Biogeosciences Discuss., 8, 2467–2522, 2011
www.biogeosciences-discuss.net/8/2467/2011/
doi:10.5194/bgd-8-2467-2011

Biogeosciences Discussions

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

Y. Zhao[1], P. Ciais[1], P. Peylin[2], N. Viovy[1], B. Longdoz[3], J. M. Bonnefond[4], S. Rambal[5], K. Klumpp[6], A. Olioso[7], P. Cellier[8], F. Maignan[1], T. Eglin[1], and J. C. Calvet[9]

[1]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVS, UMR 1572, 91191 Gif-sur-Yvette, France
[2]Laboratoire de Biogéochimie Isotopique, LBI, Bâtiment EGER, 78026 Thiverval-Grignon, France
[3]INRA, Centre INRA de Nancy, UMR1137 Ecologie et Ecophysiologie Forestières, 54280 Champenoux, France
[4]INRA, UR1263 EPHYSE, 33140 Villenave d'Ormon, France
[5]Dream CEFE-CNRS, 1919 route de Mende, 34293 Montpellier, France
[6]INRA, Grassland Ecosystem Research, 234 Avenue du Brézet, 63039 Clermont-Ferrand, France
[7]INRA-CSE, Domaine Saint-Paul, 84914 Avignon Cedex 9, France
[8]INRA, Environnement et Grandes Cultures, 78850 Thiverval-Grignon, France

[9] CNRM/GAME, Météo-France, CNRS, URA 1357, 42 avenue Coriolis, 31057 Toulouse Cedex 1, France

Received: 29 November 2010 – Accepted: 21 February 2011 – Published: 9 March 2011

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

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

BGD
8, 2467–2522, 2011

Abstract

We analyze how biases of meteorological drivers impact the calculation of ecosystem CO_2, water and energy fluxes by models. To do so, we drive the same ecosystem model by meteorology from gridded products and by "true" meteorology from local observation at eddy-covariance flux sites. The study is focused on six flux tower sites in France spanning across a 7–14 °C and 600–1040 mm yr^{-1} climate gradient, with forest, grassland and cropland ecosystems. We evaluate the results of the ORCHIDEE process-based model driven by four different meteorological models 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 between meteorological models and local tower meteorology. The 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 near the coast and influenced by sea-breeze, or located in altitude, the misfit of meteorological drivers from gridded dataproducts 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 modeled grid point and measured local meteorology going from 0.35 (REMO model) to 0.70 (SAFRAN model). The bias of meteorological models 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 than observed local meteorology drivers to ORCHIDEE is quantified for the six studied sites and 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 the largest on a daily time scale, for which it is as large as the model error. The forcing error incurring from using gridded meteorological model to drive vegetation models is therefore an important component of the uncertainty budget of regional CO_2, water and energy fluxes simulations, and should be taken into consideration in up-scaling studies.

BGD

8, 2467–2522, 2011

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 positive feedbacks with future climate change (Friedlingstein et al., 2003). Process oriented Terrestrial Biosphere Models (TBM) are useful tools to quantify and understand carbon fluxes and pools 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 Precipitation (NCEP) or by optimal data interpolation schemes (Mitchell et al., 2004), are commonly used to drive TBMs for regional and global applications. Weather is the main driver of variations in CO_2, H_2O 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, in controlling fluxes on 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 study 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 seasonal time scale, and of 40–60% on interannual time-scale, given different drivers. However, how bias of 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 (Sczcypta et al., 2011).

To tackle this problem, we use continuous measurements of CO_2, H_2O and heat fluxes made by 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 drivers biases can be justified because a high-resolution

BGD
8, 2467–2522, 2011

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 resolutions global weather analysis 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 half-hour. This allows to calculate explicitly the diurnal variation of ecosystem fluxes, which is not the norm in many biosphere models, and allows to be consistent 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 online together with CO_2, H_2O, and heat fluxes. Site observed meteorology will be considered in the study as the truth against which meteorological model products can be benchmarked. For applications limited to sites, local meteorology is obviously the best possible driver for TBMs. But for modeling regional carbon budgets, meteorological forcing is needed on a grid, hence modellers cannot avoid using imperfect model datasets. 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 questions addressed in this study are:

How different is meteorology at flux tower locations between local observation and gridded atmospheric model fields?

Is the uncertainty of modeled meteorology random or systematic?

What is the uncertainty of simulated ecosystem CO_2, water vapour and heat fluxes (here using the ORCHIDEE model) induced by errors in meteorological drivers at different time scales?

What is the sensitivity of ecosystem fluxes simulated by ORCHIDEE to each specific meteorological driver? How does the "forcing error" incurring from meteorological forcing compares with other model errors related to model structure and parameters value?

BGD
8, 2467–2522, 2011

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 multi-years, yet with focus on the growing season period. In the following, we describe the eddy-covariance data, the meteorological gridded data-products and the ORCHIDEE terrestrial biosphere model (Sect. 2). Then, we compare meteorological drivers from gridded data-products with local tower observation (Sect. 3). The performances of ORCHIDEE to simulate the variability of ecosystem fluxes are analyzed in Sect. 4, and the forcing error is estimated and compared to other model errors in Sect. 5.

2 Material and methods

2.1 Eddy-covariance data from six flux towers

The six sites 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 (Laqueille) and two intensive cropland sites, in the Paris region (Grignon) and in the south of France (Avignon), respectively. The sites location and climate space distribution over France are shown in Fig. 1. The entire dataset represents a total of 42 site *x* years (Table 1). Site equipment, data acquisition and processing (gap-filling) are described in Appendix A.

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

- 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

BGD
8, 2467–2522, 2011

round. The site is managed according to standard local management strategy and was thinned in 1991, 1996 and in 2004.

- Puéchabon (PUE) is a 70-year 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, and 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 grazed continuously with a mean stock of 0.6 livestock unit ha^{-1} per year and no fertilizer application.
- Avignon (AVI) is a long established agricultural site located in Southeastern France. Durum wheat, Peas and Durum wheat are the rotation 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 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 year 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

BGD
8, 2467–2522, 2011

period after GPP reaches its peak, spans from 1 July to 31 August for HES, LBR and LQE, and 1 May to 30 June for PUE, AVI and GRI.

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 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 half-hourly time step, and plant growth, mortality, soil carbon decomposition and phenology at daily time-step. Leaf onset is calculated as a function of growing degree-days and chilling days requirements, or 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

BGD
8, 2467–2522, 2011

of temperature-dependent maintenance respiration processes and GPP-dependent growth respiration processes. Litter and soil organic matter decomposition are calculated on daily time-step using first order kinetics 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 meteorological data. The fractional coverage of PFT at each site is prescribed according to site species data (Table 1). Soil type is set identically at all sites with a 20% sand, 45% loam and 35% clay content. Each simulation includes an equilibrium spin-up run followed by a transient run. In the spin-up run, observed meteorology for the period of observations are used cyclically to drive the model for 1500 years until equilibration of carbon (and water) pools is attained, with $< 0.05\%$ yearly increment. Biomass pools reach equilibrium within typically 200 years. 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 2008: available online via anonymous FTP to ftp.cmdl.noaa.gov, path: ccg/co2/GLOBALVIEW). Given this approach, the long-term modeled carbon balance (NEE) is by construction equal to zero on a multi-year basis, unlike observed at each flux tower. Therefore, ORCHIDEE NEE is always biased because all the sites have a net CO_2 uptake, of 300 g C m^{-2} $year^{-1}$ on average for the forest sites, and 200 g C m^{-2} $year^{-1}$ on average for the grass and crop sites (Table 1). At the cropland sites, where harvested biomass is exported away from ecosystem, 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 (Appendix D).

BGD
8, 2467–2522, 2011

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 last three drivers have no impact on the model output, and are thus discarded in the follow.

Observed meteorology (OBS) is recorded continuously on top of each tower, averaged every half-hour and quality checked (Aubinet et al., 2000). Missing values are due to instrument failure or violent weather conditions, with data-gaps of 1% for T_{air}, Q_{air}, rainfall and SW_{down} and of up to 12-days in duration at few sites. Because running ORCHIDEE requires continuous forcing, these gaps were filled with the SAFRAN model, using a linear regression between tower and SAFRAN daily data. 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 and minimum temperature at 05:00 LT (Campbell and Norman, 1998). 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 case of missing measurement, LW_{down} is gap-filled by the empirical relationship:

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

where cloudy (cloud cover) is the ratio of SW_{down} to maximum SW_{down} under clear sky condition, based on solar zenithal angle, Q_{air} the specific humidity ($kg\,kg^{-1}$), Stefans the Stefan-Boltzmann constant ($5.67 \times 10^{-8}\,W\,m^{-2}\,K^{-4}$) and T_{air} the air temperature in K. Tests during period with observations show that this Eq. (1) overestimates mean

BGD
8, 2467–2522, 2011

annual LW_{down} by 5–15 W m^{-2}. Comparison of ORCHIDEE fluxes driven by OBS or by modeled meteorology is focused on the period 2004–2007 during which gaps in OBS are minimal.

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 stations 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 dataset 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 dataset 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 B.

2.4 Model-data comparison statistics

Statistical criteria applied to the differences between OBS and modeled meteorology (Sect. 3), and to modeled and observed ecosystem fluxes (Sect. 4) are (1) the Mean Absolute bias Error (MAE), (2) the squared correlation coefficient (R^2) and (3) the Root Mean Squared Error (RMSE). The model-data comparison is performed on featured time scales: hourly, daily, monthly and annual (see Baldocchi et al., 2001; Stoy et al., 2005; Mahecha et al., 2007). We first aggregate time series of half-hourly into hourly,

BGD
8, 2467–2522, 2011

daily, monthly and annual scales, then new time series Y for each time scale is constructed using the following decomposition:

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

where X_i is the hourly ($i = 1$), daily ($i = 2$), monthly ($i = 3$) and annual ($i = 4$) averaged value of hourly time series X_1. Each statistical criteria is calculated for Y at each time scale.

3 Benchmarking gridded meteorology forcing against flux tower data

Figure 2 shows a comparison between observed (OBS) and modeled meteorology at HES and PUE from hourly to inter-annual time scales. For the other sites, the comparison is given in Appendix C. Due to the short length of records, some sites were excluded: GRI where the first year of observations is 2005 whereas EC-OPERA forcing data is only available till 2006. Figure 3 shows the R^2 of OBS vs. gridded meteorology, for 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 benchmark

The average diurnal cycle of T_{air} and SW_{down} is simulated by all gridded products well, with R^2 from 0.51 to 0.97 (Figs. 2 and C). The daytime value of T_{air} between 06:00 and 19:00 UMT appears however to be overestimated at the forest sites, but within 2 °C of local observation at the crop sites. This may reflect local evaporative cooling over forests (Zaitchik et al., 2006; Teuling et al., 2010). There is a positive bias of atmospheric T_{air} in gridded products at the upland LQE site, 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 models (Sczcypta et al., 2011). The pronounced diurnal cycle of Q_{air}

BGD
8, 2467–2522, 2011

simulated by REMO is found neither in OBS, 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 ($\approx 40\,W\,m^{-2}$) at LBR, PUE, AVI and GRI is underestimated by SAFRAN, EC-OPERA and ERA-I that give values of 22, 26 and $34\,W\,m^{-2}$, respectively, while it is overestimated by Eq. (1) 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 benchmark

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 the summer 2005 at LQE, AVI and GRI (Figs. 2 and C). The main result is that the synoptic-scale variability of daily T_{air}, Q_{air} and SW_{down} is well captured by all gridded dataproducts when compared to OBS. For daily variability of T_{air} during July–August 2003 or 2005, the mean R^2 of OBS and models is of 0.87 (from 0.79 in REMO to 0.94 in SAFRAN). 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 OBS 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 OBS. But the rain gauge data quality is poor during 2002 to 2006 (Loustau, personal communication, 2010). The mean summer 2003 rainfall is 71 mm in SAFRAN and 75 mm in OBS 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 mean R^2 values between models and OBS across

BGD
8, 2467–2522, 2011

the six sites and four years is 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 ($R^2 = 0.90$). The LQE site has the smallest R^2 (0.20), 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 OBS and models is lower for SW_{down} than for T_{air} and Q_{air}, and thus errors in SW_{down} will be concern in driving TBM models like ORCHIDEE by gridded products (see Sect. 5.4).

The LW_{down} daily variability is correctly represented by SAFRAN, EC-OPERA and ERA-I, with R^2 values going from 0.55 in SAFRAN to 0.78 in EC-OPERA across the sites at which LW_{down} measurements were collected during summer 2003, although REMO gives poor performances ($R^2 = 0.25$). Observed LW_{down} in summer 2003 (365 $W m^{-2}$) is slightly underestimated by SAFRAN (350 $W m^{-2}$), EC-OPERA (350 $W m^{-2}$) and ERA-I (348 $W m^{-2}$), but largely overestimated by REMO (430 $W m^{-2}$).

3.3 Seasonal benchmark

The mean seasonal cycle of T_{air}, Q_{air} and of SW_{down} is well captured by all gridded products (Figs. 2 and C), with a mean $R^2 = 0.98$. The MAE in the seasonal amplitude of T_{air} between models and OBS is 0.2 °C across sites and models. The seasonality of rainfall is correctly represented by gridded dataproducts at PUE and AVI where the Mediterranean summer is very dry, and most rainfall delivered in autumn ($R^2 = 0.60$ in REMO to 0.85 in SAFRAN). The agreement between observed and modeled seasonal rainfall at HES is better in SAFRAN than in other models. The seasonal cycle of LW_{down} in all data-products is in agreement with OBS, but this variable is overestimated by REMO (by 60 $W m^{-2}$) and slightly underestimated by the other models (5 $W m^{-2}$ in SAFRAN, and 15 $W m^{-2}$ in EC-OPERA, ERA-I). The seasonal phase of LW_{down} in

BGD
8, 2467–2522, 2011

REMO correlates well with observations ($R^2 = 0.86$), the positive bias of LW_{down} being correlated with the wet bias of this model.

3.4 Interannual benchmark

Considering HES, LBR and PUE that have records longer than 8 years, and excluding years with gaps in measurements, the average interannual R^2 values between annual modeled and observed meteorology are of 0.80, 0.40, 0.48, 0.38 and 0.24 for T_{air}, Q_{air}, rainfall, SW_{down} and LW_{down}, respectively. The main result is that the models are only able to reproduce correctly the interannual variablity of T_{air}, but the variability of other drivers is poorly captured by models. In particular, the interannual variability of SW_{down} and LW_{down} is not well reproduced, with an average R^2 of 0.38 for SW_{down} and 0.24 for LW_{down} across the 3 sites. The interannual variability of rainfall is faithfully reproduced at HES ($R^2 = 0.95$) and PUE ($R^2 = 0.96$) by SAFRAN. This gives higher confidence in SAFRAN meteorology to drive carbon flux on a year-to-year basis. At the LBR site, the interannual variability of rainfall is badly simulated, even with the best model SAFRAN ($R^2 = 0.47$, $n = 12$ years), which could be due to raingauge disfunctionning (Loustau, personal communication, 2010).

3.5 Summary of gridded data-products performance

3.5.1 Correlations between modeled and observed variability

In general, gridded data products compare better with local observation on monthly scale compared to hourly and daily scale (Fig. 3), with monthly R^2 of 0.88 to 1.00 for T_{air}, and of 0.61 to 0.98 for Q_{air}. For monthly rainfall, R^2 goes from 0.32 in REMO to 0.84 in SAFRAN. The monthly seasonal cycle of SW_{down} has high R^2 values across all data-products ($R^2 \sim 0.92$; 0.81 in REMO). The monthly cycle of LW_{down} is best simulated by EC-OPERA ($R^2 = 0.90$) and worst by REMO ($R^2 = 0.63$). The daily variability of T_{air}, Q_{air} and SW_{down} is reasonably well reproduced by SAFRAN, EC-OPERA and

BGD
8, 2467–2522, 2011

ERA-I, but the daily variability of rainfall and LW_{down} is not (Fig. 3). REMO shows poorer performances than other gridded dataproducts, and SAFRAN performs best. Not surprisingly, the diurnal cycle of SW_{down} is realistic in all products (mean $R^2 = 0.90$). The diurnal cycle of T_{air} is well captured by all data-products ($R^2 = 0.70$ in REMO; $R^2 = 0.90$ in the other three dataproducts); the diurnal cycle of Q_{air} is best reproduced by SAFRAN ($R^2 = 0.51$). By contrast, the diurnal cycle of LW_{down} and rainfall is not well captured by any of the data-products ($R^2 = 0.04$ in SAFRAN; 0.53 in EC-OPERA and ERA-I; R^2 for diurnal rainfall goes from 0 in ERA-I to 0.22 in SAFRAN). In summary, we found that time series of T_{air}, Q_{air} and SW_{down} from gridded data-products have high correlation with tower meteorology on hourly, daily and monthly time scales, but not on inter-annual scales (except T_{air}). The temporal variability of rainfall and LW_{down} is faithfully reproduced by gridded data-products on daily and monthly scale only, raising a caution flag for interannual TBMs simulations.

3.5.2 Bias of gridded data-products

Figure 4a shows that the MAE between gridded data products and tower observations. The mean bias of T_{air} and Q_{air} is 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 REMO larger hourly MAE is due to our conversion of daily values to hourly, thus not a bias of the model itself. For hourly rainfall, we note that none of the gridded data-product sperforms better for MAE than a "nul model" with even distribution of rainfall each hour during rainy days. The MAE of rainfall is maximum on daily time scale, going from a MAE = 1.8 mm day^{-1} in SAFRAN to MAE = 4.0 mm day^{-1} in REMO. By contrast, the MAE of rainfall remains moderate on longer, monthly and inter-annual scales (MAE < 1.0 mm day^{-1} across sites and models). The MAE of SW_{down} is large on hourly and daily scales (30–60 W m^{-2}), but improves on monthly and annual scales (10 W m^{-2}). The absolute bias of LW_{down} is minimum with EC-OPERA on hourly to monthly scales, and maximum with REMO on annual scales (50 W m^{-2}; other ataproducts ≈ 20 W m^{-2}).

BGD
8, 2467–2522, 2011

Figure 4b shows that MAE of daily to annual T_{air}, Q_{air} and SW_{down} is abnormally high at LQE compared to the other sites, irrespective of the gridded data-product compared with observations. The MAE of SW_{down} at LQE at hourly scale (80 $W m^{-2}$) is also larger than at any of the other sites (40 $W m^{-2}$). This is possibly due to a slope exposure of the LQE tower. Overall, the SAFRAN data-product has the best performance in terms of minimal MAE for T_{air} and rainfall, while EC-OPERA and ERA-I do well for SW_{down} and LW_{down}. REMO is worse than the other 3 models. As a general rule, we found that biases associated with gridded data products are smaller on monthly scale than on hourly to daily time scales. In a recent study comparing ERA-I and SAFRAN, Szczypta et al. (2010) 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 in 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.

3.5.3 Summary

We conclude from this comparison between gridded modeled and local meteorology that SAFRAN is the best forcing, both in terms of temporal variability (R^2) and absolute bias (MAE). SAFRAN is superior to other models for rainfall on daily and monthly scales. REMO has the lowest performance among the models in reproducing hourly to monthly variations. The quality of EC-OPERA and ERA-I forcing is in-between SAFRAN and REMO, yet with slightly lower skills than SAFRAN. The seasonal cycle of T_{air}, Q_{air} and SW_{down} is correctly reproduced by all the gridded data-products. The temporal variability of these variables is well predicted on daily time scale, except by REMO. The variability of rainfall on hourly scale is not reproduced, even for SAFRAN which contains synoptic station data. This shortcoming is however rather unimportant for CO_2 flux modeling. 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}. This raises a caution flag when Q_{air} is calculated by a weather generator to drive carbon

BGD
8, 2467–2522, 2011

flux simulations (Richardson and Wright, 1984; Krinner et al., 2005). On inter-annual time scale, T_{air} is the variable best reproduced by all gridded data-products, but SW_{down} and LW_{down} are the most problematic. The SAFRAN dataset has the highest skill for rainfall interannual variability, but the skill of SAFRAN is site dependent. SAFRAN does not show a bias of the same sign across the six flux tower sites, making a systematic correction difficult, and pointing out to local feedbacks of the vegetation. Higher than observed T_{air}, Q_{air}, rainfall and lower than observed SW_{down} values are consistently found across the gridded data-products on annual scale. Among the six sites, meteorology at the LQE grassland upland site is the most difficult to simulate by gridded products, given their coarser resolution.

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 OBS or with atmospheric analyses meteorology. The detailed description of the difference between modeled and measured carbon flux on four time scale at each site is given in Appendix D.

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

Figure 5 shows the correlation between modeled and observed fluxes for different time scales, and for different meteorological drivers. At first glance, the correlations are always rather low, independently of the meteorology used to drive the model. This suggests that model structural errors largely explain the small values of R^2. This also suggests that for site-level study, ORCHIDEE needs to be further calibrated. Generally, the R^2 values are higher for diurnal and monthly time scales compared to daily time scales (Fig. 5). Figure 5 also shows that on monthly scale, R^2 is higher for water fluxes (average $R^2 = 0.77$) than for CO_2 fluxes (average $R^2 = 0.61$). On average, forcing

BGD
8, 2467–2522, 2011

ORCHIDEE driven by OBS meteorology gives a higher R^2 than with any of the gridded data-products. Across all fluxes and time scales, R^2 is slightly higher with OBS (R^2 = 0.52) than with even the best modeled forcing SAFRAN (R^2 = 0.49). On average for daily scale, driving ORCHIDEE with OBS meteorology gives higher correlations than when using atmospheric analyzed meteorology, except for the LQE mountain grassland where using a modeled meteorology improves the value of daily R^2 over five fluxes from 0.16 to 0.28. This indicates error compensation in ORCHIDEE, where a biased forcing compensates for a structural bias.

4.2 Bias of modeled fluxes, function of driving meteorology

Figure 6a shows the MAE of modeled fluxes for ORCHIDEE driven by OBS and by each gridded data-product. Forcing ORCHIDEE with OBS meteorology compared to a gridded product delivers only a small reduction of MAE. The SAFRAN, EC-OPERA or ERA-I drivers result into more or less the 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 forcing are similar to those of GPP on diurnal and daily time scale. But on monthly scale, the MAE of NEE differs from the one of GPP, as it combines also the climate-driven misfit of the model to TER. We note that MAE of TER is on average smaller than for GPP on monthly scale (Fig. 6a).

One can also see from Fig. 6b that MAE differs largely between sites. In fact, the inter-sites 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.

BGD
8, 2467–2522, 2011

5 The effect of meteorology in the error budget of ORCHIDEE

5.1 Model error and forcing error

The effect 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:

$$\varepsilon_{\text{tot}} = \text{mean}\left[\text{RMSE}\left(F_{\text{sim}(i)} - F_{\text{obs}}\right)\right] \quad (3)$$

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:

$$\varepsilon_{\text{F}} = \text{mean}\left[\text{RMSE}\left(F_{\text{sim}(i)} - F_{\text{sim(OBS)}}\right)\right] \quad (4)$$

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

$$\varepsilon_{\text{mod}} = \text{RMSE}\left(F_{\text{sim(OBS)}} - F_{\text{obs}}\right) \quad (5)$$

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

$$\varepsilon_{\text{tot}}^2 = \varepsilon_{\text{F}}^2 + \varepsilon_{\text{mod}}^2 + 2 \times \text{cov}(\varepsilon_F \varepsilon_{\text{mod}}) \quad (6)$$

We are interested here in cov, the covariance term that denotes correlated errors between the model structure 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

BGD
8, 2467–2522, 2011

the simulated flux closer to the data. This may happen for instance if a too high NPP in summer implied by a bias in meteorology creates too many assimilates, which in turn increase litter respiration in the fall, and compensate 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. (6) and analyzed below. In this section, TER and NEE are modeled using optimized TER (see Appendix D).

5.2 Comparing model and forcing errors for different fluxes

Figure 7 provides the unbiased error distribution for each flux across the six sites. Unbiased error is defined from Eqs. (3–6) using unbiased RMSE. Unbiased RMSE is obtained from the analyzed time series based on Eq. (2) for hourly, daily and monthly scale, and the centred annual mean time series for annual scale. One can see that for GPP and TER, the total error ε_{tot} (blue) is largely explained by model structural error ε_{mod} (red). For GPP, the share of the structural error in the total error is 1.01, 0.80, 1.02 and 0.64 on hourly, daily, monthly and annual time scale, respectively. For TER, the forcing error ε_F (orange) on hourly scale is negligible, because soil temperature and soil humidity that control soil respiration in ORCHIDEE exhibit no diurnal variability. By contrast, on daily to annual time scale, the forcing error takes a significant share of the total uncertainty budget of TER, as shown by the ratio $\varepsilon_F/\varepsilon_{tot} = 0.63$, 0.39 and 0.35, respectively. For NEE, the share of forcing error to total error is of 0.57, 0.82, 0.64 and 0.37 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 $\varepsilon_F/\varepsilon_{tot}$ on daily and annual time scale is 0.87 and 1.03 for LH, 1.03 and 1.10 for SH, respectively. This forcing error is due primarily to biases in SW_{down} between the different forcings (see Sect. 5.4, Fig. 9).

The most interesting result is that the forcing error is not negligible, compared to other model errors. This comes a bit as a surprise because meteorology is generally

BGD
8, 2467–2522, 2011

assumed in vegetation modelling to be well enough known not to create a misfit in modeled fluxes.

We note also that $\varepsilon_{\mathrm{mod}}$ and ε_F decrease in absolute value with increasing temporal averaging scale, from typical errors values of 1.8 and 3.0 $\mathrm{g\,C\,m^{-2}\,year^{-1}}$ on diurnal scale for ε_F and $\varepsilon_{\mathrm{mod}}$, down to 0.7 and 0.5 $\mathrm{g\,C\,m^{-2}\,year^{-1}}$ on annual scale for GPP errors. For LH errors, the estimates of $\varepsilon_{\mathrm{mod}}$ decrease from about 20 $\mathrm{W\,m^{-2}}$ on hourly scale down to 6.0 $\mathrm{W\,m^{-2}}$ on annual scale, while ε_F decreases from 15 $\mathrm{W\,m^{-2}}$ to 10 $\mathrm{W\,m^{-2}}$. 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 structural model errors are partly compensated by biases of meteorological forcing. This result may be due to similar biases among different forcings compared to OBS, which act to shift the simulated TER and GPP closer to the observed fluxes. One can see in Fig. 7 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 same magnitude on TER. The forcing errors of NEE are as large as those of GPP on diurnal to daily scales. But on 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 same magnitude for NEE and GPP, because the climate sensitivity of TER is probably as large as that of GPP on long time scale, 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 structural errors. On interannual time scales, the forcing error is an overlooked source of poor model performance in tackling the simulation of interannual GPP and

BGD
8, 2467–2522, 2011

TER. For inter-annual flux variability, the ratio ($\varepsilon_F/\varepsilon_{\mathrm{mod}}$) averaged across 6 sites is 0.80, compared to 0.55 for monthly time scale, 0.86 for daily scale and 0.63 for hourly scale.

5.3 Comparing model and forcing errors across sites

Figure 8 provides a comparison of the forcing and model error at each site. Typical values of the ratio $\varepsilon_F/\varepsilon_{\mathrm{mod}}$ range from 0.54 at HES to 0.92 at LQE. The ratio $\varepsilon_F/\varepsilon_{\mathrm{mod}}$ for SH is larger than or close to 1, which likely reflects biases in $\mathrm{LW_{down}}$ forcing and a high sensitivity of SH to that driver.

Generally, the cropland and grassland sites have a larger forcing error than the forests for $\mathrm{CO_2}$ fluxes, with $\varepsilon_F = 716$, 286 and 644 $\mathrm{g\,C\,m^{-2}\,year^{-1}}$ at crop sites against 390, 155 and 291 $\mathrm{g\,C\,m^{-2}\,year^{-1}}$ at forests for GPP, TER and NEE, respectively. The LQE grassland in altitude has obviously the largest ε_F at all time scales ($\varepsilon_F/\varepsilon_{\mathrm{mod}} = 0.90$), because the forcing used to drive ORCHIDEE at this site have different spatial resolution over a complex and heterogeneous terrain, thus giving a larger spread of simulated fluxes.

The error identification per time scale can thus be used as a tool to detect sites that have a more "difficult" meteorology, impacting the simulation of ecosystem fluxes.

5.4 Separate contribution of each meteorological variables 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 the ith driver, $\varepsilon_F(i)$, are calculated by Eq. (4) with ORCHIDEE being run with OBS forcing, except for the ith driver that is taken from a gridded dataproduct (here SAFRAN). We also take the simulation where all the five meteorological variables are taken from SAFRAN, defining a total forcing error ε_F (Eq. 4). Figure 9 provides a comparison of the contribution from each meteorological driver to the total forcing error. The relative

BGD
8, 2467–2522, 2011

contribution $C(i)$ of the ith driver to the total forcing error is defined by:

$$C(i) = \varepsilon_F(i)/\varepsilon_F \times 100 \quad (7)$$

A contribution that exceeds 100%, indicates compensations between errors induced by different SAFRAN variables. It is seen on the two diagonals of Fig. 9 (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 CO_2 and water fluxes. On 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 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 inter-annual time scale, the relative contribution of each meteorological driver to ε_F becomes comparable. This adds to the difficulty of reducing errors in the simulation of inter-annual 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 OBS 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. 9 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 for 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:

$$\varepsilon_F(ij)^2 = \varepsilon_F(i)^2 + \varepsilon_F(j)^2 + 2 \times \mathrm{cov}(ij) \quad (8)$$

BGD
8, 2467–2522, 2011

Individual forcing errors $\varepsilon_F(i)$, $\varepsilon_F(j)$ and $\varepsilon_F(ij)$ are calculated by Eq. (4). 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 cov(ij) between drivers i and j is defined by:

$$C(ij) = \mathrm{cov}(ij)/\varepsilon_F^2 \times 100 \quad (9)$$

The $C(ij)$ contributions are summarized by the upper and lower triangles of Fig. 9 (upper one calculated with annual fluxes and lower one with summer fluxes). Most of the covariance 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 dataproduct overestimating SW_{down} should also overestimate T_{air} and underestimate Q_{air}. Secondly, the simulated ORCHIDEE fluxes in response to variation in drivers is likely to be a concave curve, implying that the effects of two biased variables is lower than the sum of the bias in each variable. Figure 9 shows that for inter-annual variability, however, the curvature of ORCHIDEE fluxes could be convex or concave, 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 Conclusions

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

– How different is meteorology at flux tower sites between local observation and atmospheric analyses fields sampled at the same site?

We found large differences among atmospheric analyses, and between weather-models and site meteorology. These differences are particularly important on interannual time scales, and particularly large for radiation parameters LW_{down} and SW_{down}.

BGD
8, 2467–2522, 2011

– Is the uncertainty of modeled meteorology random or systematic?

We found significant bias in atmospheric analyses compared to site-observed meteorology. The best forcing is SAFRAN that has also the highest spatial resolution, and is produced by optimal data interpolation instead of by a numerical atmospheric analyses. Biases are larger for annual time scales than daily or monthly time scales, and are also site-dependent. Sites that are difficult to reproduce by atmospheric analyses are mountain sites (Laqueuille) because of coarsely-resolved topography in atmospheric analyses.

– What is the uncertainty of modeled ecosystem fluxes – here using ORCHIDEE – induced by differences between meteorological drivers at different time scales? and What is the error induced meteorological forcing compared to error related to model structural and parameters value?

Using atmospheric analyses instead of local observations to drive ORCHIDEE, we developed a framework to estimate the forcing error, reflecting uncertainties in drivers, separately from the model structural error, reflecting uncertain parameter values and parameterizations or initial conditions. The forcing error is on average smaller than, but still comparable to the model structural error. The relative contribution of the forcing error increases with increasing time scale, indicating that modeling interannual flux variability using atmospheric analyses products as input to ORCHIDEE, is limited as much by meteorology than by imperfect model parameterizations. This result is likely to be generalized for other ecosystem models than ORCHIDEE. On diurnal and daily timescales, we found that there is negative error covariance between forcing and structural errors, indicating error compensation mechanisms in ORCHIDEE.

– What is the sensitivity of ORCHIDEE modeled ecosystem fluxes to each meteorological driving variable?

This question was addressed through factorial experiments in which one meteorological driver at a time is taken from a weather forecast model, the others being from

BGD
8, 2467–2522, 2011

observed meteorology. We found that SW_{down} had dominant contributions to the forcing errors on daily scales for most fluxes. In general, forcing errors associated to pairs of meteorological variables are found to be negatively-correlated, that is partly compensate for each other in the resulting flux. On inter-annual time scales, 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 short, uncertainty in meteorology is a limitation to the accurate modeling of flux variability. Maybe when calculating regional budgets, there are spatial error compensations in meteorological forcing that will make the situation better, and diminish the contribution of forcing errors. But this source of uncertainty, often overlooked or considered small, is in fact significant when trying to use data-driven or process-model to upscale fluxes, in particular on annual to inter-annual scales.

Appendix A

Eddy covariance equipment and data processing

All sites are equipped to measure NEE, SH and LH by covariance technique at every 30 min time step (Baldocchi et al., 2001). Meteorological data were continuously measured and averaged every half-hour. 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 method (Reichstein et al., 2005), for which uncertainties were quantified in gap filling by Moffat et al. (2007). The data were downloaded from the Carboeurope-IP database (http://gaia.agraria.unitus.it/database) for CO_2 fluxes, SH, LH and all meteological parameters except longwave radiation which were complementarily provided by each site manager. Level 4 data used for the study.

BGD
8, 2467–2522, 2011

Appendix B

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 with a hour time step using ground data observations and meteorological simulations from the French weather service. The spatial resolution is 8 km × 8 km. Validation of the SAFRAN product by Quintana-Segui et al. (2008) indicated a general good agreement 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 observation 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 hourly time step. In addition, LW_{down} was found to be underestimated at these two sites by SAFRAN by 8 to 32 $W\,m^{-2}$.

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 km × 40 km horizontal resolution. Temporal resolution of the dataset is 6-h (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 operational system can be found at www.ecmwf.int/products/data/operatioanl_system.

BGD
8, 2467–2522, 2011

ERA-I is a global consistent reanalysis of the meteorological fields with the same version of ECMWF forecasting model (T255L60), starting in 1989 and continuing in real time. Spatial resolution is 79 km by 79 km. Temporal resolution is 6-h, like ECOPERA. 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 refers to a specific simulation by the regional climate model REMO (REgionalMOdel, Jacob et Podzun, 1997) over Europe, forced by 6-h NCEP reanalysis (Kalnay et al., 1996) at lateral boundaries from 1948 to 2007. The REMO physics is 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 scales 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 et al., 2001). The atmospheric fields are archived on daily step at 25 km by 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 to reproduce precipitation and radiation, with a dry bias in Mediterranean regions.

Appendix C

Comparison between observed (OBS) and modeled meteorology at LBR, LQE, AVI and GRI from hourly to inter-annual time scales

See figure section.

BGD
8, 2467–2522, 2011

Appendix D

Comparison between modeled and measured carbon flux

We study for each flux, the effect of forcing ORCHIDEE either with OBS or with atmospheric analyses meteorology.

D1 GPP

Figure D1 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 OBS 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 HES, LBR and at PUE forest sites (Fig. D1). At sites with long-enough observation period, a negative GPP anomaly in 2003 compared to other years is modeled, in agreement with the observation (Fig. D1; 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

BGD
8, 2467–2522, 2011

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.

D2 TER

At all the sites, the annual mean simulated TER is above the observed value (Fig. D2). This is expected from the equilibrium assumption of the ORCHIDEE simulations, where soil carbon pools are at their maximum value, for the given climate. This overestimate of soil C pools and mean TER also has 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 ($RA_{M\,model}$) by multiplication by a "biomass disquilibrium factor" defined as the ratio of measured to simulated average total biomass ($\langle B_{obs}\rangle/\langle < B_{model}\rangle$), and assume that GPP-dependent growth respiration ($RA_{G\,model}$) is perfectly simulated,

Step 2: Estimate the mean average heterotrophic respiration (RH_{obs}) from the difference between observed TER and the sum $RA_{M\,model\,new} + R_{G\,model}$, 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:

$$RA_{M\,model\,opt} = RA_{M\,model}\,\langle B_{obs}\rangle/\langle B_{model}\rangle \quad \text{(D1)}$$

$$RH_{model\,opt} = RH_{model}\left\langle TER_{obs} - RA_{M\,model\,opt} - RA_{G\,model}\right\rangle/\langle RH_{model}\rangle \quad \text{(D2)}$$

$$TER_{model\,opt} = RH_{model\,opt} + RA_{G\,model} + RA_{M\,model\,opt} \quad \text{(D3)}$$

$$NEE_{model\,opt} = TER_{model\,opt} - GPP_{model} \quad \text{(D4)}$$

BGD
8, 2467–2522, 2011

We apply this optimization procedure to all sites except PUE. The observed total biomass provided by each site manager is of 7000, 7850, 300, 400 and 225 g C m^{-2} $year^{-1}$ at HES, LBR, LQE, AVI and GRI, respectively. The biomass disequilibrium factors are respectively of 0.35, 0.55, 0.45, 0.58 and 0.27, and the soil disequilibrium factors of 0.60, 0.77, 0.85, 0.67 and 0.30 at these five sites. In other words, heterotrophic respiration are overestimated by ORCHIDEE from 18% at LQE to 220% at GRI.

This optimization procedure is not applied to Mediterranean forest site PUE because the overestimated TER at this site is caused by both discrepancies in carbon allocation between root and aboveground reservoir and the equilibrium assumption: calculation from the above procedure would give negative $RH_{model\,opt}$, 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 about 67% at PUE.

D3 NEE

NEE is optimized according to Eq. (D4). Thus the problem of "disequilibrium" is overcomed, however, we should bear in mind that the optimized NEE tends to be systematically underestimated (Fig. D3) due to the overestimated GPP by ORCHIDEE (Fig. D1).

At HES, the model performs best, both on daily and seasonal scale. 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 reduced CO_2 uptake is well captured by ORCHIDEE although NEE still keeps as 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 overstimated at HES, PUE and LQE, but underestimated at AVI. At the PUE Mediterranean oak forest, ORCHIDEE overestimates the seasonal uptake of CO_2 throught the year. At the temperate forest HES,

BGD
8, 2467–2522, 2011

CO_2 uptake is overestimated during the growing season forward between May and November. The seasonal phase of NEE is well represented at the forest sites HES and LBR, with respective $R^2 = 0.94$ and 0.61, but not at PUE ($R^2 = 0.26$). Modelling inter-annual NEE variability is not satisfactory in ORCHIDEE. The best interannual R^2 is 0.78 ($n = 7$ years) at PUE when the model is driven by OBS meteorology. The worst R^2 is 0.16 ($n = 10$ years) at HES. This could partly explained by the influence of thinning and the difficulties to reproduce the impact of the large fluctuations of the relative extractable water (REW) soil content (Granier et al., 2007).

D4 Latent and sensible heat flux

Comparison between observed and modeled latent heat flux (LH) is given in Fig. D4. On diurnal time scale, there is an overestimation of LH by ORCHIDEE at HES and PUE. This positive bias of LH is logically reflected on 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 OBS 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 overrestimates LH during the winter growing season from October to March, which parallels the overestimated GPP seen in Fig. D1, indicating 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 (Fig. D1) during the dry summer. 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 seasonal time scale seems to be driven by the bias on diurnal scale. The modeling assumption of equilibrium for carbon fluxes and pools, critical to explain the NEE model-data misfit

BGD
8, 2467–2522, 2011

on seasonal scale seems has here a negligible impact on the LH misfit. On seasonal time scale, LH is predominantly controlled by soil moisture availability, atmospheric dryness and leaf area index, which are rather independent on the slow C pool values that set up the value of TER and NEE (Carvalais et al., 2010). On interannual time scale, the variability of LH is not captured by ORCHIDEE. The best interannual R^2 is 0.48 ($n = 7$ years) at PUE when the model is driven by OBS meteorology. The worst R^2 is 0.11 ($n = 7$ years) at LBR. On average at the 3 forest sites (more than 5 years observation), the interannual variability of LH is higher in the eddy-flux observations (CV = 0.18) than in the model (CV = 0.07).

Figure D5 shows that the mean annual sensible heat flux (SH) is overestimated by ORCHIDEE independent of the time scale considered. The REMO forcing tends to produces the largest SH flux. SAFRAN, EC-OPERA and ERA-I forcings produce slightly larger SH flux than obeservation. SH flux is obviously overetimated by OBS forcing, in particular at HES, PUE before 2005 and LQE. This discrepacy between modeling and 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 OBS. LW_{down} in OBS 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 meterological 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. D5. 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 years observation), the interannual variability of SH is higher in the eddy-flux observations (CV = 0.44) than in the model (CV = 0.16).

BGD
8, 2467–2522, 2011

Acknowledgements. This study was conducted within the framework of the CarboFrance Project “Impact des extrêmes climatiques sur les flux de carbone” (GICC, French project).

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

References

Aubinet, M., Grelle, A., Ibrom, A., Rannik, S., Moncrie?, J., Foken, T., Kowalski, A., Mar tin, P., Berbigier, P., Bernhofer, C., Clement, R., Elbers, J., Granier, A., Grünwald, 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., Fuentes, J., Goldstein, A., Katul, G., Law, B. E., Lee, Z., Malhi, Y., Meyers, T., Munger, W. J., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K. B., and Wofsy, S.: FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities, B. Am. 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, Progr. Photosynth. Res., 4(1), 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, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, 2009.

Botta, A., Viovy, N., Ciais, P., Friedlingstein, P., and Monfray, P.: A global prognostic scheme of leaf onset using satellite data, Glob. Change Biol., 6(7), 709–725, doi:10.1046/j.1365-2486.2000.00362.x, 2000.

BGD

8, 2467–2522, 2011

Campbell, G. S. and Norman, J. M.: An Introduction to Environmental Biophysics, Springer-Verlag, New York, 1998.
Carvalhais, 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, Glob. Change Biol., 16(10), 2813–2829, 2010.
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., Ogee, 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(7058), 529–533, doi:10.1038/nature03972, 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.
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(3–4), 105–119, doi:10.1016/j.agrformet.2005.01.002, 2005.
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., 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.
Durand, Y., Brun, E., Mérindol, L., Guyomarch, G., Lesaffre, B., and Martin, E.: A meteorological estimation of relevant parameters for snow models, Annals of Glaciology, 18, 65–71, 1993.
Farquhar, G. D., Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO_2 assimilation in leaves of C3 species, Planta, 149(1), 78–90, doi:10.1007/BF00386231, 1980.
Feser, F., Weisse, R., and von Storch, H.: Multi-decadal atmospheric modeling for Europe

BGD
8, 2467–2522, 2011

yields multi-purpose data, EOS Trans., 82, 305, 2001.
Friedlingstein, P., Joel, G., Field, C. B., and Fung, I. Y.: Toward an allocation scheme for global terrestrial carbon models, Glob. Change Biol., 5(7), 755–770, doi:10.1046/j.1365-2486.1999.00269.x, 1999.
Friedlingstein, P., Dufrene, J., Cox, P. M., and Rayner, P.: How positive is the feedback between climate change and the carbon cycle?, Tellus B, 55(2), 692–700, doi:10.1034/j.1600-0889.2003.01461.x, 2003.
Granier, A., Reichstein, M., Bréda, N., Janssens, I. A., Falge, E., Ciais, P., Grünwald, T., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Facini, O., Grassi, G., Heinesch, B., Ilvesniemi, H., Keronen, P., Knohl, A., Köstner, B., Lagergren, F., Lindroth, A., Longdoz, B., Loustau, D., Mateus, J., Montagnani, L., Nys, C., Moors, E. J., Papale, D., Peiffer, M., Pilegaard, K., Pita, G., Pumpanen, J., Rambal, S., Rebmann, C., Rodrigues, A., Seufert, G., Tenhunen, J., Vesala, T., and Wang, Q.: Evidence for soil water control on carbon and water dynamics in European forests during the extremely dry year: 2003, Agric Forest Meteorol., 143(1–2), 123-145, 2007.
Jacob, D. and Podzun, R.: Sensitivity studies with the regional climate model REMO, Meteorol. Atmos. Phys., 63(1), 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(10), 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., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, B. Am. Meteorol. Soc., 77(3), 471, 437, 1996.
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 CO_2 and water fluxes through combined in situ measurements and ecosystem modelling, Biogeosciences, 6, 1423–1444, doi:10.5194/bg-6-1423-2009, 2009.

BGD
8, 2467–2522, 2011

Knorr, W., Prentice, I. C., House, J. I., and Holland, E. A.: Long-term sensitivity of soil carbon turnover to warming, Nature, 433(7023), 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. Wea. 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:10.1029/2003GB002199, 2005.
Loiseau, P., Vuichard, N., Ceschia, E., Balesdent, J., Chevallier, T., Arrouays, D., and Soussana, J. F.: Carbon cycling and sequestration opportunities in temperate grasslands, Soil Use Manage., 20(2), 219–230, 2004.
Loubet, B., Laville, P., Lehuger, S., Larmanou, E., Flechard, C., Mascher, N., Génermont, S., Roche, R., Ferrara, R. M., Stella, P., Personne, E., Durand, B., Decuq, C., Flura, D., Masson, S., Fanucci, O., Rampon, J. N., Siemens, J., Kindler, R., Schrumpf, M., Gabriele, B., and Cellier, P.: Carbon, nitrogen and Greenhouse gases budgets over a four years crop rotation in northern France, Plant Soil, doi:10.1007/s11104-011-0751-9, 2011.
Longdoz, B., Gross, P., and Granier, A.: Multiple quality tests for analysing CO_2 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.
Medlyn, B. E., Dreyer, E., Ellsworth, D., Forstreuter, M., Harley, P. C., Kirschbaum, M. U. F., Le Roux, X., Montpied, P., Strassemeyer, J., Walcroft, A., Wang, K., and Loustau, D.: Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data, Plant Cell Environ, 25(9), 1167–1179, 2002.
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(1–4), 381–413, doi:10.1016/0378-1127(90)90150-A, 1990.
Mitchell, S., Beven, K., and Freer, J.: Multiple sources of predictive uncertainty in modeled estimates of net ecosystem CO_2 exchange, Ecol. Model., 220(23), 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., Gove, J. H., Heimann, M.,

BGD
8, 2467–2522, 2011

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(3–4), 209–232, doi:10.1016/j.agrformet.2007.08.011, 2007.

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, Irrig. Drain. Syst., 19(3), 377–412, doi:10.1007/s10795-005-8143-z, 2005.

Parton, W., Stewart, J., and Cole, C.: Dynamics of C, N, P and S in grassland soils: a model, Biogeochemistry, 5(1), 109–131, doi:10.1007/BF02180320, 1988.

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.

Peylin, P., Bousquet, P., Quéré, C. L., Sitch, S., Friedlingstein, P., McKinley, G., Gruber, N., Rayner, P., and Ciais, P.: Multiple constraints on regional CO_2 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(3), 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., Kheshgi, H. S., Le Quéré, C., Scholes, R. J. D., and Wallace, W. R.: The carbon cycle and atmospheric carbon dioxide, Chapt. 3, in: Climate Change 2001, the Scientific Basis, edited by: IPCC, Cambridge University Press, Cambrigde, 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, doi:10.1175/2007JAMC1636.1, 2008.

Rambal, S., Ourcival, J., Joffre, R., Mouillot, F., Nouvellon, Y., Reichstein, M., and Rocheteau, A.: Drought controls over conductance and assimilation of a Mediterranean evergreen ecosystem: scaling from leaf to canopy, Glob. Change Biol., 9(12), 1813–1824, doi:10.1111/j.1365-2486.2003.00687.x, 2003.

Reichstein, M., Tenhunen, J., Roupsard, O., Ourcival, J., Rambal, S., Miglietta, F., Peressotti, A., Pecchiari, M., Tirone, G., and Valentini, R.: Inverse modeling of seasonal drought effects on

BGD
8, 2467–2522, 2011

canopy CO_2/H_2O exchange in three Mediterranean ecosystems, J. Geophys. Res.-Atmos., 108, 4726, 2003.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, 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, Glob. Change Biol., 11(9), 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.

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, Glob. Change Biol., 13(4), 788–803, doi:10.1111/j.1365-2486.2007.01330.x, 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, J. and Venevsky, S.: Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems, Nature, 414(6860), 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 al.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Glob. Change Biol., 9(2), 161–185, doi:10.1046/j.1365-2486.2003.00569.x, 2003.

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(7), 887–902, doi:10.1093/treephys/25.7.887, 2005.

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, 2011, http://www.hydrol-earth-syst-sci.net/15/647/2011/.

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., Grnwald, T., Klumpp, K.,Montagnani, L., Moureaux, C., Sottocornola, M., and Wohlfahrt, G.: Contrast-

BGD
8, 2467–2522, 2011

ing response of European forest and grassland energy exchange to heatwaves, Nat. Geosci., 3(10), 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(1–4), 185–222, doi:10.1016/S0168-1923(02)00108-9, 2002.

Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., 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., Holm, E., Hoskins, B. J., Isaksen, L., 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. R. Meteorol. Soc., 131(612), 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, 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:10.1029/2004GB002395, 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(6), 743–769, doi:10.1002/joc.1280, 2006.

BGD
8, 2467–2522, 2011

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

Site name	Hesse (HES)	Le Bray (LBR)	Puechabon (PUE)	La queuille (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 year)	Maritime pine (40 year)	Mediterranean green oak (70 year)	Extensively grazed grassland	Rotation, wheat-peas-sorghum	Rotation maize-wheat-barley
Mean annual biomass ($g\,C\,m^{-2}$)	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	20004–2007	2005–2007
Annual NEE ($g\,C\,m^{-2}$)	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)	Olioso et al. (2005)	Loubet et al. (2002)

BGD
8, 2467–2522, 2011

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)

BGD
8, 2467–2522, 2011

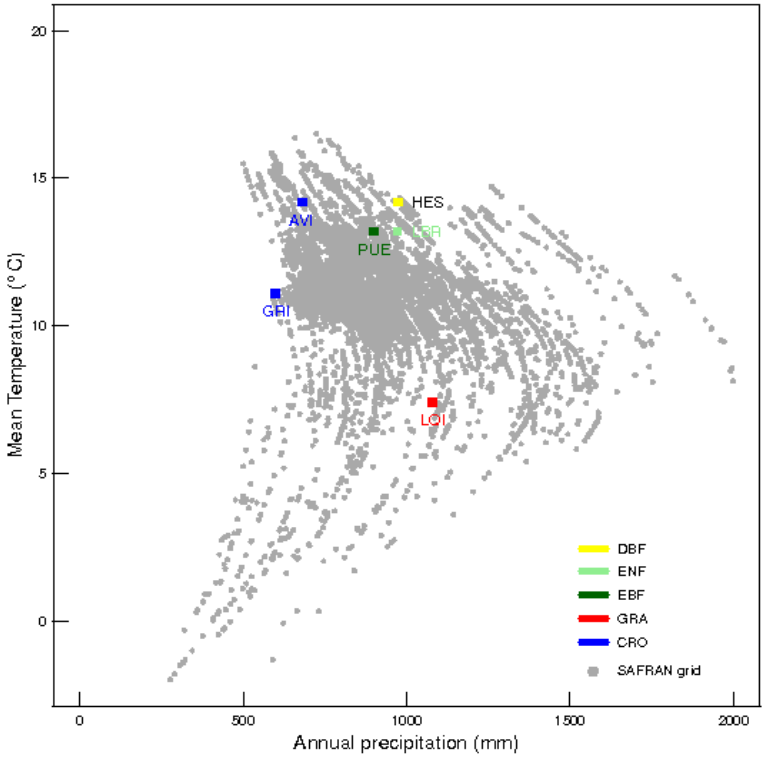


Fig. 1. The ecosystem types (DBF = deciduous broadleaves forest, EBF = evergreen broadleaves forest, ENF = evergreen needle leafs forest, CRO = cropland, GRA = grassland) of six selected flux sites over France in the climatic space: Mean temperature versus annual precipitation, which are calculated as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

BGD
8, 2467–2522, 2011

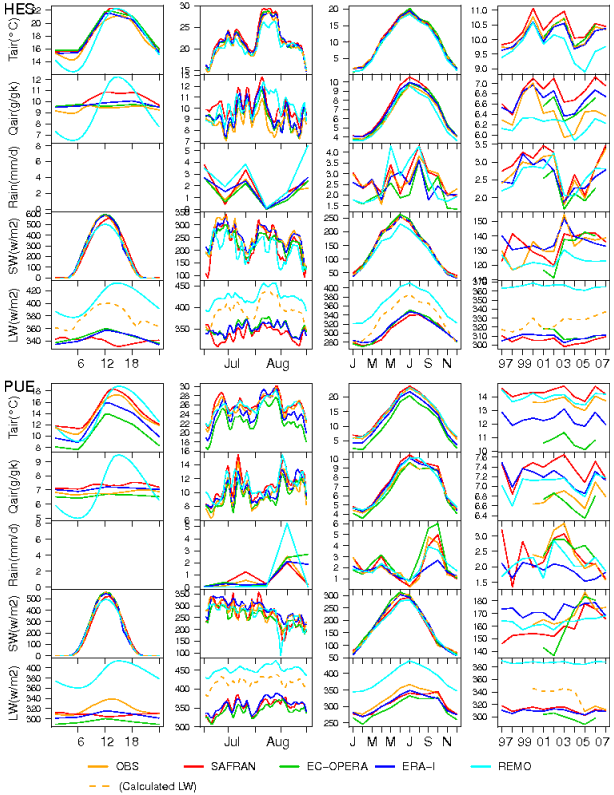


Fig. 2. Meterological drivers in in situ and in gridded data sets at six HES and PUE. First column: hourly mean diurnal cycle over peak growing season (PGS). Second column: daily mean with a running mean of 3 days for July–August of 2003 at HES and PUE. Rainfall is calulated as 5-day aggregated values; third column: monthy 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 but in dash lines. See text Sect. 2.1 for the definiton of PGS.

BGD
8, 2467–2522, 2011

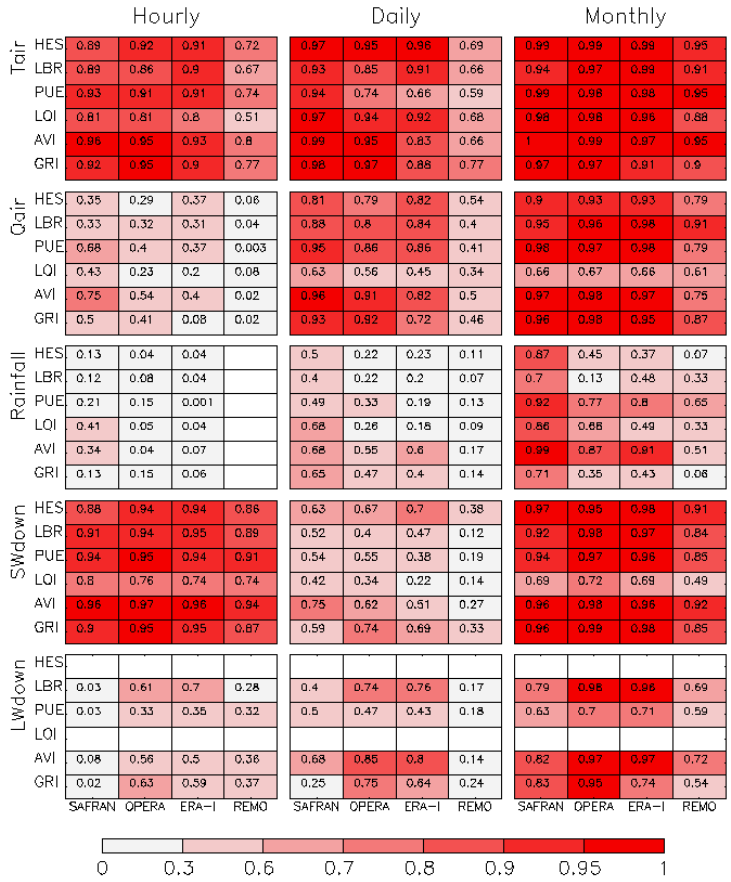


Fig. 3. Squared correlation (R^2) between meterological gridded data and in situ data over 2004 to 2007 except for EC-OPERA which covers 2004–2006. Panels of R^2 from left to right are for hourly, daily and monthly time scale, respectively. Time series to calculate R^2 correspond to growing season (GS). See text Sect. 2.1 for the definiton of GS.

BGD
8, 2467–2522, 2011

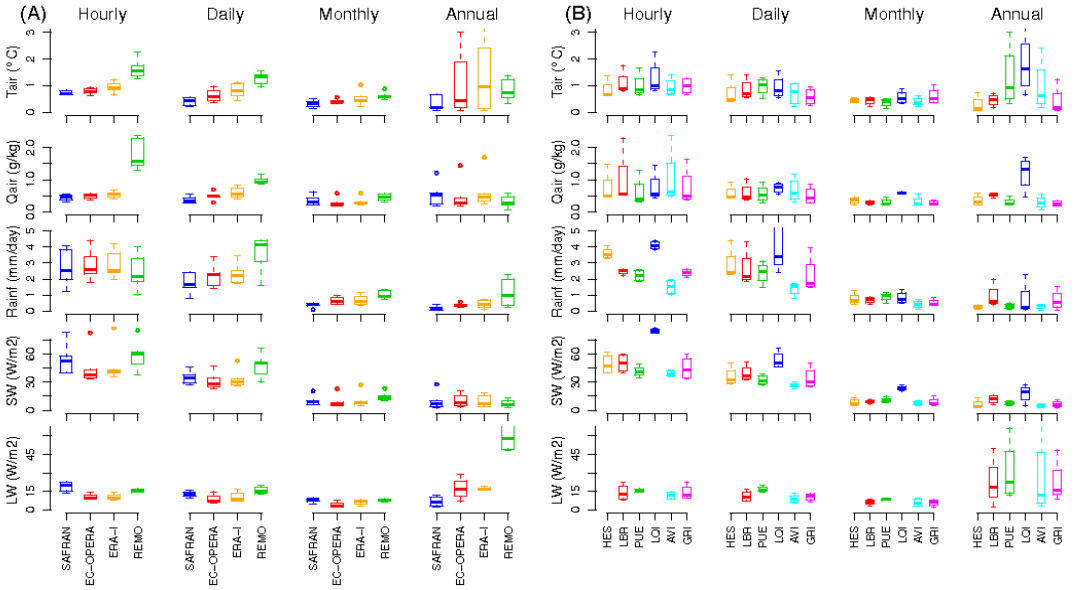


Fig. 4. Mean absolute error (MAE) between meterological gridded data and in situ data over 2004 to 2007 except for EC-OPERA which covers only 2004–2006. Left, middle and right column are for hourly, daily scale and monthly scale. Time series to calculate MAE correspond to growing season.

BGD
8, 2467–2522, 2011

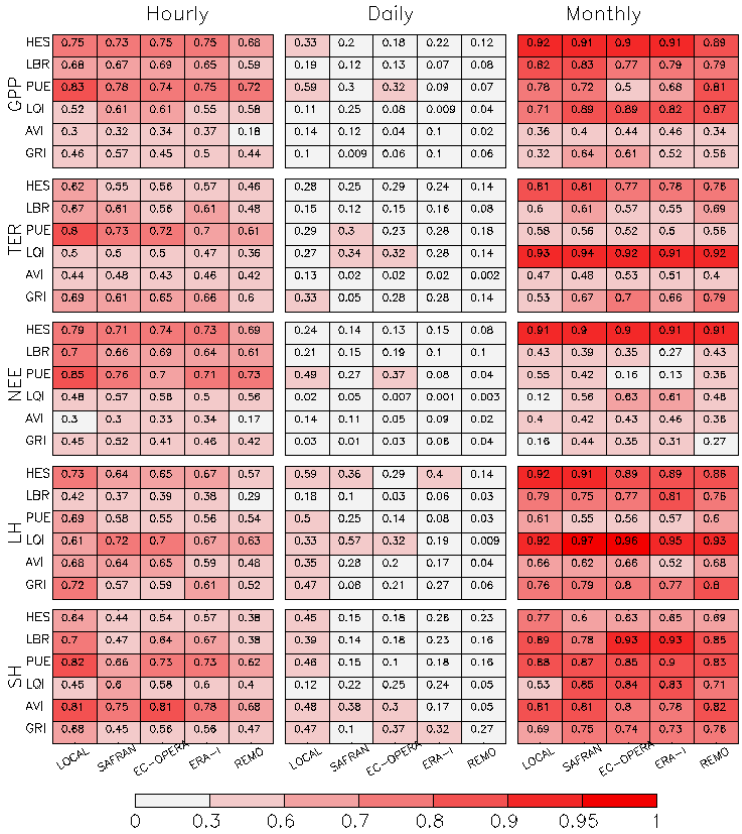


Fig. 5. Squared correlation (R^2) between simulated and measured flux data and in situ data over 2004 to 2007 except for EC-OPERA which covers 2004–2006. Panels of R^2 from left to right are for hourly, daily and monthly time scale, respectively. Time series to calculate R^2 correspond to growing season. For crop sites, we only take the years of winter-wheat-growing, that is, 2004 and 2006 at AVI, 2006–2007 at GRI.

BGD
8, 2467–2522, 2011

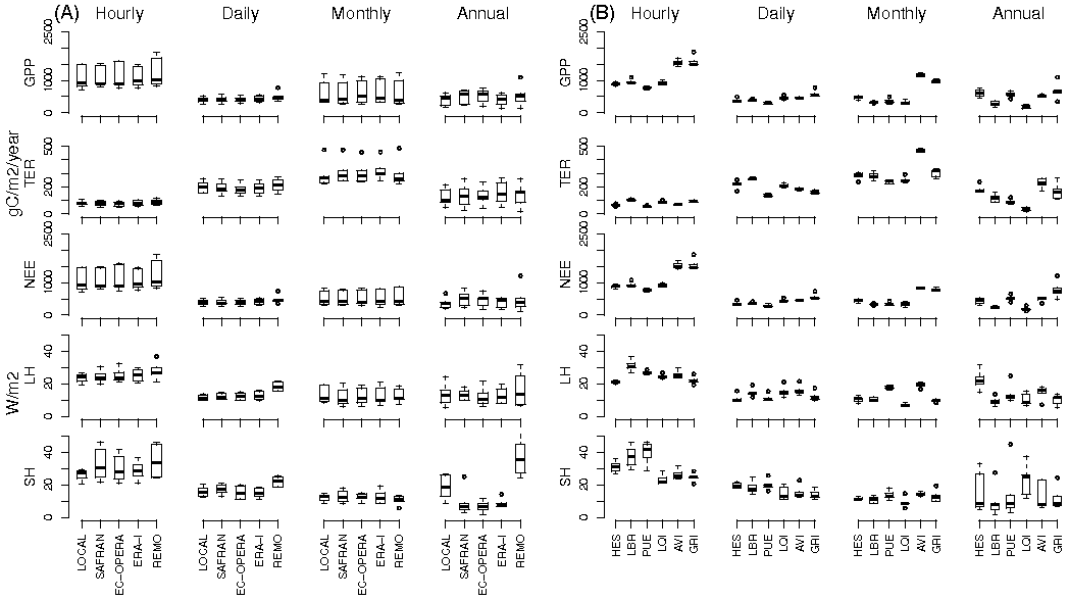


Fig. 6. Box plot of MAE between simulated and measured flux data over 2004 to 2007 except for EC-OPERA which covers only 2004–2006. **(A)** MAE across six sites; **(B)** MAE across 5 series. Column from left to right is hourly, daily, monthly and annual mean. The bottom and top of the box are the 25th and 75th percentile, respectively, and the band near the middle of the box is the median. The low and upper ends of whiskers represent the minimum and maximum, respectively.

BGD
8, 2467–2522, 2011

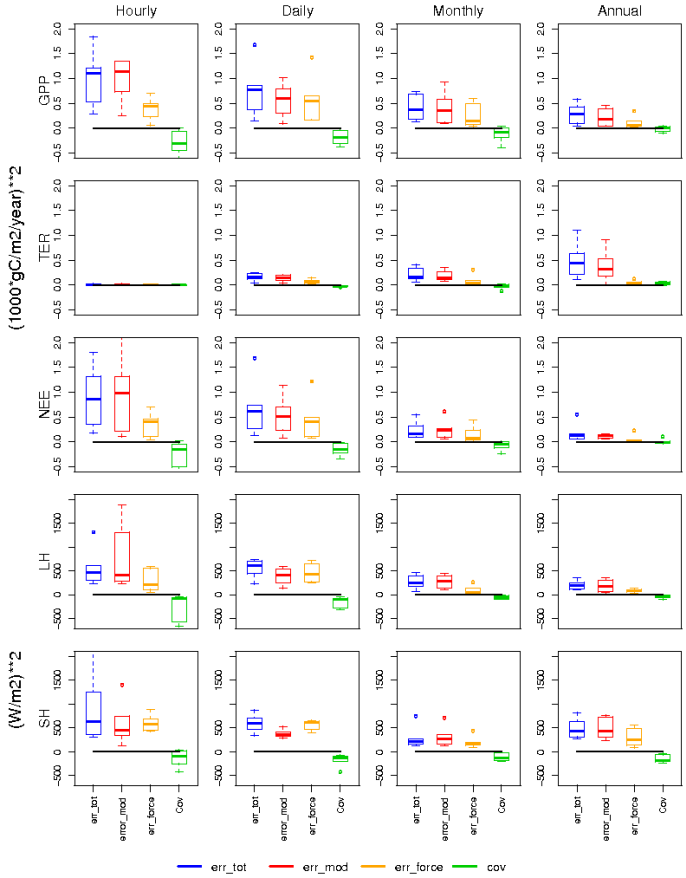


Fig. 7. Box plot of error estimation cross 6 sites during peak growing season. The bottom and top of the box are the 25th and 75th percentile, respectively, and the band near the middle of the box is the median. The low and upper ends of box represent the minimum and maximum value. err_tot, err_mod, err_force and cov denote total model error, model error, forcing error and covariance, respectively.

BGD
8, 2467–2522, 2011

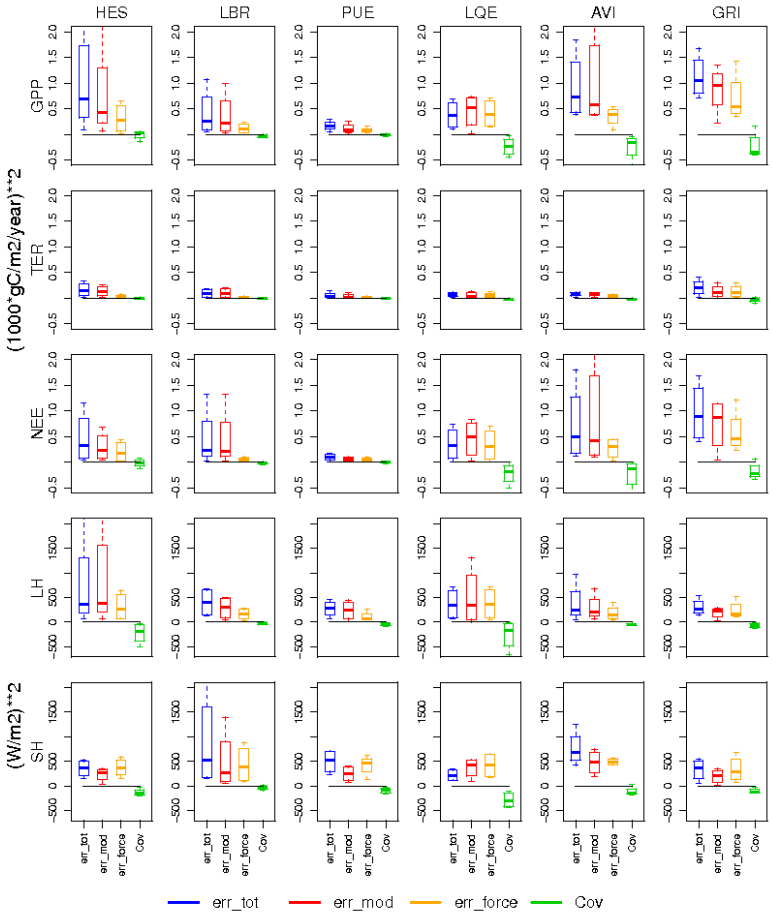


Fig. 8. Box plot of error estimation cross 4 time scales at each site during peak growing season. The sympols of boxplots are defined as the same in Fig. 7.

BGD
8, 2467–2522, 2011

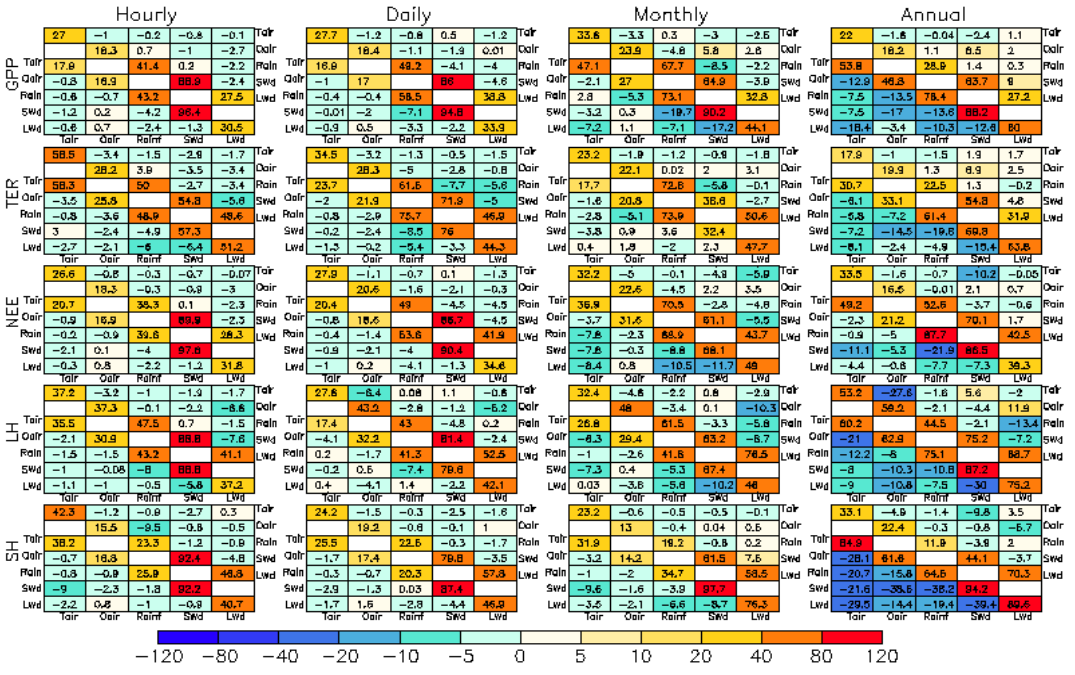


Fig. 9. Summary of forcing error caused by meteorological drivers and the covarance between the meterorological variables. Diagonal terms: contribution of each meterological variables; non-diagonal terms: contribution of paris of meterological variables. The upper-right triangle is calculated based on annual flux and the lowerleft diagnonal at peak growing season. See text for the details of calculation.

BGD
8, 2467–2522, 2011

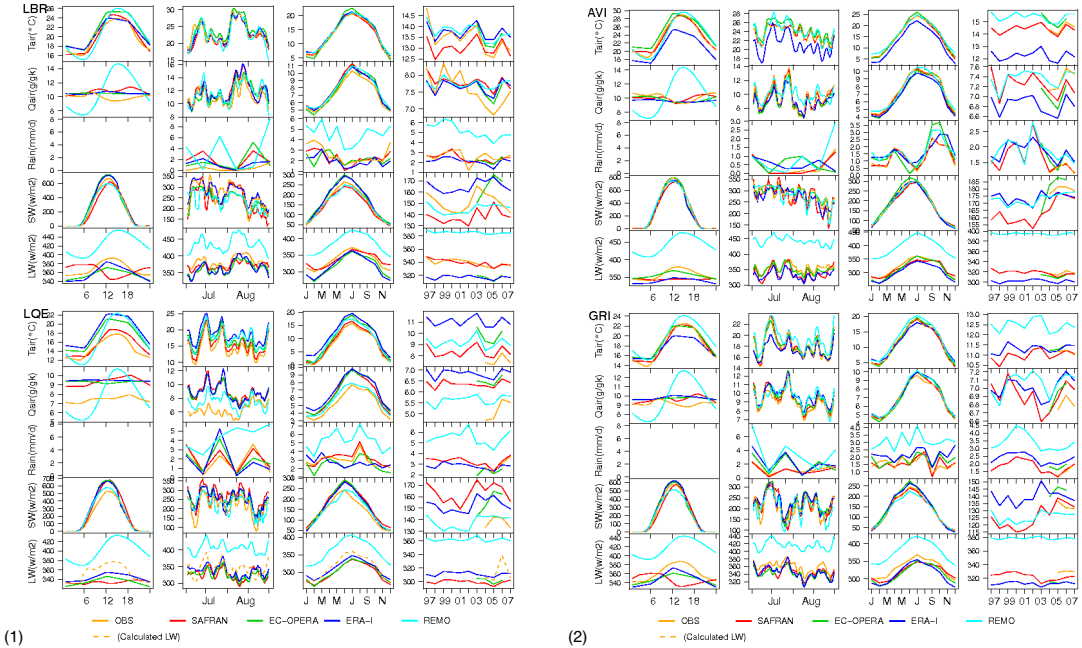


Fig. C. Meterological drivers in in situ and in gridded data sets at **(1)** LBR and LQE; **(2)** AVI and GRI. First column: hourly mean diurnal cycle over PGS. Second column: daily mean with a running mean of 3 days for July–August of 2003 at LBR, of 2005 at LQE, AVI and GRI. Rainfall is calulated as 5-day aggregated values; third column: monthy 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 but in dash lines. See text Sect. 2.1 for the definiton of PGS.

BGD
8, 2467–2522, 2011

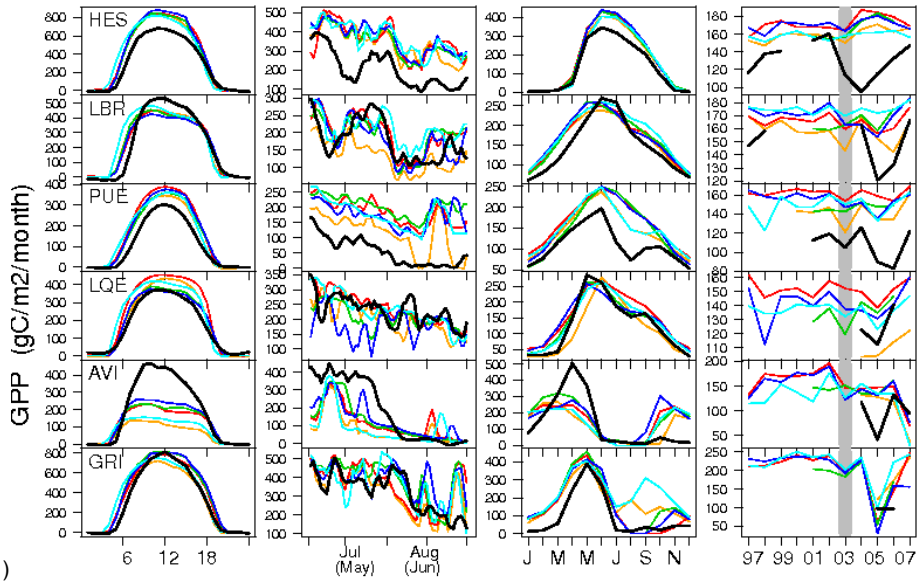
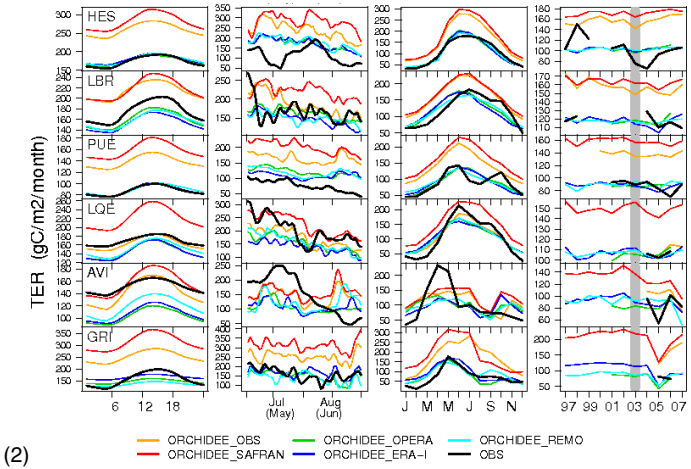
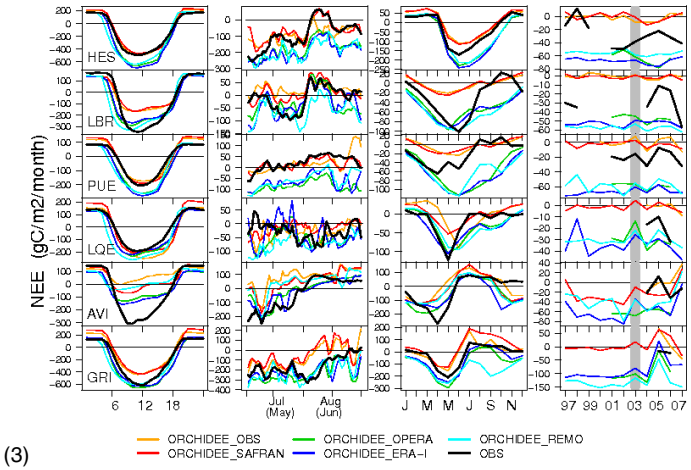


Fig. D. Simulated and measured carbon fluxes ($gc\,m^{-2}\,month^{-1}$) and water fluxes ($W\,m^{-2}$). **(1)**: GPP; **(2)**: TER; **(3)**: NEE; **(4)**: LH and **(5)**: SH. Hourly mean diurnal cycle over peak growing season correspnd to 2004–2007; daily mean with a running mean of 3 days for PGS of 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; "LOCAL" stands for the simulation driven by observed meteorological forcing. The measured flux is refered as "OBS".

BGD
8, 2467–2522, 2011



(2)



(3)

Fig. D. Continued.

BGD
8, 2467–2522, 2011

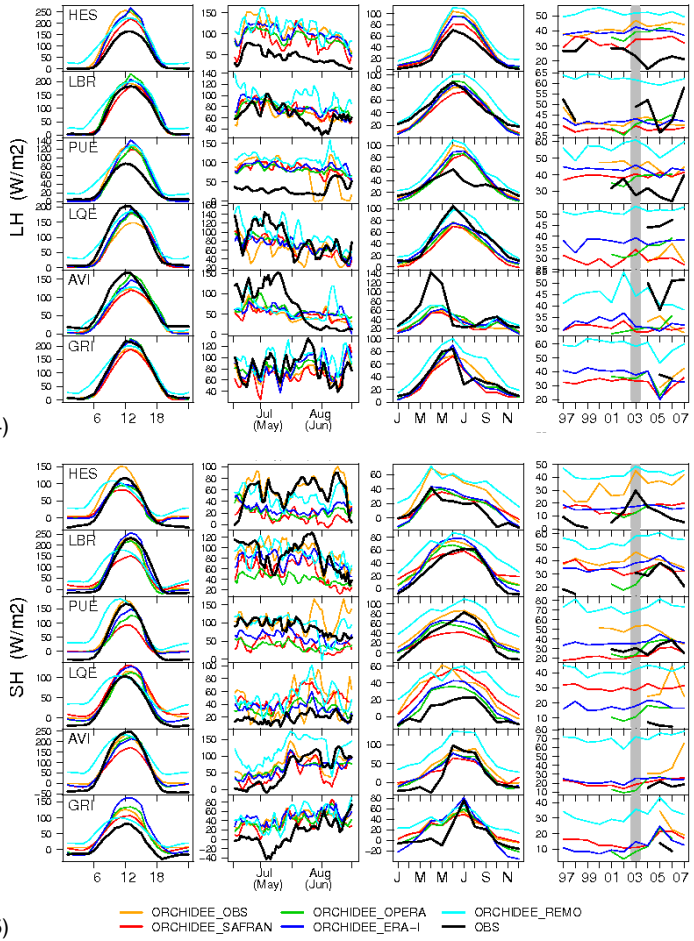


Fig. D. Continued.

BGD
8, 2467–2522, 2011