in Rainfall can not be solved even in a fine resolution data set such as SAFRAN (Quintana-segui et al., 2008). On the monthly scale, gridded data sets display improved skill to capture Rainfall variability ($R^2=0.57 \pm 0.28$, $p < 0.01$), but are still lower than other variables ($R^2=0.89 \pm 0.12$, $p < 0.01$). Concerning interannual variability, only T_{air} is well captured by gridded data sets ($R^2=0.82 \pm 0.15$, $p < 0.01$).

Among the 5 meteorological variables, T_{air} is best captured by all gridded data sets at time scales from hourly to annual ($R^2=0.88 \pm 0.12$, $p < 0.01$), while Rainfall is the worst ($R^2=0.38 \pm 0.29$, $p < 0.01$); standing between them are SW_{down} ($R^2=0.72 \pm 0.26$, $p < 0.01$), Q_{air} ($R^2=0.65 \pm 0.26$, $p < 0.01$) and LW_{down} ($R^2=0.55 \pm 0.27$, $p < 0.01$). The diurnal cycle of SW_{down} is well captured by all gridded data sets, but the challenge is at daily and annual scales, contrary to Q_{air} in which daily variability is reasonably captured but difficult in diurnal cycle and interannual variability.

Among the 6 sites, the mountain grassland site LQE is the most difficult to simulate by the gridded products, in that monthly R^2 of Q_{air} and SW_{down} at LQE is significantly lower than at any other site ($df \approx 5$, $df \approx 4$, $p < 0.01$).

3.5.2 Bias of gridded data sets

Figure 4a shows the MAE between gridded data sets and tower observations. The mean biases of T_{air} and Q_{air} are small, with median MAE values ranging from 0.5°C to 2°C for T_{air} and from 0.2 to 1.5 g kg^{-1} for Q_{air} . The larger REMO hourly MAE is due to our conversion of daily values to hourly, and thus is not a bias of the model itself. For hourly Rainfall, we note that none of the gridded data sets performs better for MAE than a “null model” with even distribution of Rainfall each hour during rainy days. The MAE of Rainfall is maximized on a daily time scale, going from a $\text{MAE}=1.8 \text{ mm day}^{-1}$ in SAFRAN to $\text{MAE}=4.0 \text{ mm day}^{-1}$ in REMO. By contrast, the MAE of Rainfall remains moderate on longer, monthly and inter-annual scales ($\text{MAE} < 1.0 \text{ mm day}^{-1}$ across sites and models). The MAE of SW_{down} is large on hourly and daily scales ($30\text{--}60 \text{ Wm}^{-2}$), but improves on monthly and annual scales (10 Wm^{-2}). The MAE of LW_{down} is minimized with EC-OPERA on hourly to monthly scales, and maximized with REMO on annual scales (50 Wm^{-2} ; other data sets $\approx 20 \text{ Wm}^{-2}$).

Concerning the four gridded data sets, SAFRAN shows the least MAE for T_{air} , but the difference from other analyzed data sets is not significant ($p > 0.50$). Moreover, the lower bias for T_{air} in SAFRAN is likely due to the gain in resolution (see Appendix B). To the contrary, the MAE of Rainfall at monthly and annual scales is significantly smaller ($p < 0.01$) in SAFRAN than any other gridded data-product, irrespective of the spatial resolution (see Appendix B). REMO has the largest MAE for Rainfall at scales from daily to annual, indicating a systematic bias in this data-product. A wet bias is reflected in Fig. 2, in particular at sites of LBR, LQE, AVI and GRI, and the overestimated Rainfall was also identified by a previous study (Chen et al., 2007). The largest MAE for LW_{down} in REMO reflects the systematic overestimation of this variable as shown in Fig. 2.

Figure 4b shows that MAE of monthly to annual T_{air} , Q_{air} and SW_{down} are abnormally high at LQE compared to the other sites, irrespective of the gridded data-product compared with observations, which reflects the overestimate of the three variables at PUE by all the gridded data sets (Fig. 2).

As a general rule, we found that biases associated with gridded data sets are smaller on monthly and annual scales than on hourly and daily time scales (Fig. 4b). But on the annual scale, the divergency of MAE among sites and gridded data sets is larger than at any other time scale, indicating the difference among gridded data set and the observation is not negligible and site dependent. For example the spread of annual T_{air} among gridded data sets can be over 2.0°C at PUE,

LQE, AVI and GRI (Fig. 2). The high divergency of meteorological drivers among gridded data on an annual time scale is also highlighted by Ricciuto et al. (2009). They estimate the deviation of annual T_{air} among different data sets as $\sim 1^\circ\text{C}$, which is sizable compared to our findings ($\sim 0.8^\circ\text{C}$).

3.5.3 Summary

We found from the comparison at six selected sites between gridded data sets and local observed meteorology that the performance of different gridded data sets varies largely with meteorological variables and time scale. SAFRAN does a good job in reproducing observed Rainfall both in terms of temporal variability (R^2) and absolute bias (MAE), which is thanks to the inclusion of measured daily Rainfall from numerous rain gauges (Quintana-sugui et al., 2007). The performance of REMO is considerably worse than the other gridded data sets for most variables ($p < 0.01$), in particular for Rainfall, SW_{down} and LW_{down} . The problem of REMO in producing Rainfall and radiation has been reported in a previous study (Chen et al., 2007). For T_{air} , Q_{air} and SW_{down} the performances of SAFRAN, EC-OPERA and ERA-I are not significantly different from each other. While not statistically significant, EC-OPERA and ERA-I are slightly better than SAFRAN in simulating SW_{down} . In a recent study comparing ERA-I and SAFRAN, Szczypta et al. (2011) found that (1) the consistency between these two products is good for T_{air} and Q_{air} ; (2) Rainfall in ERA-I does not match SAFRAN in mountainous areas and on the Mediterranean coast; and (3) ERA-I produces better SW_{down} than SAFRAN, which underestimates SW_{down} by about 5% in France overall. These results are essentially consistent with our findings.

Relatively, T_{air} from different gridded data sets and the observation have the best consistency over other variables. Although the interannual patterns are consistent, the deviation of annual mean T_{air} among gridded data sets and the observation is $\sim 0.8^\circ\text{C}$. Rainfall is difficult to simulate at hourly and daily time scales, but it is better captured at monthly and annual scales. Q_{air} and SW_{down} are the most problematic at the annual time scale ($R^2 \leq 0.5$, $p < 0.05$). Hourly Q_{air} and LW_{down} are not well reproduced by any gridded data set.

Among the six sites, meteorology at the mountain grassland LQE is the most difficult to simulate by gridded products due to the coarsely-resolved topography in atmospheric analyses. As a whole, meteorological drivers are best reproduced at the monthly scale and most problematic at the annual scale. At hourly and daily scales, the performance of gridded data sets are site and meteorological driver dependent.

We also found that our method to extrapolate daily REMO values into hourly data (as needed to drive ORCHIDEE) is a source of bias for Q_{air} and LW_{down} . This raises a caution flag when Q_{air} is calculated by a weather generator to drive carbon flux simulations (Richardson and Wright, 1984; Krinner et al., 2005).

4 Impact of driving meteorology on the simulation of ecosystem fluxes

In this section we study for each flux, the effect of forcing ORCHIDEE either with LOCAL, or with atmospheric analyses meteorology. Figure 5 shows a comparison between measured and modeled carbon and water fluxes from hourly to annual time scales based on aggregated raw time series X_i , ($i = 1, 4$, see Eq. 2). Figure 6 shows the R^2 of observed vs. gridded fluxes on different time scales. Figure 7 shows the MAE. Hourly and daily statistics are calculated only during the peak growing season, and monthly statistics during the growing season.

4.1 Comparison between modeled and measured fluxes

4.1.1 GPP

Figure 5a shows that the ORCHIDEE model has the following biases irrespective of the forcing used. Firstly, the GPP summer peak and thus the GPP diurnal cycle amplitude are overestimated at HES, PUE and LQE. By contrast, GPP is underestimated at the AVI southern crop site, even though the generic phenology parameterization of C3-crops in ORCHIDEE seems to reproduce rather well the early-season GPP peak of winter crops (wheat or peas) grown at AVI. The GPP seasonal amplitude is correctly captured at the GRI northern crop site, but the GPP increase in the spring is modeled too early compared to the observed flux. ORCHIDEE seems to overestimate the daily summertime variability of GPP at LQE and AVI even when the LOCAL, meteorology is used to drive it. An encouraging result is the ability of ORCHIDEE to capture the water stress induced decrease of GPP between early July and late August 2003 (Ciais et al., 2005) at the forest sites of HES, LBR and PUE (Fig. 5a). At sites with long-enough observation periods, a negative GPP anomaly in 2003 compared to other years is modeled, in agreement with the observation (Fig. 5a; gray bar). But the annual magnitude of the GPP anomaly of year 2003 is usually smaller than observed, suggesting an underestimated model sensitivity to drought. At the HES site, annual mean GPP is smaller in 2004 than 2003 because of lagged effects from diminished reserves (Granier et al., 2007), a process clearly lacking in ORCHIDEE, which predicts in 2004 a return to normal GPP values. We find that annual GPP is overestimated by ORCHIDEE about 8% in grassland (LQE), 25% in forest sites and over 40% in crop sites. These discrepancies might be explained by various reasons including underestimation of the drought limitation and its lag effect for forest sites, overestimated photosynthetic capacity at PUE, and overestimated length of the growth season at crop and grassland sites. In addition, the strategy of stomatal conductance reacting to soil water stress employed in ORCHIDEE may be an inefficient approach to simulating soil water stress at Mediterranean sites. When it was replaced

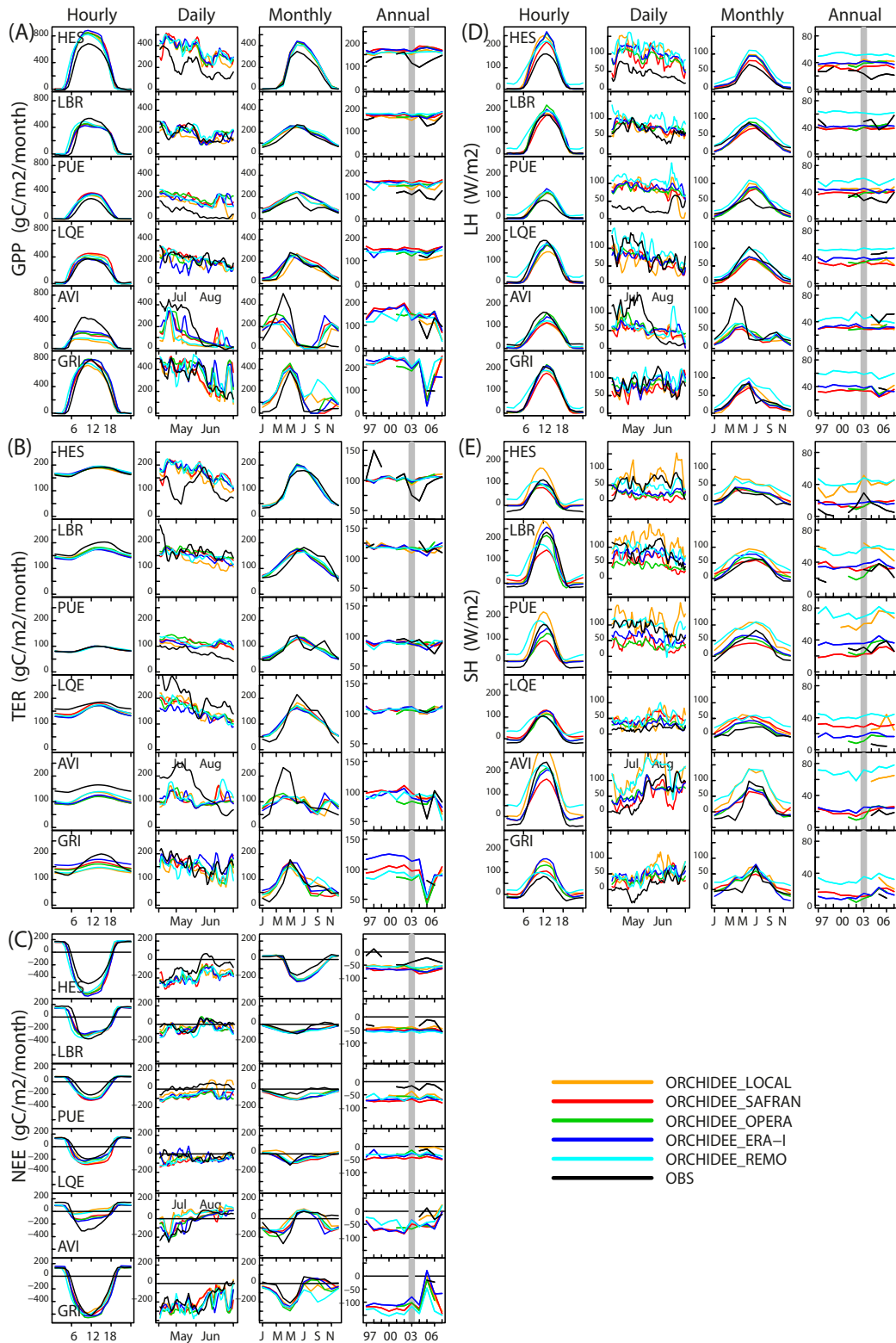


Fig. 5. Simulated and measured carbon fluxes ($\text{gC m}^{-2}\text{month}^{-1}$) and water fluxes (Wm^{-2}). (a): GPP; (b): TER; (c): NEE; (d): LH and (e): SH. Hourly mean diurnal cycle over peak growing season corresponds to 2004–2007; daily mean with a running mean of 3 days for PGS of 2003 at HES, LBR and PUE, of 2005 for LQE, AVI and GRI; monthly mean seasonal cycle over 2004–2007, except for AVI at 2005 and for GRI at 2006 when winter wheat was grown.

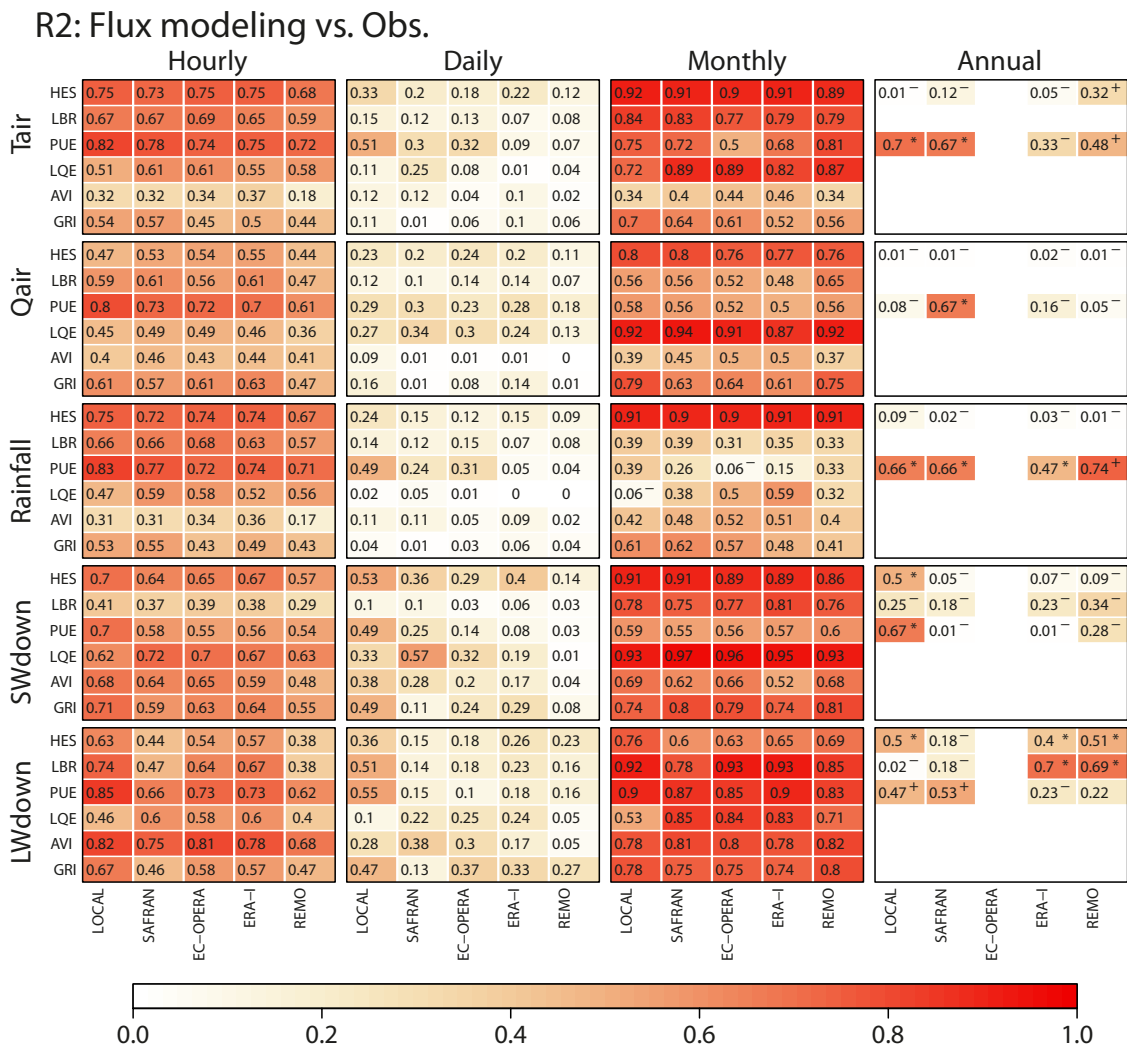


Fig. 6. Squared correlation (R^2) between simulated and measured flux data and in situ data over 2004 to 2007, except for EC-OPERA which covers 2004–2006. Panels of R^2 from left to right are for hourly, daily, monthly and annual time scales, respectively. Time series used to calculate R^2 correspond to growing season. For crop sites, we only take the years of winter wheat growing; that is, 2004 and 2006 at AVI, and 2006–2007 at GRI. The default statistical confidence level of R^2 is $p < 0.01$. Otherwise, the signal at the upper-right indicates a confidence level: “*” for $p < 0.05$, “+” for $p < 0.10$, and “-” for $p > 0.10$.

by empirically-derived non-stomatal soil moisture responses, Keenan et al. (2009) found that the performance (R^2) of ORCHIDEE in reproducing the measured seasonal cycle of GPP was substantially improved from 0.68 to 0.88.

4.1.2 TER

After optimization (see Sect. 2.5), the amplitude of the seasonal cycle of TER is correctly captured at the 3 forest sites, but underestimated at 3 non-forest sites. At HES, TER is underestimated in the early growing season and overestimated in summer and autumn, which is attributed to overestimated TER at night. TER at LBR is overestimated in winter and spring and underestimated in summer and autumn; thus TER is overestimated during its peak growing season. At Mediter-

anean site PUE, the modeled TER is too high during summer drought and too low in autumn compared to observations, although the year-round mean diurnal cycle is very well simulated. At grassland LQE, TER is overestimated in winter-spring and underestimated in summer-autumn. At the two crop sites, TER is significantly underestimated during the peak growing season April–May. Daily evolution of TER in a specific year is most difficult to simulate at crop and grassland sites. The reduced TER in early July of 2003 at HES is captured by ORCHIDEE, but it is too weak compared to the observed flux. Encouragingly, reduced TER in 2003 is more or less reproduced at HES and PUE where the observed flux data are available.

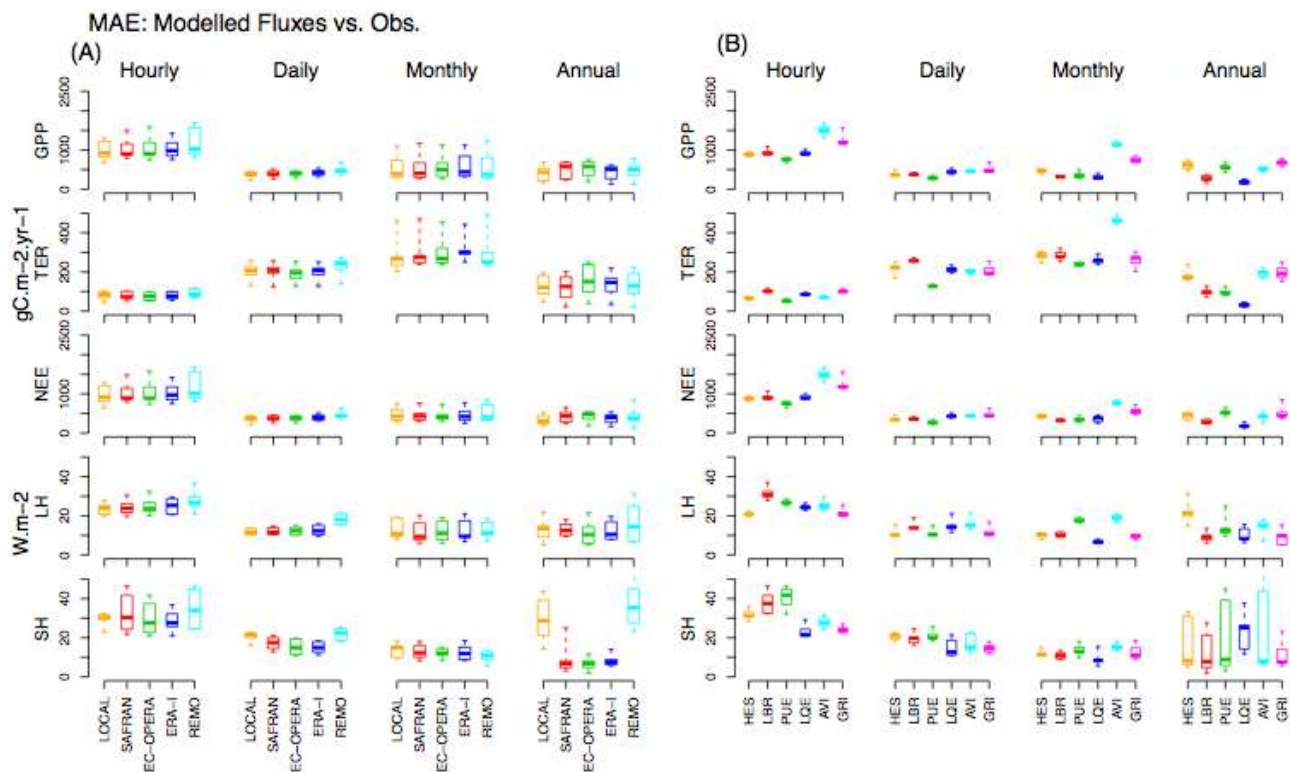


Fig. 7. Box plot of MAE between simulated and measured flux data over 2004 to 2007, except for EC-OPERA which covers only 2004–2006. (A) MAE across six sites; (B) MAE across 5 series. Columns from left to right are hourly, daily, monthly and annual means. The bottom and top of the box are the 25th and 75th percentile, respectively, and the band near the middle of the box is the median. The lower and upper ends of whiskers represent the minimum and maximum, respectively.

4.1.3 NEE

NEE is optimized according to Eq. (6). Thus the problem of “disequilibrium” is overcome; however, we should bear in mind that the optimized NEE tends to be systematically underestimated (Fig. 5c) due to the overestimated GPP by ORCHIDEE (Fig. 5a).

At HES, the model performs best, both on daily and monthly scales. The daily NEE variability is characterized in 2003 by an abrupt shift from sink to source by early August (see also Ciais et al., 2005). This large reduction in CO₂ uptake is well captured by ORCHIDEE, although NEE remains a sink due to TER optimization. At LBR, the modeled diurnal cycle amplitude of NEE is slightly small and the NEE uptake in the morning occurs 2 h earlier in the model, even though the diurnal NEE asymmetry, with a morning maximum between 10:00 and 12:00, is rather well reproduced by the model. The modeled NEE diurnal amplitude is overestimated at HES, PUE and LQE and slightly overestimated at GRI, but slightly underestimated at LBR and much underestimated at AVI. At the PUE Mediterranean oak forest, ORCHIDEE overestimates the seasonal uptake of CO₂ throughout the year. At the temperate forest HES, CO₂ uptake is overestimated during the growing season between

May and November. The seasonal phase of NEE is well represented at the forest sites HES and LBR, respectively ($R^2 = 0.94$ ($p < 0.001$) and 0.61 ($p < 0.01$)), but not at PUE ($R^2 = 0.26$, $p > 0.01$). Modeling interannual NEE variability is not satisfactory in ORCHIDEE. The best interannual R^2 is 0.78 ($n = 7$ yr) at PUE when the model is driven by LOCAL, meteorology. The worst R^2 is 0.16 ($n = 10$ yr) at HES. This could be partly explained by the influence of thinning and the difficulties in reproducing the impact of the large fluctuations of the relative extractable water soil content (Granier et al., 2008). The low skill of ORCHIDEE in reproducing interannual variability of carbon flux was also confirmed by Anav et al. (2010). There are problems in either leaf-level physiological responses to water or nutrient stress, and/or in scaling algorithms that take average leaf photosynthesis and convert it to a canopy value.

Krinner et al. (2005) examined the diurnal cycle and seasonal cycle of NEE simulated with an early version of ORCHIDEE for each PFT. In their study, ORCHIDEE simulated an amplitude that is smaller than the measured one at needle-leaf forests and greater at broad-leaf forest sites. The overall phasing of NEE was well simulated, but not so good for non-forest sites. For agricultural sites, both the diurnal and seasonal variations of NEE were badly reproduced. The

amplitude of diurnal cycle at grasslands was larger than the measured one in their study, but is smaller in ours. Given the differences in meteorological forcing and studied sites, the two studies are highly consistent.

4.1.4 Latent and sensible heat flux

Comparison between observed and modeled latent heat flux (LH) is given in Fig. 5d. On the diurnal time scale, there is an overestimation of LH by ORCHIDEE at HES and PUE. This positive bias of LH is logically reflected on the seasonal time scale during the growing season at these sites. Oppositely at LQE, the modeled diurnal cycle amplitude of LH is underestimated when the most realistic LOCAL, meteorology is used to drive ORCHIDEE. In particular, the increase of LH in the morning is delayed by roughly 1 h compared to the observations. At LBR, the model overestimates LH during the winter growing season from October to March, which parallels the overestimated GPP seen in Fig. 5a, indicating that winter acclimation of photosynthesis at this site (Medlyn et al., 2002) is missed by ORCHIDEE. At PUE, the model severely overestimates LH and also GPP during the dry summer (Fig. 5a). This suggests that the regulation of transpiration in response to water stress at this Mediterranean forest is too weak in ORCHIDEE. Patchy stomatal closure (Reichstein et al., 2003) has been suggested to limit transpiration losses at PUE, a process not incorporated in ORCHIDEE. Another model structural bias is the single-layered soil bucket model, which allows moisture to remain in the soil too long after each rain event (Keenan et al., 2009), and sustains simulated LH and GPP in the dry season. For a majority of sites, the LH bias on the seasonal time scale seems to be driven by the bias on the diurnal scale. The modeling assumption of equilibrium for carbon fluxes and pools, critical to explaining the NEE model–data misfit on the seasonal scale, seems here to have a negligible impact on the LH misfit. On the seasonal time scale, LH is predominantly controlled by soil moisture availability, atmospheric dryness and leaf area index, which are rather independent of the soil C pool values that set up the value of TER and NEE (Carvalais et al., 2010). On the annual time scale, the variability of LH is not captured by ORCHIDEE. The best interannual R^2 is 0.67 ($n=7$, $p < 0.05$) at PUE when the model is driven by LOCAL, meteorology. The worst R^2 is ~ 0.10 ($n=7$, $p > 0.10$) at LBR. On average at the 3 forest sites (more than 5 yr of observations), the interannual variability measured by coefficient of variation (CV) of LH is higher in the eddy-flux observations (CV = 0.18) than in the model (CV = 0.07).

Figure 5e shows that the mean annual sensible heat flux (SH) is overestimated by ORCHIDEE, independent of the time scale considered. The REMO forcing tends to produce the largest SH flux. SAFRAN, EC-OPERA and ERA-I forcings produce slightly larger SH fluxes than observed. SH flux is obviously overestimated by LOCAL, forcing, in particular at HES, PUE before 2005, and LQE. This discrepancy

between modeling, observation and inter-modeling is probably attributed to the difference in LW_{down} . REMO shows the largest positive bias of LW_{down} compared to that in LOCAL. LW_{down} in LOCAL at HES, PUE before 2005, and LQE is actually calculated according to Eq. (1), which tends to give an overestimated value. When overestimated LW_{down} combines with other meteorological parameters such as T_{air} and SW_{down} , this error is further enlarged in SH flux as shown in LQE. This shows that LW_{down} has a significant impact on the modeled SH at least for ORCHIDEE. The known positive bias of SH in ORCHIDEE, evidenced at nighttime in former versions of the model (Krinner et al., 2005), is still present in Fig. 5e. The day-to-day variability of SH is particularly well captured at HES during the dry summer 2003, but peaks of SH during the early July and early August 2003 heatwaves are overestimated. The seasonal variability of SH is poorly simulated, even after removing the positive SH bias. On average at the 3 forest sites (more than 5 yr of observations), the interannual variability of SH is higher in the eddy-flux observations (CV = 0.44) than in the model (CV = 0.16).

4.2 Correlations between modeled and observed fluxes, function of driving meteorology

Figure 6 shows the correlation between modeled and observed fluxes for different time scales, and for different meteorological drivers. Generally, the overall R^2 values are higher for diurnal and monthly time scales (0.57 ± 0.14 and 0.68 ± 0.20 , respectively) compared to daily and annual time scales (0.17 ± 0.13 and 0.28 ± 0.25 , respectively). The correlations at the daily time scale are always rather low, independent of the meteorology used to drive the model. This suggests that model errors largely explain the small values of R^2 at a synoptic scale. Figure 6 also shows that on the monthly scale, R^2 is higher for water fluxes (overall average R^2 across six sites and all simulations is $\sim 0.78 \pm 0.12$) than for CO_2 fluxes ($R^2 = 0.61 \pm 0.22$). This may suggest that water fluxes are more sensitive to weather variability than carbon fluxes in ORCHIDEE. On average, forcing ORCHIDEE driven by LOCAL meteorology gives a higher R^2 than with any of the gridded data sets. Across all fluxes and time scales, R^2 is slightly but not significantly higher with LOCAL ($R^2 = 0.52 \pm 0.26$) than with even the best modeled forcing, SAFRAN ($R^2 = 0.49 \pm 0.27$). On average for the daily scale, driving ORCHIDEE with LOCAL meteorology gives higher correlations than when using atmospheric analyzed meteorology, except for the LQE mountain grassland where using a modeled meteorology SAFRAN improves slightly but not significantly the value of daily R^2 over the five fluxes from 0.16 to 0.28. This indicates error compensation in ORCHIDEE, where a biased forcing compensates for a model bias.

4.3 Bias of modeled fluxes, function of driving meteorology

Figure 7a shows the MAE of modeled fluxes for ORCHIDEE driven by LOCAL and by each gridded data-product. Forcing ORCHIDEE with LOCAL meteorology compared to a gridded product delivers only a small reduction of MAE. The SAFRAN, EC-OPERA and ERA-I drivers result in more or less similar MAE values. On the other hand, REMO gives a higher MAE than all other model drivers. Differences in the MAE of NEE between different meteorological forcings are similar to those of GPP on diurnal and daily time scales. But on the monthly scale, the MAE of NEE differs from the one of GPP, as it also combines the climate-driven misfit of the model to TER. We note that MAE of TER is on average smaller than for GPP on the monthly scale (Fig. 7a).

One can also see from Fig. 7b that MAE differs largely between sites. In fact, the inter-site MAE differences are larger than the inter-meteorology MAE differences. This indicates that poorly captured ecosystem processes that control the model–data misfit differ at each site, and can be characterized despite biases in the meteorology used to drive ORCHIDEE.

5 The effect of meteorology in the error budget of ORCHIDEE

5.1 Model error and forcing error

The effects of uncertain meteorological forcing on ORCHIDEE modeled fluxes can be characterized by comparing the distance between flux simulations forced by different meteorology with the distance between simulated and observed flux. The total model error ε_{tot} is defined by:

$$\varepsilon_{\text{tot}} = \text{mean}[\text{RMSE}(F_{\text{sim}}(i) - F_{\text{obs}})] \quad (7)$$

where $F_{\text{sim}}(i)$ is the time series of simulated flux with ORCHIDEE driven by meteorology i , and F_{obs} is the observed flux. The forcing error is defined by:

$$\varepsilon_{\text{F}} = \text{mean}[\text{RMSE}(F_{\text{sim}}(i) - F_{\text{sim}}(\text{LOCAL}))] \quad (8)$$

where $F_{\text{sim}}(\text{LOCAL})$ is the flux simulated by ORCHIDEE driven by observed (LOCAL = true) meteorology at each site. The model error due to erroneous assumptions in the ORCHIDEE equations, due to errors in the value of parameters (Thornton et al., 2002; Zaehle et al., 2005; Mitchell et al., 2009) and due to incorrect initial conditions such as equilibrium spin up values of soil C and biomass pools (Carvalhais et al., 2010) are defined by:

$$\varepsilon_{\text{mod}} = \text{RMSE}(F_{\text{sim}}(\text{LOCAL}) - F_{\text{obs}}). \quad (9)$$

Estimating the above-defined ε values by 1-sigma standard deviations, it follows that:

$$\varepsilon_{\text{tot}}^2 = \varepsilon_{\text{F}}^2 + \varepsilon_{\text{mod}}^2 + 2 \times \text{COV}(\varepsilon_{\text{F}} \varepsilon_{\text{mod}}). \quad (10)$$

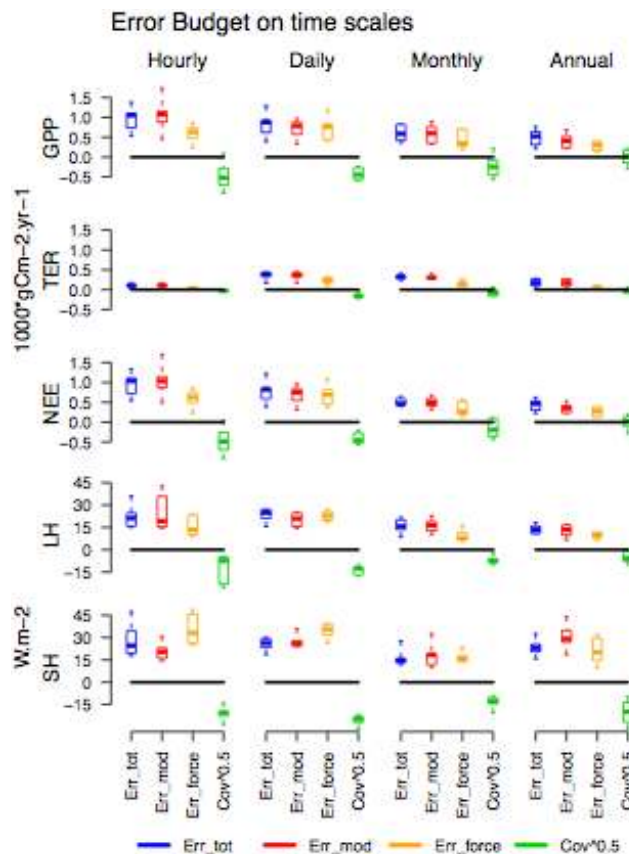


Fig. 8. Box plot of error estimation across 6 sites during peak growing season. The bottom and top of the box are the 25th and 75th percentile, respectively, and the band near the middle of the box is the median. The lower and upper ends of the box represent the minimum and maximum value. Err_tot, Err_mod and Err_force denote total model error, model error and forcing error, respectively. For the reason of plotting, covariance is expressed as the squared absolute covariance ($\text{cov}^{0.5}$) bearing the sign of covariance. Unit is $\text{kg Cm}^{-2} \text{yr}^{-1}$ and Wm^{-2} for carbon and water fluxes, respectively.

We are interested here in cov, the covariance term that denotes correlated errors between the model and biased meteorology. A positive covariance indicates that a biased meteorological driver will further degrade the model error, whereas a negative covariance indicates that a biased meteorology will compensate for model error to bring the simulated flux closer to the data. This may happen for instance if a too high GPP in summer, implied by a bias in meteorology, creates too many assimilates, which in turn increases litter respiration in the fall, and compensates for an underestimated respiration sensitivity to temperature (structural model error). The value of cov was calculated at each site from the ε_{tot} , ε_{F} , and ε_{mod} according to Eq. (10) and analyzed below. It is also worth noting that $\varepsilon_{\text{tot}} \neq \varepsilon_{\text{F}} + \varepsilon_{\text{mod}}$. In the case that cov is negative, ε_{F} and/or ε_{mod} can be larger than ε_{tot} . In this section, TER and NEE are modeled using the optimized procedure (Sect. 2.5).

5.2 Comparing model and forcing errors for different fluxes

Figure 8 provides the unbiased error distribution for each flux across the six sites. Unbiased error is defined from Eqs. (7)–(10) using unbiased RMSE, which is obtained from the analyzed time series based on Eq. (2) for hourly, daily, monthly and annual scales. One can see that for GPP and TER, the total error ε_{tot} (blue) is largely explained by model error ε_{mod} (red). For GPP, the share of model error in the total error is 1.01, 0.87, 0.92 and 0.82 on hourly, daily, monthly and annual time scales, respectively. For TER, the forcing error ε_{F} (orange) on the hourly scale is small ($\sim 75 \text{ gC m}^{-2} \text{ month}^{-1}$) because the diurnal variability is weak both in the modelings and in the measurement ($\sim 35 \text{ gC m}^{-2} \text{ month}^{-1}$). By contrast, on daily, monthly and annual time scales, the forcing error takes a significant share of the total uncertainty budget of TER, as shown by the ratio $\varepsilon_{\text{F}}/\varepsilon_{\text{tot}} = 0.63, 0.39$ and 0.35 , respectively. For NEE, the share of forcing error to total error is 0.50, 0.82, 0.64 and 0.50 on hourly, daily, monthly and annual time scales, respectively. Note that because of negative covariance between ε_{F} and ε_{mod} , the contribution of both error sources to ε_{tot} can be larger than 0.5. It is also seen that the contribution of forcing error to the total error is on average larger for LH and SH than for CO_2 fluxes, in particular on daily and annual time scales. The ratio $\varepsilon_{\text{F}}/\varepsilon_{\text{tot}}$ on daily and annual time scales is 1.04 and 0.78 for LH, 1.35 and 0.96 for SH, respectively. This forcing error is due primarily to biases in SW_{down} between the different forcings (see Sect. 5.4, Fig. 10).

The most interesting result is that the forcing error is not negligible compared to model errors. This comes a bit as a surprise because meteorology is generally assumed in vegetation modeling to be well enough known not to create a misfit in modeled fluxes. Using a simple daily model of coupled carbon and water fluxes, Spadavecchia et al. (2011) studied the uncertainty caused by meteorological drivers and concluded that driver uncertainty is relatively small, accounting for $\sim 10\%$ of the total flux, which is obviously lower than our estimation. We note that only the uncertainty in T_{air} and Rainfall were included in their study, which may neglect the forcing error caused by radiation.

We note also that ε_{mod} and ε_{F} decrease in absolute value with increasing temporal averaging scale, from typical errors values of 606 and $1077 \text{ gC m}^{-2} \text{ yr}^{-1}$ on the diurnal scale for ε_{F} and ε_{mod} , down to 407 and $293 \text{ gC m}^{-2} \text{ yr}^{-1}$ on the annual scale for GPP errors. For LH errors, the estimates of ε_{mod} decrease from about 20 W m^{-2} on the hourly scale down to 13.0 W m^{-2} on the annual scale, while ε_{F} decreases from 15 W m^{-2} to 10 W m^{-2} . It is intriguing to see that the covariance between forcing error and structural model error is negative for the six sites included in this study. This indicates that model errors are partly compensated by biases of meteorological forcing. This result may be due to similar biases among different forcings compared to LOCAL, which act

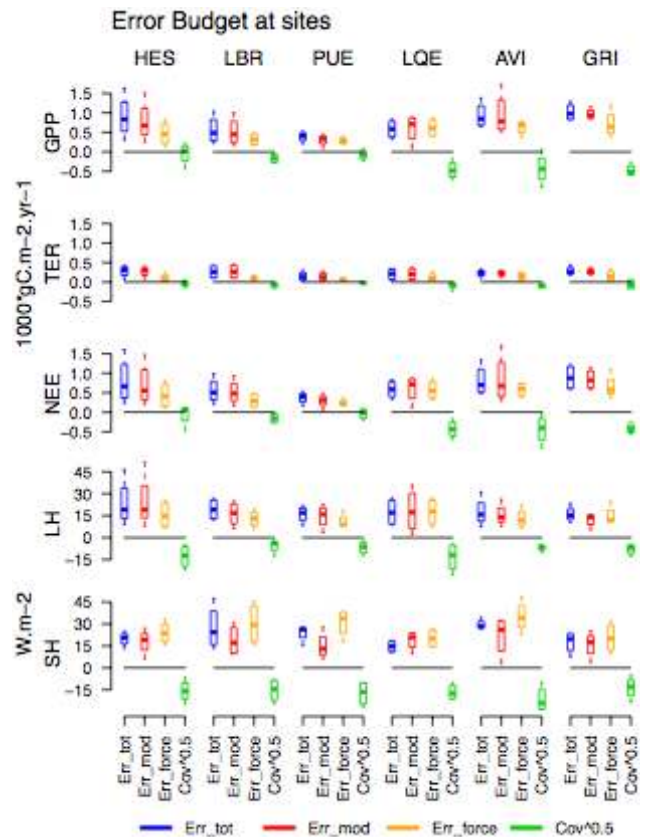


Fig. 9. Box plot of error estimation across 4 time scales at each site during peak growing season. The symbols of boxplots are defined the same as in Fig. 8.

to shift the simulated TER and GPP closer to the observed fluxes. One can see in Fig. 8 that the unbiased errors of TER are smaller than those of GPP, because TER is less sensitive to weather variability than GPP in ORCHIDEE. The NEE forcing errors are equally as large as the GPP forcing errors. This indicates that, even if NEE is the difference between GPP and TER, which both have similar sensitivities to weather variability, a forcing error on GPP will not be compensated by an error of the same magnitude on TER. The forcing errors of NEE are as large as those of GPP on diurnal to daily scales. But on the monthly scale, the forcing errors on NEE become smaller than those on GPP, indicating that compensation by TER errors, via labile pools impacted by GPP errors, might occur on this time scale, but not on shorter time scales. On an annual scale, forcing errors are of the same magnitude for NEE and GPP, because the climate sensitivity of TER is probably as large as that of GPP on long time scales, and the annual anomalies of each gross flux are partly decoupled because of their different seasonality (Piao et al., 2010; Vesala et al., 2010). Therefore at site scale, forcing errors reflecting discrepancies between atmospheric analyses and local meteorology impede an accurate simulation of the interannual fluctuations of NEE as much as the model errors.

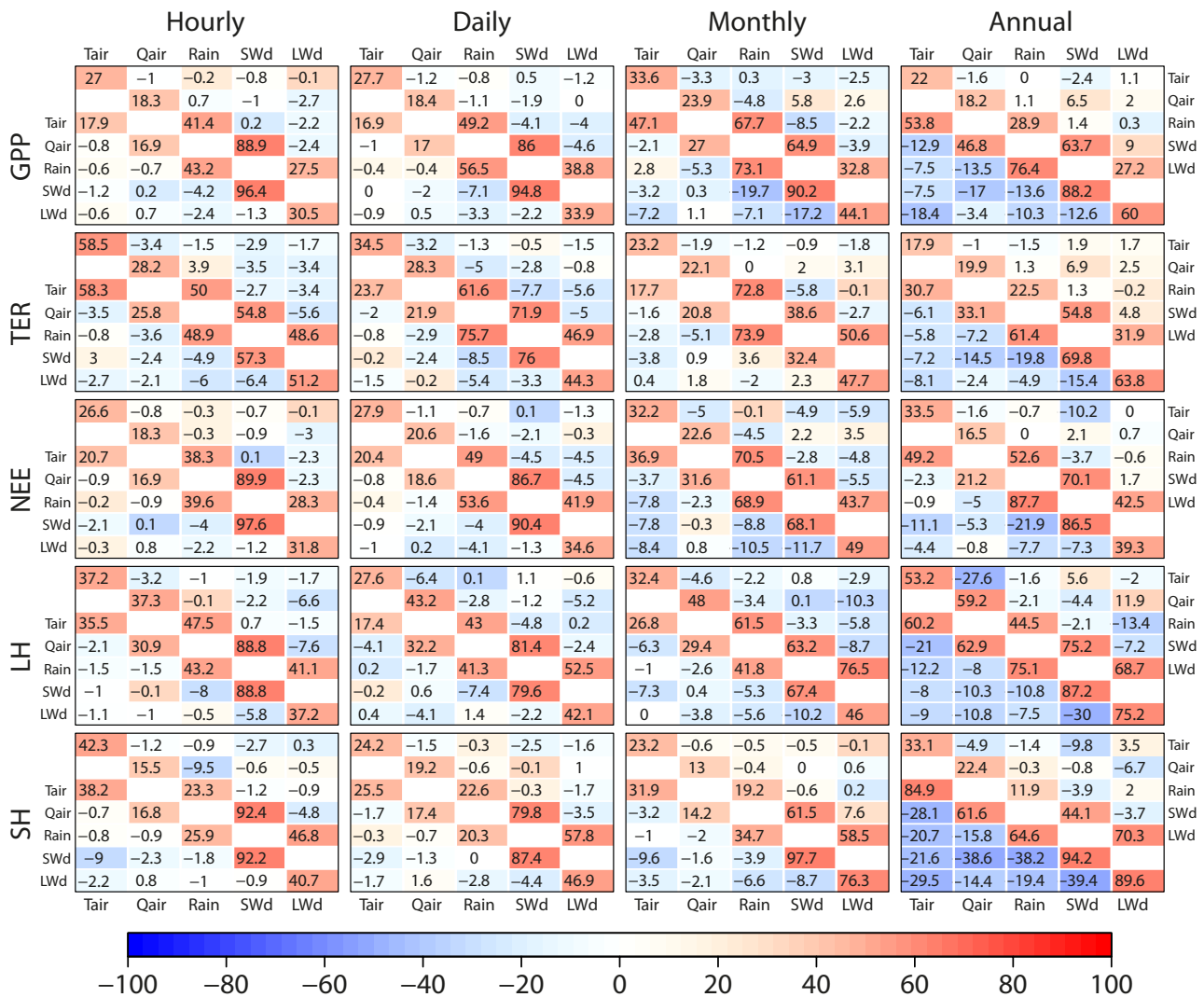


Fig. 10. Summary of forcing errors caused by meteorological drivers and the covariance between the meteorological variables. Diagonal terms: contribution of each meteorological variables; non-diagonal terms: contribution of pairs of meteorological variables. The upper-right triangle is calculated based on annual flux and the lower-left diagonal at peak growing season. See text for the details of calculation.

On annual time scales, the forcing error is an overlooked source of poor model performance in tackling the simulation of interannual GPP and TER. For interannual variability of carbon fluxes, the ratio ($\epsilon_F/\epsilon_{mod}$) averaged across the 6 sites is 0.80, compared to 0.55 for the monthly time scale, 0.86 for the daily scale and 0.63 for the hourly scale. The striking effects of meteorological forcing uncertainty on producing interannual variability of carbon fluxes was highlighted in the study of Jung et al. (2007), which showed that the total GPP over the European domain was 20% higher in the simulations driven with ECMWF than with REMO.

The study of Fagle et al. (2001) showed that errors introduced by gap-filling were directly proportional to the percentage of gaps filled during a period. For daytime, maximum observed errors were $\pm 0.20 \text{ gC m}^{-2}$ per cent of

day filled, $\pm 0.64 \text{ gC m}^{-2}$ per cent of month filled, and $\pm 0.83 \text{ gC m}^{-2}$ per percentage of year filled. The errors for nighttime gaps were in general smaller. They thus estimated that for a data set with 37% gaps (the typical percentage of gaps in flux tower data), the maximum error introduced would amount to $\pm 50 \text{ gC m}^{-2} \text{ yr}^{-1}$, which is comparable with the estimation ($\pm 25 \text{ gC m}^{-2} \text{ yr}^{-1}$) by Moffat et al. (2007). The forcing error ϵ_F for annual sums of NEE in our study is about $255 \sim 420 \text{ gC m}^{-2} \text{ yr}^{-1}$ at forest sites, largely beyond the measurement uncertainty. This conclusion is also applied to model error ϵ_m which is 20–50% larger than forcing error. On daily and monthly scales, forcing error and model error for NEE are also much larger than measurement uncertainty.

5.3 Comparing model and forcing errors across sites

Figure 9 provides a comparison of the forcing and model error at each site. Typical values of the ratio $\varepsilon_F/\varepsilon_{\text{mod}}$ for CO₂ fluxes range from 0.54 at HES to 0.92 at LQE. The ratio $\varepsilon_F/\varepsilon_{\text{mod}}$ for SH is larger than or close to 1, which likely reflects biases in LW_{down} forcing and a high sensitivity of SH to that driver.

Generally, the cropland and grassland sites have a larger share of forcing error to total error than the forests for CO₂ fluxes, with the ratio $\varepsilon_F/\varepsilon_{\text{tot}}$ at non-forest sites of 0.89, 0.61 and 0.88 against 0.55, 0.48 and 0.52 at forests for GPP, TER and NEE, respectively. In particular on the annual time scale, $\varepsilon_F/\varepsilon_{\text{tot}}$ at the LQE grassland in altitude are 1.90, 1.24 and 1.78 for GPP, TER and NEE, respectively, which are significantly higher than at any other site. The large forcing errors at this site are because the forcings used to drive ORCHIDEE at this site have different spatial resolutions over a complex and heterogeneous terrain, thus giving a larger spread of simulated fluxes.

5.4 Separate contribution of each meteorological variable to the forcing error

We carried out a series of factorial experiments in order to identify which meteorological variable has the largest impact on the forcing error ε_F . Individual forcing errors of the i th driver, $\varepsilon_F(i)$, are calculated by Eq. (8) with ORCHIDEE being run with LOCAL forcing, except for the i th driver that is taken from a gridded data-product (here SAFRAN). We also run the simulation where all the five meteorological variables are taken from SAFRAN, defining a total forcing error ε_F (Eq. 8). Figure 10 provides a comparison of the contribution from each meteorological driver to the total forcing error. The relative contribution $C(i)$ of the i th driver to the total forcing error is defined by:

$$C(i) = \varepsilon_F(i)/\varepsilon_F \times 100. \quad (11)$$

A contribution that exceeds 100 %, indicates compensations between errors induced by different SAFRAN variables. It is seen on the two diagonals of Fig. 10 (upper diagonal calculated with fluxes all year round and lower diagonal with fluxes covering peak growing season only) that SW_{down} has a dominant relative contribution to the forcing error associated with GPP, NEE, LH and SH. Error in this driver is thus critical in the uncertainty budget of simulated CO₂ and water fluxes. On the monthly time scale, the drivers that contribute to ε_F by order of decreasing importance (across the six sites) are SW_{down}, Rainfall, LW_{down}, T_{air} and Q_{air} . On the monthly scale, bias in LW_{down} becomes as important as bias in SW_{down} in the forcing error of SH. The importance of LW_{down}, a driver not systematically measured at flux tower sites and often overlooked in model studies, should not be underestimated. On the interannual time scale, the relative contribution of each meteorological driver to ε_F becomes com-

parable. This adds to the difficulty of reducing errors in the simulation of interannual flux variations, because uncertainty in each meteorological driver contributes significantly.

We now estimate the effect of using SW_{down} and T_{air} drivers from SAFRAN instead of LOCAL on the NEE and GPP forcing errors. Firstly, the effect of each variable taken separately is not additive. This is proven by calculating the covariance of ε_F between a simulation where both drivers are from SAFRAN, and factorial simulations where only one driven is from SAFRAN (Eq. 8). The results in Fig. 10 show that SW_{down} and T_{air} have distinct contributions to ε_F . SW_{down} is a dominant source of forcing error for GPP and NEE, explaining 95 % of ε_F , against 45 % only by T_{air} . Interestingly, the contributions of errors induced by each driver do not sum up to 100 %, implying covariance. The forcing error arising from bias in the pair of meteorological drivers i and j can be decomposed by:

$$\varepsilon_F(ij)^2 = \varepsilon_F(i)^2 + \varepsilon_F(j)^2 + 2 \times \text{cov}(ij). \quad (12)$$

Individual forcing errors $\varepsilon_F(i)$, $\varepsilon_F(j)$ and $\varepsilon_F(ij)$ are calculated by Eq. (8). The combined forcing error $\varepsilon_F(ij)$ is obtained by driving ORCHIDEE by with both drivers i and j being taken from SAFRAN.

The relative contribution $C(ij)$ to the error covariance $\text{cov}(ij)$ between drivers i and j is defined by:

$$C(ij) = \text{cov}(ij)/\varepsilon_F^2 \times 100. \quad (13)$$

The $C(ij)$ contributions are summarized by the upper and lower triangles of Fig. 10 (upper one calculated with annual fluxes and lower one with summer fluxes). Most of the covariances between pairs of drivers are found to be negative, indicating forcing error compensation effects in ORCHIDEE. We hypothesize that negative covariances between drivers arises from two effects. Firstly, biases in pairs of weather variables are likely to be correlated. For instance, a data-product overestimating SW_{down} should also overestimate T_{air} and underestimate Q_{air} . Secondly, the simulated ORCHIDEE fluxes in response to variation in drivers are likely to be in the same direction, implying that the effect of two biased variables is lower than the sum of the bias in each variable. Figure 10 shows that for interannual variability, however, the response of ORCHIDEE fluxes to errors in drivers could be different, thus causing positive or negative covariance in the contribution of pairs of meteorological drivers to the total forcing error. However, the positive values are rather small and less than 15 % for all fluxes.

6 Summary and discussion

In conclusion, we provide summary answers to the questions raised in the introduction.

1. We found large differences among analysis data sets, and between analysis data sets and site meteorology.

These differences are particularly important on annual time scales in terms of both amplitude and variability, and are particularly large for Q_{air} and SW_{down} . All analysis data sets well capture the phase of seasonal cycle in observed meteorology, but have problems in capturing hourly and daily variability depending on the meteorological variable and the analysis data set.

2. There is significant bias in analysis data sets compared to site-observed meteorology. The overall performance of SAFRAN, which has the highest spatial resolution and included observed meteorology by optimal interpolation, is slightly better than EC-OPERA and ERA-I. SAFRAN does a good job in reproducing observed Rainfall but is wrong with diurnal LW_{down} . REMO has a wet bias and overestimates LW_{down} by 15–20%. Meteorology at mountain sites (Laqueuille) is difficult to reproduce by atmospheric analysis because of the coarser representation of topography.
3. The forcing error is on average smaller than, but still comparable to the model error. In particular, the ratio $\varepsilon_F/\varepsilon_{\text{mod}}$ is 0.86 on the daily time scale and 0.80 on the annual time scale, indicating that modeling interannual flux variability using atmospheric analysis products as input to ORCHIDEE is limited as much by meteorology as by imperfect models. This result is likely to be generalized for other ecosystem models (Jung et al., 2007).
4. Forcing errors associated with pairs of meteorological variables are found to be negatively-correlated; that is, they partly compensate each other in the resulting flux. In our study SW_{down} has dominating contributions to the forcing errors for most fluxes on daily scale. On the annual scale, unfortunately, each meteorological variable appears to contribute evenly to the forcing error, making it more difficult to design a strategy for improving the simulation of fluxes.

In making the comparison of meteorological drivers among data sets, we take the local observation as a benchmark which is the best possible approximation to the “true” meteorology. However, we do not think the observations are error-free; moreover the comparison between atmospheric analysis data and single point measured data may not be so reasonable since the analysis data represent the area average, not a single point as the measured data does. Nevertheless, the data comparison carried out at least is better than without any data comparison. Our conclusions are essentially consistent with the previous findings – which included the comparison among REMO, ECMWF products and the local observation (Chen et al., 2007); between SAFRAN and ERA-I (Szczypta et al., 2011); and annual mean meteorology among several data sets (Ricciuto et al., 2009) – despite different benchmark data sets used in each study. The scale mismatch problem is addressed by conducting a comparison between SAFRAN

and its resolution-reduced version “SAFRAN-LOW” (see Appendix B). We found that the impact of spatial resolution on meteorology is negligible on hourly, daily and monthly scales. On the annual time scale, it is still smaller than but comparable to the differences among data sets, in particular for T_{air} and Rainfall.

Our study demonstrates that the estimation of LW_{down} is a major source of error. This variable is seldom validated against the measurements because it is not often observed either at climatological networks or flux towers. However, our study shows that the forcing error caused by LW_{down} on sensible heat flux (SH) is likely as much as those caused by other meteorological drivers, at least it is the case for ORCHIDEE. Another source of bias lies in the method used to convert daily meteorology to hourly values, which it is especially problematic for Q_{air} and LW_{down} . An improved conversion method is obviously needed for ORCHIDEE. Given the fact that most ecosystem models are driven by meteorology at hourly or higher time steps while the atmospheric analysis data are often at lower temporal resolutions, such as 6-h or daily, the conversion problem can be common for most ecosystem modeling groups and should be paid enough attention.

As an indispensable step but not the primary task of this study, we evaluated the performance of ORCHIDEE on different time scales in the context of carbon and water fluxes. We found that ORCHIDEE captures well the observed diurnal and seasonal cycles ($R^2 = 0.58 \pm 0.14$ and $R^2 = 0.68 \pm 0.20$, respectively), but shows pronounced model–data disagreement on daily and annual time scales ($R^2 = 0.17 \pm 0.13$ and $R^2 = 0.29 \pm 0.25$, respectively). The good agreement of model–data for diurnal and seasonal cycles is consistent with the results of previously published studies with ORCHIDEE (Anav et al., 2010; Keenan et al., 2009; Krinner et al., 2005; Santaren et al., 2007), and is also supported by the conclusions of Braswell et al. (2005), who made spectral analyses of a simple ecosystem carbon flux model. The substantial disagreement of model–data on daily and annual time scales, representing respectively synoptic and interannual variability, was reported by Mahecha et al. (2010), who made model–data comparison with spectral analyses considering two ecosystem models (one is ORCHIDEE). Recently Dietze et al. (2011) used wavelet analyses to synthesize the performance of 21 ecosystem models (including ORCHIDEE) at 9 eddy flux towers in North America. They concluded that model error is (1) largest at synoptic scale; (2) largest at seasonal course; and (3) large at the diurnal scale. Their results are consistent with ours for (1) but opposite for (2) and (3). The discrepancy in conclusions between our studies and theirs is not likely attributed to the analyses over time-frequency (i.e. spectral analyses). Nevertheless, both studies suggested that model development should focus on the understanding of synoptic process.

In short, uncertainty in meteorology is a limitation to the accurate modeling of flux variability. But this source

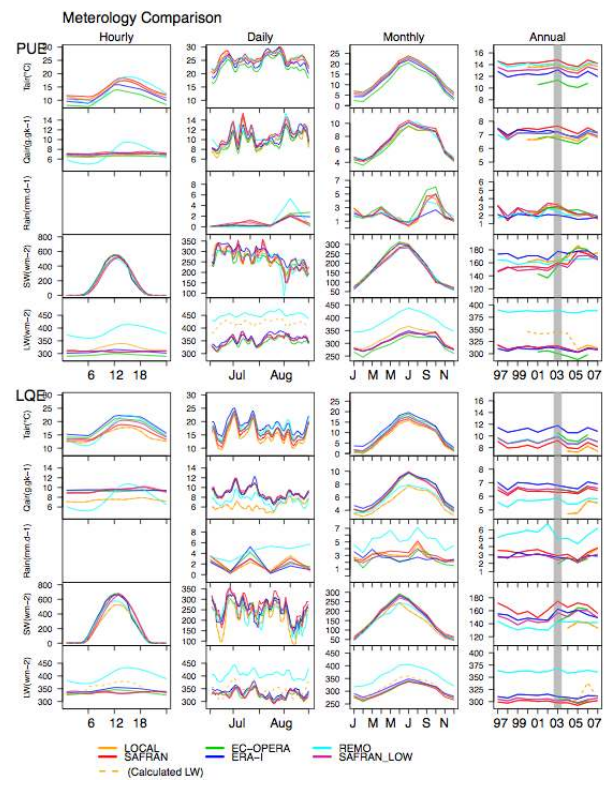
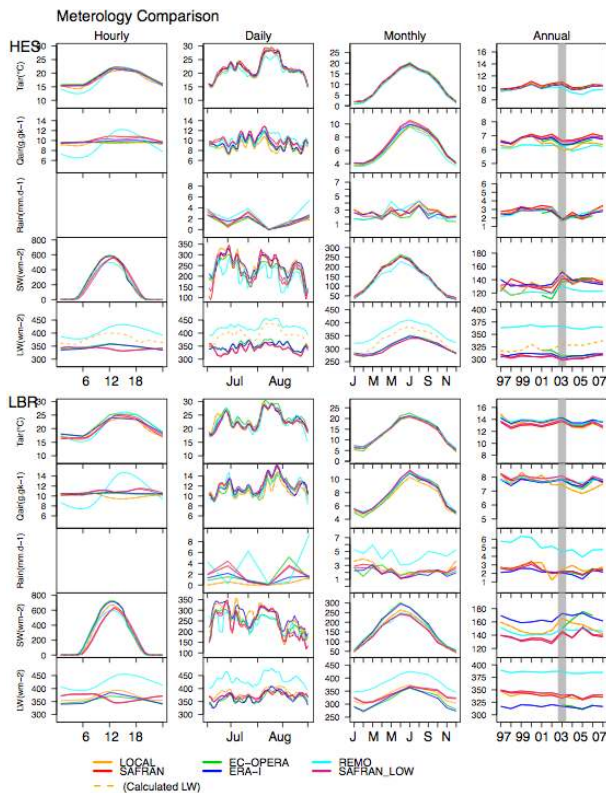


Fig. B1. Meteorological drivers in in situ and in gridded data sets at six sites. First column: hourly mean diurnal cycle over peak growing season (PGS). Second column: daily mean with a running mean of 3 days for July–August of 2003 at HES, LBR and PUE and of 2005 for LQE, AVI and GRI. Rainfall is calculated as 5-day aggregated values; third column: monthly mean seasonal cycle; fourth column: annual mean. The hourly mean diurnal cycle and monthly mean seasonal cycle correspond to 2004–2007 except for EC-OPERA (2003 to 2006). In the case of site-year without measured LW_{down} , calculated LW_{down} is plotted in dash lines.

Fig. B1. Continued.

of uncertainty, often overlooked or considered small, is in fact significant when trying to use data-driven process-based models to upscale fluxes, in particular on the annual time scale.

Appendix A

Description of gridded meteorological products

SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige) is a mesoscale atmospheric analysis system for surface variables covering France. Analyses are generated at spatial resolution of 8 km × 8 km with an hour time step using ground data observations and meteorological simulations from the French weather service. Validation of the SAFRAN product by Quintana-Segui et al. (2008) indicated a general good agree-

ment with individual station data. Temperature, precipitation and relative humidity were found to be faithfully reproduced, with $R^2 > 0.85$ and negligible systematic bias. SW_{down} has some bias, especially in coastal areas. The annual mean bias of SW_{down} is about 2 %, and the RMSE was found to be significant. The LW_{down} field has a positive bias during the winter and a negative bias during the rest of the year. This variable was independently evaluated against observations at two Météo-France long-term radiation monitoring sites during two years (these sites are independent from the flux towers of this study). It was concluded that LW_{down} from SAFRAN compares correctly with daily mean in situ observations, but that there was a discrepancy at the hourly time step. In addition, LW_{down} was found to be underestimated at these two sites by SAFRAN by 8 to 32 Wm^{-2} .

EC-OPERA is the result of the ECMWF operational forecasting system (<http://www.ecmwf.int>) used in this study between 2001 and 2006. The data are produced by the version T511L60 of ECMWF system, with approximately 40 × 40 km horizontal resolution. Temporal resolution of the data set is 6-hourly (00:00, 06:00, 12:00 and 18:00 UTC). We linearly interpolated all variables into half-hourly values to drive ORCHIDEE. The ECMWF operational system has been updated through time. Further information on the evolution of the operational system can be found at: http://www.ecmwf.int/products/data/operational_system/evolution/index.html.

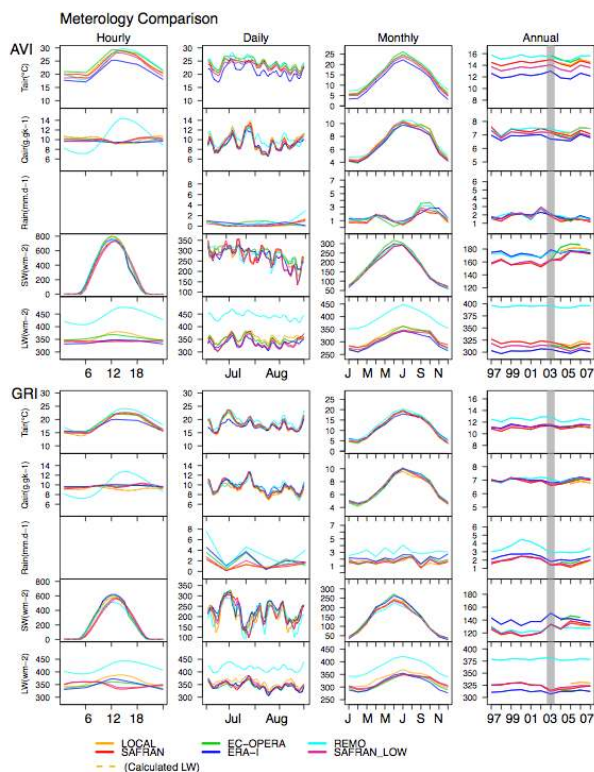


Fig. B1. Continued.

ERA-I is a globally consistent reanalysis of the meteorological fields which was produced by the Optimum Interpolation analysis and the assimilating model (T255L60) at ECMWF, starting in 1989 and continuing in real time. Spatial resolution is 79×79 km. Temporal resolution is 6-hourly, like EC-OPERA. Compared to the former ECMWF reanalysis product ERA-40 (Uppala et al., 2005), the new ERA-I benefits from several improvements of the ECMWF forecasting system (Berrisford et al., 2009).

REMO here refers to a specific simulation by the regional climate model REMO (REgionalMOdel, Jacob and Podzun, 1997) over Europe, forced by 6-hourly NCEP reanalysis (Kalnay et al., 1996) at lateral boundaries from 1948 to 2007. The REMO physics are adapted from the ECHAM4 global model of the Max-Planck-Institute for Meteorology (Koch and Feser, 2006). The specific REMO simulation used here was performed with nudging of large scale meteorological fields (von Storch et al., 2000) to keep the simulated state close to the driving state at larger scales, while allowing the model to freely generate regional-scale weather within Europe (Feser and Weisse, 2001). The atmospheric fields are archived on a daily step at 25×25 km resolution. Daily REMO data were disaggregated into half-hourly values for driving ORCHIDEE by using the gap-filling process described in Sect. 1.3. The same REMO forcing was used for an inter-comparison of vegetation models in the CARBOEUROPE project (Vetter et al., 2008). A brief evaluation

of the data by Chen et al. (2007) showed that REMO reproduced well the observed temperature, but had difficulties in reproducing precipitation and radiation, with a dry bias in Mediterranean regions.

Appendix B

Gridded meteorological products and spatial resolution

In evaluating gridded meteorological forcing against flux tower measurements (LOCAL), we note that the spatial resolution in LOCAL and gridded data sets are much different. In particular, the spatial resolution in the gridded data sets varied from high resolution (8 km) in SAFRAN to coarse resolution in ERA-I (79 km). The values in the grid represent the mean meteorological state over an area of 8×8 km or of 79×79 km. The comparison between LOCAL and gridded data and among gridded data may suffer from the problem of scale mismatch. To address this question, we create a “low” resolution SAFRAN data set; that is, centred on the interested site, meteorological variables in SAFRAN are averaged over the grids included in an area of 80×80 km. This newly created meteorological data set is referred to as “SAFRAN-LOW” to differ the high resolution “SAFRAN”. Obviously “SAFRAN-LOW” at each site represents the mean meteorological state over 80×80 km. “SAFRAN-LOW” is then treated as an independent data set, along with LOCAL, EC-OPERA, ERA-I and REMO, being evaluated against SAFRAN at four time scales (Fig. B1). The statistics are shown in Figs. B2 for R^2 and B3 for MAE, indicating how these data sets match or mismatch against SAFRAN.

At hourly, daily and monthly time scales, the differences between SAFRAN-LOW and SAFRAN are negligible. In Fig. B1, the curves of each meteorological variable between the high and low resolution SAFRAN data sets (red and violet lines, respectively) nearly overlap or are very close to each other. The overall R^2 between SAFRAN-LOW and SAFRAN (Fig. B2) across all meteorological variables and sites is about 0.87, 0.92 and 0.97 on hourly, daily and monthly time scales, respectively, significantly higher ($p < 0.01$) than R^2 between SAFRAN with any other data sets, including LOCAL. Rainfall is the variable impacted most by spatial resolution in that R^2 between SAFRAN-LOW and SAFRAN is about 0.72 ($p < 0.01$) at the hourly scale and 0.82 ($p < 0.01$) at the daily scale, significantly higher ($t = .34$, $df = 9.26$, $p < 0.01$) than R^2 between SAFRAN and LOCAL ($R^2 \approx 0.20$ for hourly and 0.55 for daily scales). On the monthly scale, the difference in R^2 is not significant any more ($t = 1.22$, $df = 6.3$, $p = 0.26$). The difference between each data set and SAFRAN as measured by MAE shows (Fig. B3) that the SAFRAN-LOW is significantly lower than any other data set on hourly, daily and monthly scales.

On the annual time scale, however, we find that the difference in meteorological variables between SAFRAN-LOW

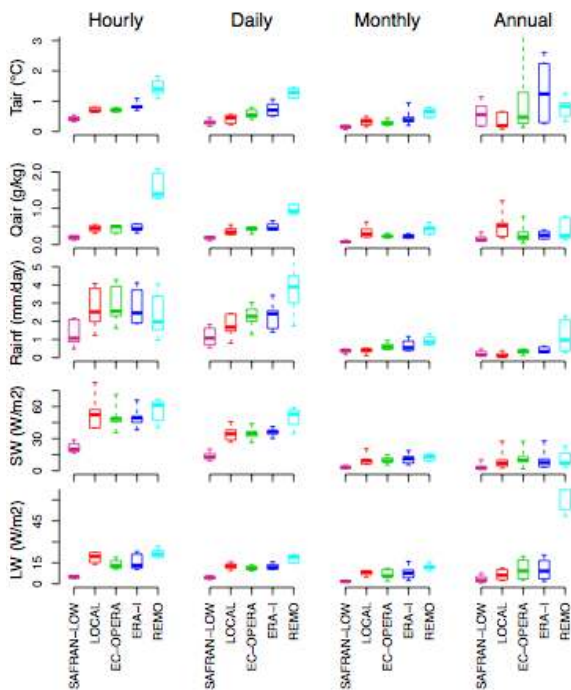


Fig. B3. Mean absolute error (MAE) between meteorological data sets and SAFRAN over 2004 to 2007, except for EC-OPERA which covers only 2004–2006. Time series used to calculate MAE correspond to growing season except for at the annual time scale, which covers continuously from 1997 to 2007, whenever data available.

In summary, the uncertainty in meteorological variables caused by spatial resolution in the cases of SAFRAN and SAFRAN-LOW is significantly smaller than that caused by reanalysis models, but it is not negligible for T_{air} and Rainfall on the annual time scale. Simulations of carbon and water fluxes by ORCHIDEE driven by SAFRAN and SAFRAN-LOW are not significantly different at our examined time scales (not shown), indicating forcing and model error compensation mechanisms in ORCHIDEE.

Acknowledgements. This study was conducted within the framework of the CarboFrance Project “Impact des extrêmes climatiques sur les flux de carbone” (GICC, French project). We wish to thank Frederic Chevallier and three anonymous reviewers for their constructive comments.

Edited by: E. Falge



The publication of this article is financed by CNRS-INSU.

References

- Anav, A., D'Andrea, F., Viovy, N., and Vuichard, N.: A validation of heat and carbon fluxes from high-resolution land surface and regional model, *J. Geophys. Res.*, 115, G04016, doi:10.1029/2009JG001178, 2010.
- Aubinet, M., Grelle, A., Ibrom, A., Rannik, S., Moncrieff, J., Foken, T., Kowalski, A., Martin, P., Berbigier, P., Bernhofer, C., Clement, R. Elbers, J., Granier, A., Grunwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., and Vesala, T.: Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology, *Adv. Ecol. Res.*, V30, 113–175, 2000.
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., and Fuentes, J.: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem-Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities., *B. Am. Meteorol. Soc.*, 82, 2415–2434, 2001.
- Ball, J., Woodrow, I., and Berry, J.: A model predicting stomatal conductance and its contribution to the control of photosynthesis under different environmental conditions, *Progress in Photosynthesis Research*, 4, 221–224, doi:10.1007/BF02180320, 1987.
- Berrisford, P., Dee, D., Fielding, K., Fuentes, M., Kallberg, P., Kobayahi, S., and Uppala, S.: The ERA-Interim archive, ERA Report Series, 2009.
- Botta, A., Viovy, N., Ciais, P., Friedlingstein, P., and Monfray, P.: A global prognostic scheme of leaf onset using satellite data, *Global Change Biol.*, 6, 709–725, doi:10.1046/j.1365-2486.2000.00362.x, 2000.
- 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, *Global Change Biol.*, 11, 335–355, doi:10.1111/j.1365-2486.2005.00897.x, 2005.
- Campbell, G. S. and Norman, J. M.: An introduction to environmental biophysics, Springer-Verlag, New York, 1998.
- Carvallhais, N., Reichstein, M., Ciais, P., Collatz, J., Mahecha, M., Montagnani, L., Papale, D., Rambal, S., and Seixas, J.: Identification of vegetation and soil carbon pools out of equilibrium in a process model via eddy covariance and biometric constraints, *Global Change Biol.*, 16, 2813–2829, 2010.
- Cellier, P., Dizengremel, P., Castell, J. F., Biolley, J. P., Le Thiec, D., Bethenod, O., Roche, R., Lebard, S., and Goujet, R.: BIOPOL-LATM: a French national project to analyse the biosphere-atmosphere interactions in the context of air pollution, Workshop “Establishing ozone critical levels II”, United Nations Economic Commission for Europe, Göteborg (SWE), poster, 2002.
- Chen, Y., Churkina, G., and Heimann, M.: A comparison of regional climate variables between various data sources, Max-Planck-Institut for Biogeochemie, Jena, 2007.
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A. D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T., and Valentini, R.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, *Nature*, 437, 529–533,

- doi:10.1038/nature03972, 2005.
- Dietze, M. C., Vargas, R., Richardson, A. D., Stoy, P. C., Barr, A. G., Anderson, R. S., Arain, M. A., Baker, I. T., Black, T. A., Chen, J. M., Ciais, P., Flanagan, L. B., Cough, C. M., Grant, R. F., Hollinger, D., Izaurralde, R. C., Kucharik, C. J., Laffeur, P., Liu, S. G., Lokupitiya, E., Luo, Y. Q., Munger, J. W., Peng, C. H., Poulter, B., Price, D. T., Ricciuto, D. M., Riley, W. J., Sahoo, A. K., Schaefer, K., Suyker, A. E., Tian, H. Q., Tonitto, C., Verbeeck, H., Verma, A. B., Wang, W. F., and Weng, E. S.: Characterizing the performance of ecosystem models across time scales: A spectral analysis of the North American Carbon Program site-level synthesis, *J. Geophys. Res.*, 116, G04029, doi:10.1029/2011JG001661, 2011.
- Delzon, S. and Loustau, D.: Age-related decline in stand water use: sap flow and transpiration in a pine forest chronosequence, *Agr. Forest Meteorol.*, 129, 105–119, doi:10.1016/j.agrformet.2005.01.002, 2005.
- de Rosnay, P. and Polcher, J.: Modelling root water uptake in a complex land surface scheme coupled to a GCM, *Hydrol. Earth Syst. Sci.*, 2, 239–255, doi:10.5194/hess-2-239-1998, 1998.
- Ducoudré, N. I., Laval, K., and Perrier, A.: SECHIBA, a New Set of Parameterizations of the Hydrologic Exchanges at the Land-Atmosphere Interface within the LMD Atmospheric General Circulation Model, *J. Climate*, 6, 248–273, 1993.
- Durand, Y., Brun, E., Mérindol, L., Guyomarch, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, *Ann. Glaciol.*, 18, 65–71, 1993.
- Durand, Y., Giraud, G., Brun, E., Mérindol, L., and Martin, E.: A computer-based system simulating snowpack structures as a tool for regional avalanche forecast, *J. Glaciol.*, 45, 469–484, 1999.
- Falge, E., Baldocchi, D., Olson, R., Anthoni, P., Aubinet, M., Bernhofer, C., Burba, G., Ceulemans, R., Clement, R., Dolman, H., Granier, A., Gross, P., Grunwald, T., Hollinger, D., Jensen, N. O., Katul, G., Keronen, P., Kowalski, A., Lai, C. T., Law, B. E., Meyers, T., Moncrieff, J., Moors, E., Munger, J. W., Pilegaard, K., Rannik, U., Rebmann, C., Suyker, A., Tenhunen, J., Tu, K., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: Gap filling strategies for defensible annual sums of net ecosystem exchange, *Agr. Forest Meteorol.*, 107, 43–69, doi:10.1016/S0168-1923(00)00225-2, 2001.
- Farquhar, G. D., Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO₂ assimilation in leaves of C3 species, *Planta*, 149, 78–90, doi:10.1007/BF00386231, 1980.
- Feser, F. and Weisse, R.: Multi-decadal atmospheric modeling for Europe yields multi-purpose data, *EOS Transactions*, 82, 305, doi:10.1029/01EO00176, 2001.
- Friedlingstein, P., Joel, G., Field, C. B., and Fung, I. Y.: Toward an allocation scheme for global terrestrial carbon models, *Global Change Biol.*, 5, 755–770, doi:10.1046/j.1365-2486.1999.00269.x, 1999.
- Friedlingstein, P., Cox, P., Betts, R., Bopp, L., von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., Govindasamy, B., 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 C4MIP model intercomparison, *J. Climate*, 19, 3337–3353, 2006.
- GLOBALVIEW-CO₂: Cooperative Atmospheric Data Integration Project – Carbon Dioxide. CD-ROM, NOAA ESRL, Boulder, Colorado, available on Internet via anonymous FTP to ftp.cmdl.noaa.gov, Path: ccg/co2/GLOBALVIEW, 2011.
- Garnier, E., Lavorel, S., Ansquer, P., Castro, H., Cruz, P., Dolezal, J., Eriksson, O., Fortunel, C., Freitas, H., Golodets, C., Grigulis, K., Jouany, C., Kazakou, E., Kigel, J., Kleyer, M., Lehsten, V., Lepš, J., Meier, T., Pakeman, R., Papadimitriou, M., Papanastasis, V. P., Quested, H., Quéfier, F., Robson, M., Roumet, C., Rusch, G., Skarpe, M., Sternberg, M., Theau, J.-P., Thébault, A., Vile, D., and Zarovali, M.: Assessing the effects of land use change on plant traits, communities and ecosystem functioning in grasslands: a standardized methodology and lessons from an application to 11 European sites, *Ann. Bot.-London*, 99, 967–985, 2007.
- Jacob, D. and Podzun, R.: Sensitivity studies with the regional climate model REMO, *Meteorol. Atmos. Phys.*, 63, 119–129, doi:10.1007/BF01025368, 1997.
- Jarosz, N., Brunet, Y., Lamaud, E., Irvine, M., Bonnefond, J., and Loustau, D.: Carbon dioxide and energy flux partitioning between the understorey and the overstorey of a maritime pine forest during a year with reduced soil water availability, *Agr. Forest Meteorol.*, 148, 1508–1523, doi:10.1016/j.agrformet.2008.05.001, 2008.
- Jung, M., Le Maire, G., Zaehle, S., Luyssaert, S., Vetter, M., Churkina, G., Ciais, P., Viovy, N., and Reichstein, M.: Assessing the ability of three land ecosystem models to simulate gross carbon uptake of forests from boreal to Mediterranean climate in Europe, *Biogeosciences*, 4, 647–656, doi:10.5194/bg-4-647-2007, 2007.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, B., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Roy, J., and Dennis, J.: The NCEP/NCAR 40-Year Reanalysis Project, *B. Am. Meteorol. Soc.*, 77, 437–471, 1996.
- Katul, G., Lai, C.-T., Schäfer, K., Vidakovic, B., Albertson, J., Ellsworth, D., and Oren, R.: Multiscale analysis of vegetation surface fluxes: from seconds to years, *Adv. Water Resour.*, 24, 1119–1132, doi:10.1016/S0309-1708(01)00029-X, 2001.
- Keenan, T., García, R., Friend, A. D., Zaehle, S., Gracia, C., and Sabate, S.: Improved understanding of drought controls on seasonal variation in Mediterranean forest canopy CO₂ and water fluxes through combined in situ measurements and ecosystem modelling, *Biogeosciences*, 6, 1423–1444, doi:10.5194/bg-6-1423-2009, 2009.
- Knorr, W., Prentice, I. C., House, J. I., and Holland, E. A.: Long-term sensitivity of soil carbon turnover to warming, *Nature*, 433, 298–301, doi:10.1038/nature03226, 2005.
- Koch, W. and Feser, F.: Relationship between SAR-Derived Wind Vectors and Wind at 10-m Height Represented by a Mesoscale Model, *Mon. Weather Rev.*, 134, 1505–1517, 2006.
- Krinner, G., Viovy, N., Noblet-Ducoudré, N. D., Ogée, 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:200510.1029/2003GB002199, 2005.
- Longdoz, B., Gross, P., and Granier, A.: Multiple quality tests for analysing CO₂ fluxes in a beech temperate forest, *Biogeosciences*, 5, 719–729, doi:10.5194/bg-5-719-2008, 2008.

- Mahecha, M. D., Reichstein, M., Lange, H., Carvalhais, N., Bernhofer, C., Grünwald, T., Papale, D., and Seufert, G.: Characterizing ecosystem-atmosphere interactions from short to interannual time scales, *Biogeosciences*, 4, 743–758, doi:10.5194/bg-4-743-2007, 2007.
- McMurtrie, R., Rook, D., and Kelliher, F.: Modelling the yield of *Pinus radiata* on a site limited by water and nitrogen, *Forest Ecol. Manage.*, 30, 381–413, doi:10.1016/0378-1127(90)90150-A, 1990.
- Medlyn, B. E., Loustau, D., and Delzon, S.: Temperature response of parameters of a biochemically based model of photosynthesis. I. Seasonal changes in mature maritime pine (*Pinus pinaster* Ait.), *Plant Cell Environ.*, 25, 1155–1165, doi:10.1046/j.1365-3040.2002.00890.x, 2002.
- Medlyn, B. E., Robinson, A. P., Clement, R., and McMurtrie, R. E.: On the validation of models of forest CO₂ exchange using eddy covariance data: some perils and pitfalls, *Tree Physiol.*, 25, 839–857, 2005.
- Mitchell, S., Beven, K., and Freer, J.: Multiple sources of predictive uncertainty in modeled estimates of net ecosystem CO₂ exchange, *Ecol. Model.*, 220, 3259–3270, doi:10.1016/j.ecolmodel.2009.08.021, 2009.
- 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., Jeffrey, H., 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, *Agr. Forest Meteorol.*, 147, 209–232, doi:10.1016/j.agrformet.2007.08.011, 2007.
- Moncrieff, J. B., 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.
- Olioso, A., Inoue, Y., Ortega-Farias, S., Demarty, J., Wigneron, J., Braud, I., Jacob, F., Lecharpentier, P., Ottlé, C., Calvet, J., and Brisson, N.: Future directions for advanced evapotranspiration modeling: Assimilation of remote sensing data into crop simulation models and SVAT models, *Irrigation and Drainage Systems*, 19, 377–412, doi:10.1007/s10795-005-8143-z, 2005.
- 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: algorithms and uncertainty estimation, *Biogeosciences*, 3, 571–583, doi:10.5194/bg-3-571-2006, 2006.
- Parton, W. J., Stewart, J. W. B., and Cole, C. V.: Dynamic of C, N, P and S in grassland soils: a model, *Biogeochemistry*, 5, 109–131, 1988.
- Peylin, P., Bousquet, P., Quéré, C. L., Sitch, S., Friedlingstein, P., McKinley, G., Gruber, N., Rayner, P., and Ciais, P.: Multiple constraints on regional CO₂ flux variations over land and oceans, *Global Biogeochem. Cy.*, 19, GB1011, doi:200510.1029/2003GB002214, 2005.
- Piao, S., Luyssaert, S., Ciais, P., Janssens, I. A., Chen, A., Cao, C., Fang, J., Friedlingstein, P., Luo, Y., and Wang, S.: Forest annual carbon cost: a global-scale analysis of autotrophic respiration, *Ecology*, 91, 652–661, doi:10.1890/08-2176.1, 2010.
- Prentice, I. C., Farquhar, G. D., Fasham, M. J. R., Goulden, M., Heimann, L. M., Jaramillo, V. J., Khashgi, H. S., Le Quéré, C., Scholes, R. J. D., and Wallace W. R.: The carbon cycle and atmospheric carbon dioxide, Chapter 3, in: *Climate Change 2001*, edited by: IPCC, the scientific basis, Cambridge University Press, 2001.
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S.: Analysis of Near-Surface Atmospheric Variables: Validation of the SAFRAN Analysis over France, *J. Appl. Meteorol. Climatol.*, 47, 92–107, 2008.
- Rambal, S., Ourcival, J., Joffre, R., Mouillot, F., Nouvellon, Y., Reichstein, M., and Rocheteau, A.: Drought controls over conduction and assimilation of a Mediterranean evergreen ecosystem: scaling from leaf to canopy, *Global Change Biol.*, 9, 1813–1824, doi:10.1111/j.1365-2486.2003.00687.x, 2003.
- Reichstein, M., Tenhunen, J., Rouspard, O., Ourcival, J., Rambal, S., Miglietta, F., Peressotti, A., Pecchiari, M., Tirone, G., and Valentini, R.: Inverse modeling of seasonal drought effects on canopy CO₂/H₂O exchange in three Mediterranean ecosystems, *J. Geophys. Res.-Atmos.*, 108, 4726, doi:10.1029/2003JD003430, 2003.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Valentini, R., Aubinet, M., Berbigier, P., Bernhofer, C., Buchman, N., Gilmanov, T., Granier, A., Grünwald, T., Havrankova, K., Janous, D., Knohl, A., Laurela, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biol.*, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.
- Ricciuto, D. M., Thornton, P. E., Schaefer, K., Cook, R. B., Davis, K. J., and Synthesis, N.: How uncertainty in gap-filled meteorological input forcing at eddy covariance sites impacts modeled carbon and energy flux (Invited), *AGU*, 90(52), Fall Meet. Suppl., Abstracts, B54A-03, 2009.
- 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., Verma, S. B., and Wofsy, S. B.: A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes, *Agr. Forest Meteorol.*, 136, 1–18, 2006.
- Richardson, A. D., Hollinger, D., Aber, J., Ollinger, S. V., and Braswell, B.: Environmental variation is directly responsible for short- but not long-term variation in forest-atmosphere carbon exchange, *Global Change Biol.*, 13, 788–803, doi:10.1111/j.1365-2486.2007.01330.x, 2007.
- Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A. M., Papale, D., Reichstein, M., Stauch, V. J., Braswell, B. H., Churkina, G., Kruijt, B., and Hollinger, D. Y.: Statistical properties of random CO₂ flux measurement uncertainty inferred from model residuals, *Agr. Forest Meteorol.*, 148, 38–50, 2008.
- Richardson, C. W. and Wright, D. A.: WGEN: A model for generating daily weather variables. US Department of Agriculture, Agricultural Research Service, ARS-8, USDA Washington, DC, 1984.
- 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.
- Schimel, D. S., House, J. I., Hibbard, K. A., Bousquet, P., Ciais, P., Peylin, P., Braswell, B. H., Apps, M. J., Baker, D., Bondeau,

- A., Canadell, Churkina, J., Cramer, W., Friedl, C. B., Friedlingstein, P., Goodale, C., Heimann, M., Houghton, R. A., Melillo, J. M., Moore, B., Murdiyarso, D., Noble, I., Pacala, S. W., Prentice, I. C., Raupach, M. R., Rayner, P. J., Steffen, W. L., Wirth, C., House, J. I., Denning, A. S., Scholes, R. J., Steffen, W. L., and Wirth, C.: Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems, *Nature*, 414, 169–172, doi:10.1038/35102500, 2001.
- 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, doi:10.1046/j.1365-2486.2003.00569.x, 2003.
- Soussana, J. F., Loiseau, P., Vuichard, N., Ceschia, E., Balesdent, J., Chevallier, T., and Arrouays, D.: Carbon cycling and sequestration opportunities in temperate grasslands, *Soil Use Manage.*, 20, 219–230, 2004.
- Spadavecchia, L., Williams, M., and Law, B. E.: Uncertainty in predictions of forest carbon dynamics: separating driver error from model error, *Ecol. Appl.*, 21, 1506–1522, doi:10.1890/09-1183.1, 2011.
- Stoy, P. C., Katul, G. G., Siqueira, M. B. S., Juang, J., McCarthy, H. R., Kim, H., Oishi, A. C., and Oren, R.: Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: a wavelet analysis, *Tree Physiol.*, 25, 887–902, doi:10.1093/treephys/25.7.887, 2005.
- Stoy, P. C., Richardson, A. D., Baldocchi, D. D., Katul, G. G., Stanovick, J., Mahecha, M. D., Reichstein, M., Detto, M., Law, B. E., Wohlfahrt, G., Arriga, N., Campos, J., McCaughey, J. H., Montagnani, L., Paw U, K. T., Sevanto, S., and Williams, M.: Biosphere-atmosphere exchange of CO₂ in relation to climate: a cross-biome analysis across multiple time scales, *Biogeosciences*, 6, 2297–2312, doi:10.5194/bg-6-2297-2009, 2009.
- Szczypta, C., Calvet, J.-C., Albergel, C., Balsamo, G., Boussetta, S., Carrer, D., Lafont, S., and Meurey, C.: Verification of the new ECMWF ERA-Interim reanalysis over France, *Hydrol. Earth Syst. Sci.*, 15, 647–666, doi:10.5194/hess-15-647-2011, 2011.
- Teuling, A. J., Seneviratne, S. I., Stockli, R., Reichstein, M., Moors, E., Ciais, P., Luyssaert, S., van den Hurk, B., Ammann, C., Bernhofer, C., Dellwik, E., Gianelle, D., Gielen, B., Grunwald, T., Klumpp, K., Montagnani, L., Moureaux, C., Sottocornola, M., and Wohlfahrt, G.: Contrasting response of European forest and grassland energy exchange to heatwaves, *Nat. Geosci.*, 3, 722–727, 2010.
- Thornton, P. E., Law, B. E., Gholz, H. L., Clark, K. L., Falge, E., Ellsworth, D. S., Goldstein, 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, *Agr. Forest Meteorol.*, 113, 185–222, doi:10.1016/S0168-1923(02)00108-9, 2002.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, I., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, *Q. J. Roy. Meteorol. Soc.*, 131, 2961–3012, doi:10.1256/qj.04.176, 2005.
- Vesala, T., Launiainen, S., Kolari, P., Pumpanen, J., Sevanto, S., Hari, P., Nikinmaa, E., Kaski, P., Mannila, H., Ukkonen, E., Piao, S. L., and Ciais, P.: Autumn temperature and carbon balance of a boreal Scots pine forest in Southern Finland, *Biogeosciences*, 7, 163–176, doi:10.5194/bg-7-163-2010, 2010.
- 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, doi:10.5194/bg-5-561-2008, 2008.
- von Storch, H., Langenberg, H., and Feser, F.: A Spectral Nudging Technique for Dynamical Downscaling Purposes, *Mon. Weather Rev.*, 128, 3664–3673, 2000.
- 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:200510.1029/2004GB002395, 2005.
- Zaitchik, B. F., Macalady, A. K., Bonneau, L. R., and Smith, R. B.: Europe's 2003 heat wave: a satellite view of impacts and land-atmosphere feedbacks, *Int. J. Climatol.*, 26, 743–769, doi:10.1002/joc.1280, 2006.