Anonymous Referee #2

Received and published: 23 December 2010

Overall Comments:

Evaluating estimates of the various components of the carbon balance from ecosystem process models can be difficult due to the lack of validation data of sufficient spatial density or extents. Moreover, inter-model comparisons require standardization of input driver data sets across models as well as considerable mobilization of resources for actually preparing and running the models. This manuscript presents a suite of techniques designed to circumvent some of these problems by exploiting the spatial autocorrelation structures inherent in output raster maps of carbon balance for three biosphere models. Variogram analysis of NEE, GPP and Re estimates for North America (and their time dependent changes in 'correlation length' and semivariances) are used to compare the spatial properties of model results that cannot be ascertained from looking at maps, while informing the next step of the study. Variable selection and geostatistical regression are used to identify those factors in the biosphere models that most influence NEE, GPP and Re without having to depend on complex model formulations and to allow ease of comparing these sensitivities across the models. I very much like the spirit of this manuscript, partly because it describes a more parsimonious approach to inter-model comparisons. But also because it brings together what has been usually up to now to separate 'knowledge arenas,' that of global ecosystem modelling and geostatistics, and here lies a vast amount of underexploited potential.

The work is well-motivated, technically sound and very well written (mercifully so).

Scientific/Technical Comments:

I notice that all variograms that the authors present are forced to pass through the origin. Is this also true of those variograms underpinning Figure 2? If the curves pass through the y-axis, and therefore have a positive semi-variance, this indicates a residual non-spatial variance. This is either interpreted as noise or the occurrence of spatial structure smaller than the sample spacing. How might this feature inform the study (e.g. computation of nugget-to-sill ratios as a way to standardize signal-to-noise ratios and intercompare model results)?

The choice of variogram model (exponential) is based both on experience, as well as examination of the experimental varigorams. An experimental variogram is created by taking an average semivariance within specific binned separation distances, then plotting those averages for each bin. The experimental variograms are not forced to go through zero (Figure 4). Thus, the experimental variograms showed no indications that a nugget was warranted, we chose an exponential model without a nugget to model the spatial variability. **Also, it is important to note that you really wouldn't expect a nugget for a biospheric model because the model is estimating flux at a 1 degree by 1 degree spatial resolution; and thus, you would not expect any variability at resolutions smaller than the resolution of the model.**

Presentation/Language/Structure:

I am a little surprised that the authors have not computed variograms for the data sets used to drive these models and used these to compare to the variograms of the model results. Particularly for the LUE models I would expect that much of the structure underpinning the variograms of

some of the carbon balance estimates (and particularly GPP) would be very similar to the NDVI data, since it is one of the only data sets that are not actually a result of interpolation and thus artificial smoothing. In fact, basic metadata about the original spatial resolution of the drivers appear to be absent. This is important information that can be used to guide the analysis and interpretations.

Please see our response to Reviewer #1's comments related to Table 1 and model driver metadata. Also see our response to Reviewer # 1's comments concerning constructing variograms for the environmental variables.

I think the manuscript would benefit from a brief introduction to the concept of geostatistics and the variogram, particularly as a diagnostic tool. It would be instructive to show an idealized variogram, label its components and describe them briefly. I mention this because I think that many reading this article (and journal) may be very interested in comparing ecosystem models but may not have a technical background in geostatistics.

We can add a brief introduction to the concept of geostatistics and variograms to the manuscript, as well as figure showing an idealized variogram.

The vector algebra is tedious to read, and is not central to your message. Section

2.3 can be reduced and moved into an appendix, which is also where you may want to place short tutorial on the variogram as a tool.

We are working to shorten the methods section and remove any detailed information that can be referenced through other manuscripts. Based on these revisions, we can determine whether an appendix would be warranted.

I wonder what the authors mean when they use the term 'spatial scale' – to me the term is ambiguous. Does it mean 'spatial extent' or 'spatial resolution,' (sampling density)?

We agree with the reviewer's comment and we are revising the manuscript to make this clearer and less ambiguous. When we refer to spatial scale, we are referring to the scale at which variability is observed, or the scale of spatial variability.

p. 7907, lines 9-10: Can the authors describe briefly (or provide and example) of how not accounting for spatial autocorrelation of modelled estimates can lead to misrepresentation of inferred relationships?

An example was provided in the text in the Introduction just after the statement in question: "For example, many environmental variables exhibit a seasonal cycle similar to NEE. When only one environmental variable is regressed against NEE, the derived relationship may be a result of correlation in their seasonal cycles rather than a true explanatory relationship. Therefore, the resultant regression represents a scaling parameter (e.g., how to scale the variable to look like NEE), rather than that variable's relationship to flux." However, we can try to make the connection between this example and the statement in question clearer.

p. 7908, lines 10-12: The objective in the paper is presented firstly as what the authors do not do, before what the authors actually do in the same sentence. This is awkward, please fix.

This has been fixed in the manuscript, as suggested by the reviewer.

p. 7911, lines 11-13: NOAA/NASA Pathfinder NDVI is from the NOAA AVHRR instrument and is essentially the same thing. Which NDVI data set (derived from NOAA AVHRR data) was therefore used in this model?

The model uses the Normalized Difference Vegetation Index (NDVI) from the Advanced Very High Resolution Radiometer (AVHRR) data processed by the Global Inventory Modeling and Mapping Studies (GIMMS) lab, version "g" (Pinzon et al., 2005; Tucker et al., 2005). We have clarified this in the manuscript.