

Interactive comment on “Evaluation of biospheric components in Earth system models using modern and palaeo observations: the state-of-the-art” by A. M. Foley et al.

Anonymous Referee #2

Received and published: 9 September 2013

The paper by Foley et al. describes structural and parameter uncertainties in models and challenges to performing model evaluation. Acknowledged to be the product of a mini-conference focusing on Earth system model (ESM) evaluation using modern and paleo observations, the paper reviews key concepts in model-data comparison and summarizes challenges tied to using observational datasets with large uncertainties (e.g., satellite data and paleo records). The authors also describe three levels of ESM evaluation metrics, identifying various distance measures and providing limited and recent examples of their application to diagnostics in biogeophysical model-data comparison. Recommendations for model evaluation and calibration are offered subsequently, but they are never demonstrated with new data and model results. Instead,

C4882

the authors rely upon a subset of existing work to provide illustrative examples, and they often summarize the discussion of other authors to make their points. Section 4 appears to contain a collection of disparate subtopics in subsections of various lengths joined together as recommendations. This section could be significantly improved by reorganization.

Overall, the paper does a good job of reviewing challenges to rigorous model-data comparison, of identifying traditional methods for calculating various distance metrics, and of describing the current state of biosphere model evaluation. The authors recognize the importance of community-based evaluation efforts, like the International Land Model Benchmarking project. The paper likely provides an accurate representation of the subject mini-conference, and, as such, is a potentially valuable conference report. The paper does not present new science results.

Specific comments:

Page 10941, lines 9–10: The multi-model mean should always perform averagely since it is an average. You may wish to more clearly explain what you wish to say about model performance for regional climate.

Page 10946, line 20: The word “of” is missing between “because” and “the”.

Page 10948, line 13: The word “as” may be extraneous here.

Page 10959, line 22: The word “that” may be extraneous here.

Page 10960, line 7: The first “an” is unneeded.

Page 10968, lines 17–21: This sentence suggests that a larger number of model tests may reduce the ability of an investigator to understand or interpret model results. This makes little sense, but the authors may have a point there that is not conveyed clearly or correctly.