Geosci. Model Dev. Discuss., 3, C811–C818, 2011
www.geosci-model-dev-discuss.net/3/C811/2011/

**Geoscientific
Model Development
Discussions**

Interactive comment on "A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale" *by* E. Blyth et al.

E. Blyth et al.

emb@ceh.ac.uk

Received and published: 11 February 2011

Author response: I thank the reviewers for their consideration of this paper. I agree with most of the comments and have attempted to address them. More analysis would be very worthwhile, but the authors wish to lay this data and these tests out to the community as a benchmark so that they can be used for further assessment. Therefore I have not attempted to add more analysis, as suggested by Reviewer 3, rather I have tried to be more explicit about the aims of the paper. Reviewer 1. It is true that how the data sets were chosen for the comparison exercise was not described in detail. This is to some extent because the subject is immense. On the one hand there are many

reported datasets available, on the other it is sometimes hard to obtain the data and get free access to it. There are no rules for choosing how many river basins to choose, or indeed which rivers. We wanted a range of climates and biomes and we needed the number to be manageable in terms of analysing the data. More research was carried out for the Fluxnet data we used in terms of checking for energy closure and this is described in the Blyth et al (2009) paper. But still the overriding principle was to have a range of climates/biomes and a manageable number. As for the offline question: that is because the climate forcing from the model may be a very significant factor in determining the response of the land surface model, not the model itself. I have added some text to that effect. The issue of partitioning different parts of the model would be good indeed, but that is another study and another paper! The question of whether the benchmark data influences the development of the model and whether that is a good thing or not, is a good one. However, we are not calibrating the model to the data. We are merely locating gross errors in the model. When we identify a region where the model is not working we can go and get some more data or some processes data and calibrate it. I think the benchmarking system is not the same as model calibration. I have added some text to the effect. Specific edits to answer the reviewer: Page 1833. Line 24: Insert at beginning of the para.: 'It was decided that a small number of FLUXNET sites would be used in this initial benchmarking system to allow researchers to see in a glance the overall performance of the model. For the present study. . . . . .' Page 1835. Line 24: Insert sentence after '. . .. FLUXNET data. A small selection was chosen so that the overall picture could be obtained with one set of graphics. Initially 8 river basins were chosen, including the Parana. However it was not possible to find a stretch of that river (in that very productive region of South America) that was not heavily managed. So we reduced the number to 7, covering the key regions of the Americas, Europe and Africa. In the Americas. . . . . .' Page 1832. Line line 17. After '. . .offline from the climate model because the climate forcing data may be the most important factor determining the response of the land surface mode, not the land surface model itself. In addition. . .' Page 1844. Line 20. Insert '. . .

researchers. It is important that the benchmark data remains somewhat independent of model development and is only used to indicate gross errors. However. . ..'

Reviewer 2. Title: The inclusion of the words 'at both the global and seasonal scale' is there to indicate to the reader that this is not just based on point observations and also that we are looking at a broad brushed assessment of the model – no interannual variations, no diurnal cycles, no trends: just seasonal and at the global scale. I am happy that this title gives the reader a good indication of whether they want to read further or not. Abstract: I agree a broader sentence could be added. Page 1830: line 2. Precede by: 'Evaluating the models we use in prediction is important as it allows us to identify uncertainties in prediction as well as guiding the priorities for model development.' I disagree that I should put in the Abstract the exact CO2 data we used. All the observations in the abstract are referred to in very broad terms. More details are included later. Introduction: Including a brief description of other aspects of land-atmosphere exchanges that may be changing over time that are modelled is a good idea. Page 1831. Line 4. Add to end ' global carbon cycle and changes in the land cover due to land used for food and fuel production can have impacts on the weather and climate (Cox et al., 2000).' Page 1832. Line 13. 'Similar initiatives, such as the CLAMP project (see Randerson et al, 2009) and the study by Cadule et al (2009), have been. . .' Page 1832. Line 15. 'MOSES-TRIFFID (Met Office Surface Exchange Scheme with the Dynamic Vegetation Model: TRIFFID – essentially the same as JULES) carbon cycle. . ...' Page 1832. Line 17. 'GSWP2 (Global Soil Wetness Project 2)..' Section 2.1 The fluxnet data was only taken from one year. This was so that we could choose a year that had some interesting weather – rather than just the mean climate. In this way, the land surface (and the model) should be responding to a state that it was not necessarily adapted to – like a dry year in the Amazon. The year that is chosen is for each site is stated in the table. Add text as follows: Page 1834. Line 3. Add a new sentence. '. . ..weather data. A single year of data is chosen for this comparison to highlight the response of the observations (and the model) to conditions that may not be the climatological mean for that region. For instance, we chose a very dry year

for the Amazon forest site comparison which enables us to assess the ability of the model to reproduce drought conditions. The problems. . ...' Section 2.2: More clarity and details are needed for the reader to comprehend the methodology we used. We will alter the text as follows: Page 1834. Lines 26 – 27. Delete from 'Using an. . .. Kaminski et al, 1999a, b).' And insert the following: 'We used the global inverse model of atmospheric tracer transport model TM2 (Kaminski et al., 1999a, b).The 'station matrices' for the adjoint model (see Kaminski et al., 1999a, b) were derived from the TM2 (Heimann et al., 1989) using the mean wind fields from ERA40 the ECMWF reanalysis for the year 1987. The monthly station records were post-processed to extract the detrended, mean monthly observations between 1980-1990 and same procedure was applied to model outputs before the net monthly fluxes were supplied to the transport model. Monthly ocean-atmosphere fluxes were taken from an ocean carbon cycle (HAMOCC3, Maier-Reimer, 1993) and monthly CO2 emission fields from fossil fuel and cement production were based on Marland et al., 1989, assuming no seasonality in emissions.. The model that produced these matrices of contributions performed reasonably well in the TRANSCOM experiment. It does not represent the state-of-the-art with respect to atmospheric transport model analyses, but is a practical alternative that can be distributed to the JULES community as part of the benchmarking exercise. Given the over-view of seasonal fluxes benchmarking that is being delivered it was felt that this was adequate.' The results of the comparison show a large anomaly in the model. This is probably due to the overestimate of the soil respiration due to incorrect initial soil carbon contents, although a thorough analysis would be needed to pin-point the exact explanation. We will add some text as follows: Page 1840. Line 10. 'This discrepancy is probably due to the overestimate of the soil respiration, also shown in the fluxnet data comparisons, which is possibly due to incorrect initial soil carbon contents, although a thorough analysis would be needed to pin-point the exact explanation.' The y-axes of these plots have been changed to ppmv. References: GLOBALVIEW-CO2 (1999)Cooperative Atmospheric Data Integration Project – Carbon dioxide. CD-ROM. Boulder, NOAA, CMDL, Colorado Marland, J Boden TA, Griffin RC et al., (1989) Esti-

mates of CO2 emissions from fossil fule burning and cement manufacturing, based on US Bureau of Mines manufacturing data. ORNL/CIAC-25, NDP-030, Carbon Dioxide Inf. Anal. Center. Oak Ridge National Laboratory, Oak Ridge, Tenn. Maier-Reimer E (1993) Geochemical cycles in an ocean general circulation model – preindustrial tracer distributions. Global Biogeochemical Cycles, 7(3), 645-677. Section 2.3 The citation for GRDC has been amended slightly and follows the form requested by GRDC (see below). We have also given the address of the website. As far as we are aware, there is no further reference for these data. We have added a separate citation for the mean monthly data for the Congo (GRDC, 2008) following the form suggested by GRDC. The GRDC-requested form of citation is "The Global Runoff Data Centre, 56068 Koblenz, Germany" so we will amend the ref at final line of p1835 to The Global Runoff Data Centre, 56068 Koblenz, Germany, http://www.bafg.de/GRDC Top of page 1836, after "1903-1983" add "(Global Runoff Data Centre, 2008)", then in the references: Global Runoff Data Centre, 2008, Long-Term Mean Monthly Discharges and Annual Characteristics of GRDC Stations, Global Runoff Data Centre, Koblenz, Germany. Section 2.4 The PFTs from Foley are aggregated using a look up table which is now included as Table 3. BL NL C3 C4 Shrub Bare soil Tropical evergreen 0.9 0.1 Tropical deciduous 0.8 0.15 0.05 Temperate BL EG 0.9 0.1 Temperate NL EG 0.8 0.15 0.05 Temperate deciduous 0.1 Boreal EG 0.8 0.15 0.05 EG/Deciduous mixed 0.4 0.4 0.1 0.1 Savanna 0.2 0.75 0.05 Grassland/Steppe 0.9 0.1 Dense shrub 0.15 0.7 0.15 Open Shrub 0.6 0.3 0.1 Tundra 0.35 0.35 0.3

Page 1851. Add to Table 2 title: Metric RMSE of monthly values Page 1853. Change label to 'Latent Heat' Page 1855 and 1856. Change to Micro moles.

Reviewer 3. This reviewer has pointed out that the promise of the abstract and introduction are not matched by any delivery of analysis in the results section. I can understand the complaint, and I think this is partly due to an overstatement of the mission of the paper due to enthusiasm of what can be done with this data. In fact, the real mission of the paper is to set out the data and the set of metrics chosen to quantify the per-

formance of the model: the intention being that future researchers will be able to use this to make more in-depth analyses. I will therefore address the issues in some cases by toning down the promises in the abstract and introduction. However, including more information about Metrics is a good suggestion and I will include that, as well as a more explicit discussion of uncertainty as follows: 1: Reviewer 3 quite correctly raises the issue that we are not performing a rigorous assessment of interannual variability. We do understand the importance of this, and we recognise that we have not considered this in our metrics of model success. For the carbon cycle in particular, we can learn things about model performance by analysing individual years. ElNino/LaNino variation and associated changes in the rate of increase of atmospheric carbon dioxide concentration can inform us of, especially, tropical response of vegetation to temperature changes. Nearly all of the Figures presented in our long paper are assessment of seasonal model performance against a broad set of measurements. We now discuss this in the new section on Metrics (see response to point 2). 2: Page 1839. Line 7. Insert a subsection '4.1 Metrics. In order to provide the user with a simple assessment of the error in the model, the same diagnostic is used for each of the datasets: the mean monthly value of the property that is observed. This allows us to compare very different types of observations. For some of the observations this was straightforward such as the river flows, the fluxnet data and the atmospheric CO2. However, for distributed datasets such as the NDVI series, it is not obvious how to reduce the observations and model output to single time-series of mean monthly values. For this system, we decided to encapsulate the seasonality of the NDVI and LAI from the model by looking at the area-average value of the selected river basins. Most of the river basins are fairly uniform in climate and vegetation type, and so this represents a simple way of representing the mean response of the plants to climate in terms of phenology, which also allows us to compare the plant response directly to the response of the water balance through the river flow. The exception is the Niger River which passes through very contrasting climates. In this analysis, a simple test of assessing the strength of the seasonality was used, as it represents a first-order test of the performance of a land-surface model. A

future version should include interannual variability as this would allow us to assess the performance of the model to extremes of weather such as El Nino or La Nina years. Such an assessment is feasible with the current model configuration. No attempt was made in the benchmarking at this stage to include uncertainties in the metrics, or to combine them in any way. Combining the metric across the various measures of performance requires a deeper understanding of uncertainty and further work is required to address this. Instead the simple Root Mean Square Error of the mean monthly quantity is presented. The idea is that researchers using the JULES model can compare the model with any changes that have made to these simple diagnostics.' 3. Page 1830. Line 13. Replace 'The results show that combined use of observations of carbon and water fluxes in essential in order to understand. . ..' With 'A few examples are chosen to demonstrate the importance of using combined use of observations of carbon and water fluxes in essential in order to understand. . ..' 4: Page 1838. Line 7. Introduce new sub-section: '3.1 The model used in this benchmarking is the JULES model. It s the land surface model used within the Hadley centre GCM. The description of the model is given in Cox et al, 1999 and Cox et al, 2001. The model has had some upgrades since then but essentially remains the same. It is now a community model and is distributed via its website: www.jchmr.org/jules which contains further content about the model. It is a mechanistic model of the land surface including linked processes of photosynthesis and evaporation, soil and snow physics as well as plant growth and soil microbial activity. These processes are all linked through a series of equations that quantify how the soil moisture and temperature govern the evapotranspiration, heat balance, the respiration, photosynthesis and carbon assimilation. It runs at a subdaily step, using meteorological drivers of rainfall, incoming radiation, temperature, humidity and windspeed as inputs.' 5: The errors in the model are exaggerated by the fluxnet analysis as I used a non-interactive LAI. What I was interested in seeing was not the absolute error (which can be overcome with calibration) but the relative error of the evaporation and the photosynthesis. I will add some text to that effect. Page 1839. Line 20. Add at end 'The errors in this analysis are rather large. This is mainly due

to the simple approach of using a fixed, predetermined LAI for these varied sites. The errors can be reduced at a site by calibrating and adjusting the parameters for the local conditions. However, what is more interesting for this assessment is the way the evaporation error agrees or contrasts the error in the GPP. It is the relative errors that allow the model user to identify process anomalies.' 6: I can see that we maybe got a bit carried away with the Introduction! Page 1831. Delete line 8-20. 'However, ...The Unified model).' Plus edit line 7.' At a minimum, land surface model s used in Climate Prediction need. . .' and edit line 21 'in the UK Hadley Centre Climate Prediction model (the Unified Model).' Specific points: Happy to delete that paragraph.. I think it repeats the discussion and doesn't add anything. Page 1844. Delete lines 1-5. I have tried to remove inconsistencies in the test. Page 1833. Line 11. Delete 'half-hourly'. Replace with 'Hourly'.

Interactive comment on Geosci. Model Dev. Discuss., 3, 1829, 2010.