

Anonymous Referee #2

Received and published: 30 March 2012

The paper raises a very important issue of evaluating land model performances. The study figures out which different aspects for land model evaluation have to be taken into account. It provides a summary about available data which are useful for benchmarking. But, unfortunately the paper fails providing a benchmarking protocol. It should be clarified that benchmarking is a measure of model performance as well as model validation is a part of this by evaluating model data against observed data. For some parameters there is even a lack of data and only a model intercomparison could overcome this to evaluate the uncertainty range of estimation. Many data listed in tables 2 and 3 are useful for validation on the local scale, but in my opinion all global data are model results. Especially should be distinguished much more precisely that some data only provide model intercomparison data. There are only few observed data which are useful for validation, but as already mentioned a benchmarking is lacking in this paper.

Thanks for the comments. It is not clear to us what the reviewer meant by benchmarking protocol. This manuscript is intended to describe a general framework for benchmarking land models. Specific procedure of benchmark analysis is offered in some of the published papers, such as that by Randerson et al. (2009).

We agree with the reviewer that benchmarking is a measure of model performance and model validation is a part of this by evaluating model data against observed data. We have made major revision in several parts of the manuscript to highlight the point.

We also agree with the reviewer that model intercomparison could evaluate uncertainty ranges of estimation. The sentences on lines 88-94 and lines 328-330 discuss the point.

Thanks to the reviewer's comment, we reorganized that not all data in Tables 2-4 can be used to benchmark land models. We changed the table captions to be "Candidate benchmarks ..." to reflect the point.

The distinction between validation against observed data and model data has to be described more clearly. Many suggestions made by this study are already considered by many studies, e.g. the comparison against flux data as well as runoff data or fpar.

We appreciate the reviewer's comment. The first two paragraphs in Section 4.2 describe the distinction between observed data and model data against which models may be evaluated.

What is really needed is a comprehensive study where all land models contribute to the estimation of the uncertainty of biophysical cycles. Even the attendance of disciplinary data models as GPP or soil respiration assessment methods would support the benchmark process. The study could contribute to this effort by structuring the benchmarking. Therefore it needs a more detailed analysis of which observations are useful and which model data at which spatial and temporal resolution should be used.

It is a good suggestion. We do need to understand fundamental properties of models and information content of data at various temporal and spatial scales. We added one paragraph in section 3 (line 246-262) to illustrate that understanding fundamental properties of land models potentially can make benchmark analysis much more effective. We hope the community will take to your suggestion soon.

Different input data has to be taken into account to capture this uncertainty as well.

This is a very critical point, which should be considered in any analysis on land models.