

Interactive  
Comment

## ***Interactive comment on “How errors on meteorological variables impact simulated ecosystem fluxes: a case study for six French sites” by Y. Zhao et al.***

### **Anonymous Referee #2**

Received and published: 4 May 2011

Zhao et al. present a model study testing the driver sensitivity of a well known terrestrial ecosystem model, ORCHIDEE. They compare model runs at six different eddy-covariance sites in France, driving the model both with observed in-situ meteorology, and meteorology derived from regional climate products. The main conclusions are that inaccuracies and biases in data from the large scale meteorology leads to differences in estimated fluxes of carbon and water from the ecosystem model. The authors argue that driver error is thus an important source of error in large scale simulations, particularly at longer time-scales.

The experiment is well conceived in that the authors clearly identify climate products

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



in which they are interested, and suitable sites within the region of availability of those products. The comparison of the climate products is rigorous and should be of interest both to people using those products, and, to a lesser extent, to those with a broader interest in the extent of variability between different climatic data sets. It study is poorly conceived, unfortunately, in that the ecosystem model is not well parameterized for the sites, leading to a model which poorly characterizes the ecosystem sensitivity to climatic drivers.

The baseline ORCHIDEE model performance is very poor, and seems much poorer than other works previously presented (e.g. Krinner et al., 2005; Jung et al., 2007a), particularly at the water stressed sites (where it has previously been shown to perform quite well (Keenan et al., 2009)). This leads to a common occurrence in the analysis where the model performs better when using regional climate data instead of climate data measured at the site. This detracts greatly from the significance of all statistical comparisons made between model output driven by local or regional based climate. The authors therefore are forced to spend much of their analysis on teasing out the difference between what they call model structural error, and forcing error. I find this unfortunate, as I would argue and sensitivity in model output to changes in drivers is fundamentally erroneous as the model sensitivity to climate was not properly characterized by the model in the first place (as evidenced by the poor model performance using site observed driver data). The results therefore are not necessarily generalizable to other models. The focus on trying to separate driver error from model error prevents a detailed analysis of the (far more interesting) relative sensitivities of GPP and RE to errors in climate drivers.

I see two different ways in which these concerns could be addressed:

1. The authors could argue that the model uses PFT parameters and therefore should not be 'tuned' to a site (i.e. site level parameters should not be introduced). In this case, all comparisons with site level fluxes should be omitted, allowing the authors to focus on a more regional analysis. This would greatly benefit the comparison of the

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

climate data products as independent meteorological stations could be incorporated (currently the six sites available to test your climate products far underrepresent the spatial heterogeneity of the region). Regional model runs could then be performed using each of the climate products to test sensitivity of model outputs to differences in drivers (as in Jung et al., 2007b, but focusing on the relative sensitivity of GPP and RE, and thus NEE).

2. The authors could consider the site level measured values of NEE and derived values of GPP and RE to be important baselines, in which case ORCHIDEE simulations would need to be rerun, ensuring that the model is accurately reproducing the observations when driven by the observed meteorology. This should not be too difficult, given that the model is already set up to run at these sites.

---

Abstract: Page 2469 Line 6: please clarify what 7-14 degrees celcius refers to.

Line 11: ‘The seasonal cycle of air temperature, humidity and shortwave downward radiation is reproduced correctly by all meteorological models (average  $R^2 = 0.90$ ).’ It is unclear what the authors mean by this. Do they mean the magnitude of change between seasons? Please be more specific.

Line 13-15: “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” This statement makes the inference that sea-breeze and altitude are responsible for the errors in the gridded data products. Such vague statements are not necessary and could be misleading. If you have shown that sea-breeze and altitude are the culprits then state so clearly. Otherwise, omit such statements.

Line 17-18: “ $R^2$  between modeled grid point and measured local meteorology going from 0.35 (REMO model) to 0.70 (SAFRAN model)” This statement could be read as the SAFRAN model is much better than the REMO model. Again, this is misleading,

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

as here you are reporting the limits? This is not clear and should be reported more precisely.

Line 23-25: “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” There is a break in the flow of logic here. The reader has the impression that you are reporting results (not methods) and are saying that model error is comparable (or of a similar magnitude to) forcing error. Please rephrase.

Line 25: “The forcing error is the largest on a daily time scale, for which it is as large as the model error.” It is to be assumed then that there is no bias in the forcing error? If there were to be a bias, this would lead to small errors in modeled NEE on the daily scale accumulating to larger errors in the annual sums. The first line of the abstract states that biases in forcings are to be analyzed, but there is no statement on the result of this analysis of biases other than that using different forcing products gives you different modeled NEE. Please clarify this and revise the whole abstract.

Page 2471

Line 8: “which is not the norm in many biosphere models”. Please remove. It could equivalently be stated as “which is the norm in many biosphere models”, and thus has no information content.

Line 21: “uncertainty” Do you mean error?

Line 19-25. Perhaps bullet point these questions to aid the reader.

Line 25-28 These are two separate questions here. Please separate.

Page 2473, Line 12: “The six sites cover” -> “The six sites (Table 1) cover”

Page 2474. Line 16: “Stomatal conductance is reduced by soil water stress (McMurtrie et al., 1990)” This was shown to be an ineffective approach for simulating soil water

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

stress in ORCHIDEE by Keenan et al. (2009). That probably explains the very poor simulation of GPP by ORCHIDEE with observed climatology at the water stressed sites included in the comparison (PUE, AVI, and GRI).

Page 2475, Line 6-7: “The last three drivers have no impact on the model output, and are thus discarded in the follow.” Surface pressure and wind speed have no effect on ORCHIDEE output? Then why are they in the model?

Line 7: “ in the follow.” in the following.

Line 10-11: “with data-gaps of 1% for Tair, Qair, rainfall and SWdown and of up to 12-days in duration at few sites.” Please rephrase.

Section 2.4: Given the length of the manuscript, this section could be omitted.

Page 2482: Line 27. “other ataproducts”

Page 2483: Line 4. “This is possibly due to a slope exposure of 5 the LQE tower”. Please clarify – a slope exposure would justify a lower than normal Swdown, but higher than normal?

Page 2483: Line 7. “REMO is worse than the other 3 models” Please be careful when making such statements. REMO does not perform as well as the other 3 models at your sites, and that is all you can say. Each model is designed differently and may perform better or worse for different conditions. A model may be generally worse than another, but you can not say that conclusively with your limited number of test sites. This is an important point and the manuscript should be revised throughout to account for it. e.g Line 18; “SAFRAN is superior to other models ” ... at our sites.

Line 24: “This shortcoming is however rather unimportant for CO2 flux modeling.” Such statements should either be proven by the results presented in the manuscript or backed up with citations. In this case, hourly rainfall variability may well be important if you are interested in modeling hourly CO2 flux.

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

Page 2484, Line 6: “and pointing out to local feedbacks of the vegetation” Variability in SAFRAN performance does not suggest local vegetation feedbacks.

Line 10: “given their coarser resolution”. This is unclear. Is the LQE grassland site not representative of the broader area contained within the coarse resolution grid cell of the gridded products? Again, this is an example of seemingly unsubstantiated statements.

Line 21: “This suggests that model structural errors largely explain the small values of R<sup>2</sup>”. This is not true. The ORCHIDEE model has been shown to perform well at these sites (too many papers to cite here). The problem most likely is that the model was not parameterized correctly for these sites. The assumption seems to be that given PFT general parameters the model should perform well ‘out-of-the-box’ so to speak. But there is little reason to believe that a particular site is representative of its corresponding PFT.

Line 22: “This also suggests that for site-level study, ORCHIDEE needs to be further calibrated.” This does seem to be the case, and in fact it would appear to me to be the main problem with this manuscript. The authors attempt to use the measured flux data to test the impact of using different climatic drivers on NEE, but make no attempt to ensure that the model is ‘behaving’ well at the sites when driven with the observed met data. Such an approach invalidates any statistical comparison of ORCHIDEE driven with different climate data to ORCHIDEE driven with observed met data. What is the meaning of better or worse in this case when ORCHIDEE can perform better with met data from a gridded product?

Page 2485, Line 4-8: “On average for daily scale, driving ORCHIDEE with OBS meteorology gives higher correlations than 5 when using atmospheric analyzed meteorology, except for the LQE mountain grassland where using a modeled meteorology improves the value of daily R<sup>2</sup> over five fluxes from 0.16 to 0.28. This indicates error compensation in ORCHIDEE, where a biased forcing compensates for a structural bias.” That ORCHIDEE with OBS met is only better than ORCHIDEE with gridded met makes for

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



a very poor comparison. With only 6 sites, it would not have been much effort to make sure ORCHIDEE was performing as well as it could at each site. If ORCHIDEE is not parameterized correctly for the sites, then its climatic sensitivity must be poorly characterized (bad parameterization) and so any quantification in terms of gC sensitivity to different climatic drivers (as presented later on in the ms) will be invalid.

Page 2485: line 20: “This indicates that poorly captured ecosystem processes that control the model-data misfit differ at each site”. The authors are to be commended for the openness in which they present the poor performance of ORCHIDEE at the sites. That said, I would strongly encourage them to rerun the analysis with better model runs. It is well known ORCHIDEE can do better than this, and I am surprised that ‘team ORCHIDEE’ have an interest in releasing such poor results. Can ORCHIDEE really not do better?

Page 2487, line 12: “model structural error”. Be careful with the use of this term throughout the manuscript. What you are talking about is a mix of model structural error and parameter error, which are two very different things. Model error may be a better term to use.

Page 2487, Line 15: “For TER, the forcing error  $\varepsilon F$  (orange) on hourly scale is negligible, because soil temperature and soil humidity that control soil respiration in ORCHIDEE exhibit no diurnal variability.” So there is no bias in the temperature in the gridded drivers?

Page 2488, Line 3: “1.8 and 3.0 g C m<sup>-2</sup> year<sup>-1</sup>” 3gC year<sup>-1</sup> is a very very small amount. There’s something I’m not getting here? Is this gC day<sup>-1</sup>? Please report errors on the scale you are referring to.

Page 2489, Line 10: “ $\varepsilon F = 716, 286$  and  $644$  g C m<sup>-2</sup> year<sup>-1</sup>” Again, I would really question what these values mean given that the model performs so poorly. If the modeled climate response to OBS met is not correct, then the model must be either over or undersensitive to some climatic driver(s). You then swap that climate driver for

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

one from a 'gridded product', and are interested in how sensitive model output is to the change in driver, but model sensitivity is poorly characterized.

Page 2492: Line 20: "by meteorology than by " as by

Page 2493, Line 8-10: "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". The situation of compensating errors can not be considered to be something positive! Please remove this statement.

Page 2495, Line 7: "REMO refers", REMO here refers to...

Page 2511; Fig 2. Please revise Axis labels and tick values – it is impossible, or at least very uncomfortable, to read this, and I have good eye sight!

Fig 4. Same comment as above. Bear in mind that a lazy reader will skip over a graph if they meet any difficulty in interpreting it.

Fig 5. Similar comment as above – for example there is lots of room within each box, the number font could be doubled and still fit in the same space. These are small details but important if you want people to pay attention to your work.

Please considered these comments when revising the other Figures in the manuscript. Also please use SI units for all axis – e.g., "gC m<sup>-1</sup>Ås<sup>-1</sup>", not "gC/m/s",

Fig. D. Please insert a legend, each figure should be independently interpretable.

---

Interactive comment on Biogeosciences Discuss., 8, 2467, 2011.

**BGD**

8, C799–C806, 2011

---

Interactive  
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

