

## ***Interactive comment on “A comprehensive benchmarking system for evaluating global vegetation models” by D. I. Kelley et al.***

**Anonymous Referee #3**

Received and published: 30 December 2012

[general comments]

This paper addresses a comprehensive benchmarking system of global DGVMs. The Authors present simple and understandable metrics to compare simulation results and benchmark dataset, and show some results with globally available dataset using two related DGVMs and one simple model. It is well emphasized that the need for using benchmark dataset which is independent of modeling, and importance of updating new benchmark product and their use to improve DGVMs modeling. However, I suggest adding more comprehensive review for available metrics for benchmarks (other than the metric used here, e.g. NME, MPD, MM), and its pros and cons will improve the manuscript for the development of comprehensive benchmarking system. Also, it will be helpful if authors discussed the priority of required benchmark dataset to improve

C6932

future DGVMs development at the end of the manuscript in relation to the Luo et al. (2012) although it may requires the subjective decisions. Please find below a number of specific comments which the authors might want to consider in a revision of the paper.

[specific comments]

- 1) page 15729, line 21: Could you add SeaWifs dataset website?
- 2) page 15731, line 1: What is the definition of the 'disturbed' sites? The data has 'Managed' category in Management variable, but there is no single 'disturbed' category in that.
- 3) page 15735, line 7; page 15735, line 16; page 15737, line 2: Is is ambiguous to refer just Table 2 for each application. It will help if you could add reference numbers or categories in the column 4 of the Table 2.
- 4) page 15737, line 11: Table 2 show 1 for perfect disagreement. But in the text, it shows 2 for complete disagreement. Could you explain the consistency about this?
- 5) page 15739, line 5: Please provide LPJ version number used in this study.
- 6) page 15742, line 24: Is underestimate of NPP for the variance? If so, please clarify this.
- 7) page 15742, line 26: I couldn't understand where the values 1.26-0.56 are come from.
- 8) page 15772, table 5: It seems first appearance of "Annual average 1998-2005" should be "Annual average 1950-2005".
- 9) page 15747, line 3-4: Table 6 shows LPJ has marginally better performance with Luyssaert et al. (2007) dataset at the site locations, however, the text says all three models performs better against the Beer et al. (2010).

C6933

10) page 15752, line 18-19: Is is related to the 'at sites' score comparison between Luysaert et al. (2007) and Beer et al. 2010 in Table3? If so, please add reference to that in the text for clarification.

---

Interactive comment on Biogeosciences Discuss., 9, 15723, 2012.

C6934