Biogeosciences Discuss., 12, C5422–C5430, 2015
www.biogeosciences-discuss.net/12/C5422/2015/

Biogeosciences
Discussions

Interactive comment on "An objective prior error quantification for regional atmospheric inverse applications" *by* P. Kountouris et al.

Anonymous Referee #1

Received and published: 15 September 2015

Review of Kountouris et al, An objective prior error quantification for regional atmospheric inverse applications.

Overall:

This paper addresses and interesting and important topic with an interesting set of data and models. Overall the methods and intent are sound and the logic of the paper is clear. I have some significant concerns about the statistical methods, clarity of the presentation, and attention to prior literature. I would recommend publication, but not without significant revisions.

First, and most important, it is clear from the figures (and noted in the discussion) that the exponential model used to fit the spatial correlation functions is biased. The argu-

ment presented for the exponential model is entirely one of convenience. This paper should note, at minimum, whether or not this function does yield an unbiased representation of the true error structure and if not, what biases exist. This should be stated clearly, including in the abstract of the manuscript. The analysis would be strengthened by a comparison of multiple models, as suggested briefly in the discussion and as performed in other publications on this topic.

Uncertainty assessments are not provided uniformly, and the methods for the uncertainty assessment are never articulated. Please include uncertainty bounds with all of the major results, note them in the abstract, and explain the uncertainty assessments methods in the manuscript.

There is limited comparison to and evaluation of prior literature at a number of points in the manuscript. See the detailed comments below. Most notable, a very similar analysis was published in Biogeosciences in 2013 but is not mentioned in this manuscript.

The figure quality is inconsistent. See detailed comments below.

Details:

Page 2, Line 9: I don't understand "substitute observed by modeled fluxes". Substitute modeled fluxes for observed fluxes?

Page 2, Line 11: Was the random error added to observed or modeled "tower" fluxes? If this was added to the observed fluxes, why? There is already random error in the measurements.

Page 2, Lines 14-15: "This difference…" isn't clear. Do the large biases exist with respect to 5PM, or the other models? And how does a large bias cause a long temporal autocorrelation?

Page 3, line 10. I don't understand the term "regularized." Can this be defined? It is used in more than one place, and I don't recognize what concept is being communicated.

Page 4, lines 4-6. Coarser scale inversions may not explicitly utilize correlation lengths, but they are implicitly imposing a large correlation length (which may be entirely inappropriate). Please clarify.

Page 4, lines 13-14. I don't understand this description. How can you derive a spatial correlation in the prior flux error from a coarse resolution inversion?

Page 4, lines 22-23. This is not correct. The pattern of fluxes was not used to evaluate the spatial correlation length. Lauvaux et al, 2012 tested the spatial correlation length scale by cross-validation of the posterior CO2 mole fractions. CO2 observations were reserved from the inversion, and the correlation length that provided the best fit to the reserved CO2 data was identified as the best choice.

Page 5, lines 1-4. This sentence is unintelligible.

Page 4, lines 12-13. Make it clear that flux measurement sensitivities are areas, not lengths. You are describing dimensions of an area measurement. This is not clear as written. I also recommend that you note that the resolution of an inversion system is not necessarily the same as the true resolution of an inverse flux estimate.

Page 4, line 15. There are many, many studies of flux upscaling with towers and spatial databases. The paper would benefit from a somewhat expanded review of this literature.

Page 7, line 25. What does "across entire daily course" mean?

Page 9, lines 13-14. Why is the year of EC data used to optimize respiration parameters singled out? What about the other parameters and years of EC data?

Page 10, line 3. English. I suggest, "it is the only model with spatial resolution (10 km) comparable to. . ."

Page 10, lines 6-8. Number of flights is repeated. Resolution of aircraft flight data contradicts earlier text that stated 2 km resolution.

Page 10, lines 15-16. Do you neglect the impact of the observation errors? Or is it just that you cannot separate these errors from errors in the model? The observation errors are already part of the observations.

Page 11, equations (2) and (3) and the "nugget" effect: Mathematically, when k=0, equation (2) = 1. What is this nugget effect? How can it exist for k=0? The numerator and denominator are identical when k=0.

Page 11, line 19. What does "distributed along the entire daily course" mean? And how can aircraft data span 36 days? Are the times in between the aircraft flights neglected?

Page 12, equation (4). Again, I don't see how a nugget is relevant with the normalized correlation equation (2).

Page 12-13. ". . .assuming that the involved prior errors for each model are identical in a sense that they share the same statistics and (are?) not correlated." What does this mean?

Page 13, line 8. Richardson et al., 2008, only deals with random sampling error. More recent papers add gap filling errors and friction velocity screening errors to the observational error assessment.

There are many years of flux tower data. Only 2007 observations are used. Why? Could the results be strengthened with additional years of observations?

Page 14, lines 3-6. Your equations state (but do not define explicitly in the equations) that you are going to evaluate model-data, and model-model differences. The results, however, begin by stating standard deviations in observed NEE. I cannot tell what you have computed. Please clarify the methods and the associated results so that these values can be interpreted by the reader. I think you are presenting time-constant, spatially varying standard deviations across sites, then summed over all times. But the paper should not make me guess what you are computing here, and the methods say nothing about this computation.

Page 14, line 11. Why is this “in line” with the spatial standard deviation? Model-data differences at a given site are not necessarily related in any way to differences in observed fluxes across sites.

Page 14, lines 13-14. What are “site specific correlations” that are presented as a single value? If these are site specific correlations, what site is being presented? Again, the methods are not sufficiently clear.

Page 14, lines 15-18. This sentence’s English needs work. Further, the statistical comparison (see above comment) isn’t clear. Finally, the model-data difference, if I understand it, also includes a temporal component since it is summed over time. As best I can tell, the same data go into both calculations, so I don’t understand how the authors can draw this fuzzy conclusion about ability to simulate temporal variability better than spatial variability. Clear, targeted work on this topic has been published elsewhere but has been neglected in the introduction to this paper. This is also not clearly a main focus of this paper. I would suggest that you either expand the paper to address this topic properly, or delete this discussion.

Figure 2 is not clear. Is each point a different site? How is this figure related (or not) to the “site specific correlations” noted on lines 13-14 of Page 14?

Figure 3 has the same problem as Figure 2. Please specify what distribution (sites?) are being illustrated by the box and whisker plots. In addition, the sign of the bias is never defined. Finally, it would be useful to provide a conversion to gC m-2 yr-1.

Figure 4 has too many lines of similar color and tiny size to be read clearly.

Page 14, line 28. All site? Flux site? Sub-site? The terminology surrounding Figure 4 needs to be cleaned up. I cannot tell what is being plotted. The text appears to contradict the figure caption, and the terminology changes enough to be quite confusing.

Page 15, lines 2-10. Do not describe the figures in the text. The figures present these results. Discuss the significance of the figures in the text.

The exponential fit appears to be a poor choice. Based on Figure 4, most of the site autocorrelations are below the fit lines at lag times of 30-70 days. Thus the model used to fit these curves is quite biased. It is also more biased for some model-data comparisons than for others. This makes the comparison of decay times misleading, to the point perhaps of being meaningless. I would not publish results based on such a biased approximation of the site-based results. I think this functional fit must be changed. At minimum, the quality of fit must be made very clear, and the logic for keeping this function, despite its relatively poor fit, articulated.

Page 15, lines 13-15. If the correlation at zero lag is not 1, either your calculations are flawed or equation (2) does not represent your methods.

Page 15. A root mean square error of a functional fit to an autocorrelation curve is not very meaningful. Evaluating the quality of multiple potential fits to find the best fit to the data would be meaningful and improve the analysis.

Page 16, line 5. “not applicable”

Page 16, lines 9-11. The sites were screened because the bias was greater than 2.5 umol m-2 s-1. But now the text says that the bias for these sites was not greater than 2.5 umol m-2 s-1, just “larger than average.” I am confused. Please clarify.

There is no functional fit to the aircraft data (figure 5). Given the poor quality of the fits in Figure 4, and I am not convinced that there really is a difference between the two data sets. I would be much more convinced by a comparison of the mean or median values, binned by lag time.

Page 16, line 29. What is the purpose of the root mean square error?

Figure 6. The exponential fit is consistently below the median at distances of 200-400km. I would argue that your correlation computation shows consistently positive values out to approximately 200-400km, which is consistent with Hilton et al., (2013), who performed a similar calculation for North American flux tower sites and model-

data differences using VPRM. Again, your exponential fit appears to be biased. I do not believe that quoting the results of a biased fit is sound.

Hilton et al., (2013), published a paper using very similar methods using North American flux towers, a much longer time series of data, and more evaluation of the robustness of the resulting length scales. This was published in Biogeosciences. The results contradict the results presented here in that Hilton et al (2013) found significantly larger length scales for their variogram fits. The similarity is so great that the Hilton et al (2013) paper really should be cited and evaluated with respect to these results.

Figure 7. How is the confidence interval computed? This is not a simple case of computing the standard deviation of a Gaussian. Please explain the methods. I am still dubious of the value of the exponential fit, but in any case the methodology for the confidence interval estimate must be explained.

Figure 7 points out something that is lacking from the primary results reported in the abstract – uncertainty bounds. These results suggest that the uncertainties in the computed correlation lengths are very large. This should be reported in the abstract.

The standard spatial statistical method for Figures 6 and 7 would be a variogram. Why have the authors chosen a different approach?

Page 18, lines 5-6. English. 'it difficult to determine... where the asymptote lies" perhaps?

Figure 8 illustrates again how poorly the exponential model fits the data. And the exponential model is not shown on the figure, which is inconsistent with figures 6 and 7. D=35km with a 95% confidence interval of 26-46 km is clearly biased given that none of the aircraft data reaches 1/e of the zero correlation anywhere within that range. The exponential model is poor and should not be used, or only with serious caveats about the biased nature of the fit.

Some of the colors in Figure 9 are nearly indistinguishable when used to plot very thin lines. Please either reduce the content of the figure or find a way to distinguish the different model-model pairs more clearly.

Figure 9: Why are the individual points not shown, as for the model-data comparison? I can understand reducing the information shown, but I am concerned about the quality of the exponential fit, and it is impossible to evaluate from this figure.

Figure 10. Again, the blue-green-purple lines are difficult to distinguish (all dark), and the red-green lines will be indistinguishable for those who are red-green colorblind.

Discussion and conclusions. The results are compared only to Chevallier's publications, and that comparison is limited to the temporal autocorrelation. There is insufficient effort to put these results into the context of prior work on this topic.

Page 21, lines 26-28. The observational errors are in your calculations. It is not neglected. It cannot be isolated and removed, but it is not neglected.

End of page 24, beginning of page 25. This is an interesting discussion. Again, the methods used for this interesting calculation are opaque. Please explain. Propagating these error estimates is not trivial. How was this done?

Same section: Comparing your aggregated error estimate to the range of existing continental-scale flux estimates (e.g. Peylin et al, 2013) would be more useful than the very limited analysis presented.

Same section: I agree that this analysis (pending evaluation of the unknown methods) would strongly suggest that the total continental-scale, annual flux errors are seriously underestimated, and I agree that this is an important issue to point out. This should be part of the abstract, as it leads to significant uncertainty regarding the validity of the correlation lengths. The current abstract suggests no such uncertainty regarding the conceptual model promoted in this paper.

Page 25, paragraph starting with "Exponentially decaying...". This paragraph begins to give reasons for using an exponential model for the correlations. Some notes below about this discussion:

1) This discussion belongs earlier in the paper. It presents the logic for using this fit.

2) The reasons given are entirely reasons of simplicity and convenience, not accuracy of the fit. I would suggest that the best job of describing the correlations should be the primary goal of this paper. Considering how to simplify these correlation functions to make them convenient is another problem. I have already noted that I believe the exponential fit is so poor that it is significantly misleading. The analysis would be improved by evaluating different fits and finding what fits are best given the data.

3) Lines 20-25. Computational simplicity is not a good reason to use the wrong correlation length. This is a disturbing discussion.

4) Equation (7) does not exist in the paper.

Page 26, lines 3-4. English needs work. And I'm not sure what is meant by "for the short spatial scales." And what studies have already used the correlation lengths derived from this study? I'm not sure that the 'future work' needs to be part of this paper.

Interactive comment on Biogeosciences Discuss., 12, 9393, 2015.