

## ***Interactive comment on “Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model” by M. Jung et al.***

**Anonymous Referee #2**

Received and published: 26 August 2009

I congratulate the authors on an interesting and innovative paper, I have no hesitation in suggesting this manuscript be published with only minor revisions. While I do have concerns that I outline below, this manuscript offers important technique developments and research findings that are only partially affected by these concerns.

The main achievement of this work is the model tree ensemble development and selection technique, which is applied to Fluxnet upscaling. I do not share Reviewer 1's opinion that the paper's shortcomings in specifying causal mechanisms in the upscaling technique constitute grounds for major revision. I suspect that the ensemble model

C1689

tree technique presented here will be remembered long after its application to Fluxnet upscaling - there are many other potential applications.

My specific concerns are as follows.

1. The synthetic data case presented here, while a good test case, will not necessarily translate as well to Fluxnet upscaling (although I suspect it would). This may be in part a point reviewer 1 was making. While the upscaling technique assumed LPJ simulations were truth, training it on 0.5 by 0.5 degree grid squares (already a region) to upscale to a larger region may not tell us how the technique trained on single site data (i.e Fluxnet sites) would upscale to the region. This is hinted at in the conclusion but should perhaps be stated explicitly in the discussion. In general though, this synthetic data study is a good mechanism to test the technique, and care was taken in other respects to emulate the nature of Fluxnet data.
2. Another point that is mentioned in the conclusion but I felt should be part of the discussion was that the synthetic data study used the technique to simulate a closed system, whereas Fluxnet data are measurements of a real-world open system (in the sense discussed in Oreskes et al, 1994, Science).
3. I felt section 2.1 could give more detail about how TRIAL worked
4. I needed more explanation about the 10-fold cross validations in Section 2.1.2.
5. I feel that some of the terminology used should be more clearly defined before it's used. For example, "extrapolation" was used by the authors to imply the model tree was making predictions outside the range of its training set (usually termed "out of sample"), when the common usage meaning of this word does not necessarily imply this. To some people, extrapolation, prediction and interpolation are synonymous. Please make it perfectly clear when these have particular meanings.

Trivial changes:

Section 2.1.1 - 10-fold and cross-validation need to be hyphenated. Section 2.1.6 - 10-

C1690

fold and cross-validation need to be hyphenated. First sentence in Section 3.2, either use: "Providing realistic uncertainty estimates of the upscaling product is essential for its scientific use" or "Providing realistic uncertainty estimates of the upscaling products is essential for their scientific use"

---

Interactive comment on Biogeosciences Discuss., 6, 5271, 2009.

C1691