Biogeosciences Discuss., 12, C6956–C6969, 2015
www.biogeosciences-discuss.net/12/C6956/2015/

Biogeosciences
Discussions
Open Access

Interactive comment on "The ability of atmospheric data to resolve discrepancies in wetland methane estimates over North America" *by* S. M. Miller et al.

S. M. Miller et al.

scot.m.miller@gmail.com

Received and published: 23 October 2015

We would like to thank the reviewer for suggestions and comments on the manuscript. The reviewer's detailed suggestions have been very helpful in improving the manuscript. Below, we have included the reviewers comments (in bold) along with our reply and the associated changes/updates to the manuscript.

1 Summary

- **In general, a new method of statistical analysis (BIC for CH4 wetland flux model selection) needs to be tested fully to understand the abilities and limitations to give confidence to the results.**

  The statistical methods and associated applications in this paper are by no means new. Model selection based upon the BIC has been used widely across statistics and has been used in numerous top-down studies of CO_2 and CH_4 to select a flux model. Schwarz (1978) originally developed the Bayesian Information Criterion (BIC), and BIC-based model selection has subsequently been used widely in statistical modeling (Ramsey and Schafer, 2012). In more recent years, several studies have used BIC-based model selection to explain patterns in eddy-covariance CO_2 flux measurements (Mueller et al., 2010; Yadav et al., 2010). A number of atmospheric inverse modeling studies have also employed BIC-based model selection to choose bottom-up models or environmental datasets for the prior flux estimate; these inverse modeling studies have applied BIC-based model selection to both CO_2 (Gourdji et al., 2012) and CH_4 (Miller et al., 2013, 2014a) flux problems. In addition to the aforementioned applications, BIC-based model selection has also been used to investigate the detectability of atmospheric patterns from surface sources (Fang et al., 2014; Shiga et al., 2014; Fang and Michalak, 2015). The model selection methods outlined in those three papers are identical to those used in the present manuscript.

  In the revised manuscript, we have included more references to these sources to bolster the methods outlined in Section 2.2.

- **Examine the uncertainties for each model components, such as sensitivity of the methodology to transport errors, flux errors, background concentration estimates, other assumptions/approximations and their interactions, as well as the stability or robustness of the analysis. If a thorough analysis**

is not done, at least a full discussion of the potential problems is needed to put the preliminary results in the proper perspective.

The analysis in this paper is based upon several existing inverse modeling studies, studies that use the same WRF-STILT model simulations, background concentration estimate, covariance matrix parameters, etc. (Miller et al., 2013, 2014a,b). Those studies discuss potential modeling and measurement errors in great detail. For example, Miller et al. (2013) explore uncertainties in the estimated methane boundary condition, uncertainties in the estimated covariance matrix parameters (the parameters that define $\mathbf{\Psi}$), uncertainties due to atmospheric transport in WRF-STILT, uncertainties due to geological CH_4 sources, and uncertainties in the attribution of CH_4 to individual sources. Miller et al. (2014a) discuss uncertainties due to the nested meteorology domains in WRF-STILT, uncertainties in the methane boundary condition, uncertainties due to the sparsity of the CH_4 observation network, uncertainties in atmospheric transport estimated by WRF-STILT, and uncertainties in the covariance matrix parameters. Finally, Miller et al. (2014b) discuss uncertainties in the flux estimate due to the assumptions made by the statistical modeling framework.

We have added a new section to the supplement (S7) that summarizes and highlights the discussion in these previous papers.

2 General comments

- **In Section 3: 4 of the 7 models were selected from WETCHIMP for the synthetic data experiment to evaluate the BIC method. Presumably the authors expected (without sufficient reasons) the 3 remaining models to be ‘unacceptable’, including LPJ-Bern. Yet BIC results in section 4.1 indicate LPJ-Bern is selected most often when using real observations, bringing**

into question the value and correctness of this BIC method (in its current form). This suggests the rejection of LPJ-Bern could be a problematic assumption in section 3 and/or the BIC methodology is not working. Possibly other assumptions and approximations (such as uncertainty estimates, model transport errors, etc.) are incorrect, the whole new method needs to be evaluated much more thoroughly as noted above. In general, much stronger justifications are needed for model selection. The BIC analysis should be done for all WETCHIMP models in the synthetic data experiment. If the range of flux model variations is too large, then what is the range of applicability for the BIC method?

We have revised the manuscript to include all seven WETCHIMP models in the synthetic data experiment.

In the original manuscript, we did not include three models in the synthetic experiments because those models had an anomalously large magnitude relative to the other model estimates. The larger the wetland flux, the more likely that the observation network can detect a methane pattern from wetlands. Therefore, if we conduct the synthetic data experiment using a flux model that has an anomalously large magnitude (see Fig. 4), we would concomitantly obtain anomalously optimistic results. Thus, we had not included these three models in the synthetic data experiments to avoid biasing the experiment results.

The analysis in Section 4.1 examines a completely different question, one that is independent of the flux model’s magnitude. Rather, that experiment examines which models have spatial distributions that are consistent with the data. That analysis disentangles the spatial distribution of the fluxes from other confounding questions of magnitude or seasonality. Unlike the first synthetic data experiment, the analysis in Section 4.1 is insensitive to the magnitude of the flux models.

We have clarified the experimental design in Sections 2.3 – 2.4 to prevent any confusion. In the revised manuscript, we have also included all seven

WETCHIMP models in the synthetic data experiments. We have scaled the magnitude of those models to match Pickett-Heaps et al. (2011), Miller et al. (2014a), and Wecht et al. (2014); this step ensures that the BIC results in Section 3 will not be biased by anomalously large WETCHIMP flux models, as explained above.

- **The authors stated 'By contrast, the observation network is largely insensitive to spatial variability in wetland fluxes across the US; in most instances, the model selection framework favors a spatially-constant model over a wetland model for the two US regions.' (page 9352, lines 9-12). Was 'a spatially-constant model' included in the work and the BIC model 'favors' or selected it over the other models? The comparison to 'a spatially-constant model' in general needs to be explained and documented more clearly. If the BIC method actually selects the 'spatially-constant model' (a clearly wrong model), maybe this is an indication of the lack of ability of the method.**

We have added additional explanation to the manuscript to clarify this concept. The 'spatially-constant model' is analogous to an intercept term in a regression. It is standard practice to include an intercept in regression modeling; this intercept is always included in the model, irrespective of the availability or suitability of other explanatory variables. In spatial statistics, that intercept term is a spatial constant, hence the phrase 'spatially-constant model.'

The BIC-based model selection framework examines every possible combination of explanatory variables. The BIC-based framework will then score each combination of explanatory variables based upon how well each fits the observations and based upon the number of variables in each combination; combinations with a large number of variables receive a larger penalty (Eq. 1). Each candidate combination always includes an intercept term. It is plausible that none of the variables explain substantial variability in the observations. In such case, the best-scoring model may be one that only includes an intercept (the most basic

possible model). The intercept term is almost certainly not the "true", correct model; the intercept does not represent the unknown methane flux processes that occur within soils. Similarly, the intercept almost never explains all variability in the observations. Rather, the model selection framework may choose the intercept-only model if other candidate variables have limited ability to explain the atmospheric data.

A hypothetical example helps illustrate this concept. In theory, we could use model selection to explore whether observations at the LEF tower in Wisconsin can detect a methane pattern from wetland fluxes in Siberia. We could examine this question using a synthetic data experiment like the first experiment in Section 2.3. Any variability in the observations at LEF would be unrelated to methane fluxes in Siberia. Therefore, a model selection framework would not choose a Siberian flux model as an explanatory variable. That result is not necessarily a judgement on the accuracy of the Siberian flux model. Similarly, that result does not indicate that the intercept is a 'perfect model.' Rather, that result simply indicates that methane observations at LEF cannot be explained by fluxes in Siberia.

In addition, it is important to note that BIC-based analysis in Sections 3 and 4.1 address different questions, and we do not expect any one model to be selected by the model selection framework in all cases. The first synthetic data experiment explores whether the atmospheric observation network can detect an atmospheric pattern from wetland sources. The second experiment asks a more detailed question – whether the network is sensitive to the location of wetland fluxes. The network could select both the intercept and a wetland model in the first experiment but may only select the intercept (no wetland model) in the second experiment. In that case, the network can detect an atmospheric CH_4 pattern from wetlands but cannot pinpoint the geographic origin of that atmospheric pattern.

For additional discussion, refer to Shiga et al. (2014); the authors used model selection to explore the detectability of anthropogenic CO_2 fluxes across North America. Our experimental design parallels that study.

- **The 3 subsections, 4.1, 4.2, 4.3 seem to be unrelated and somewhat conflict with each other. Sections 4.2 and 4.3 did not make use of the information from 4.1 that LPJ-Bern and SDGVM performed best in the BIC analysis. In fact, the best model from 4.1, LPJ-Bern performed poorly in the flux magnitude and seasonal maximum comparison in section 4.2. There is no explanation on this discrepancy. This could be an indication of the problem in the BIC analysis in selecting a poorly performing flux model.**

  Subsections 4.1, 4.2, and 4.3 analyze the spatial distribution, magnitude, and seasonality of each WETCHIMP model, respectively. Those sections examine each of these qualities individually; each section should be unrelated to the others. A model that best matches the magnitude of the observations will not necessarily have a spatial distribution that best matches the observations (and vice versa). Similarly, a model that best matches the magnitude of the observations at annual scale will not necessarily have a seasonality that best matches the observations (and vice versa). For example, the spatial distribution of LPJ-Bern compares favorably against the data, but its magnitude must be scaled downward to match the observations. This result is neither a discrepancy in the methodology nor a problem in the model selection framework.

- **Section 4.3 is a comparison of all 7 flux models to the results of another inversion model, Miller et al. 2014. Since inversion model results are highly uncertain, this simple model to model type comparison has little useful information unless the authors can show independently that the Miller et al. 2014 results are good and can serve as a benchmark for comparison. At the minimum, the WETCHIMP fluxes should be compared to a variety of inversion model results to see the uncertainty possible for the inversion model**

results. The authors should note in the manuscript that section 4.3 is only a comparison of models, which is different than comparison to 'available data' (page 9345, lines 6-9, see above).

  We disagree with the premise that an observation-based estimate like this one provides 'little useful information.' Fig. 5 displays the uncertainties in the inverse modeling estimate. Refer to the supplement and Miller et al. (2014a) for a more detailed discussion of the uncertainties.

  There are few regional-scale inversions that estimate the seasonal cycle of wetland fluxes from the HBL using in situ or satellite methane observations. Miller et al. (2014a) is the only study that we are aware of. Three other top-down methane studies, Pickett-Heaps et al. (2011), Wecht et al. (2014), and Turner et al. (2015), examine methane fluxes from the HBL. However, those studies use the seasonality from a bottom-up model and do not estimate the seasonal cycle independently using atmospheric or satellite data. For reference, Melton et al. (2013) provides a list of existing methane inversion studies that are relevant to wetland fluxes.

3 Specific comments

- **Page 9345, line 1: 'biogeochemical models leveraged all available data', what is this vague 'all available data' referring to?**

  We have modified this line to be more specific.

- **Page 9346: "background" concentration – the methane concentration of air entering the North American regional domain? Do all STILT particles always leave the North American regional domain during the model simulations? Provide more information on the estimation of the background**

concentration and the whole model settings to enable other scientists to check and/or compare results.

Refer to bullet point #2 above and to the reply to Reviewer #1. Both discuss modeling uncertainties, including uncertainties in the estimated background concentrations.

- **Page 9346, lines 21-24: repeating lines 23-27 on page 9344.**

  We have condensed the text to remove this repetition.

- **Page 9347, line 7: change 'first term in Eq. (1)' to 'first two terms in Eq. (1)'.**

  We have updated the text accordingly.

- **Page 9348, line 29: 'We also include a spatial constant or intercept term in X'. What is the physical or numerical significance of the 'spatial constant or intercept term'. How do the results compared with or without the 'spatial constant or intercept term'?**

  Refer to the earlier discussion on the intercept for an explanation of this point.

- **Page 9353, line 26: change 'GEIMS' to 'GIEMS'.**

  We have changed the text accordingly.

- **Page 9354, lines 16-20: 'The estimated contribution of anthropogenic emissions from EDGAR v4.2FT2010 is added to this background (in red). Note that the estimated scaling factors for EDGAR (Sect. 2.4) are 1.7±0.3 at Chibougamau, 5.6±0.5 at East Trout Lake, 2.4±0.3 at Fraserdale, and 2.5±0.3 at Park Falls.' Explain the meaning of the different scaling factors and how realistic are they (up to 5.6x)? What are the spatial regions these scaling factors are applied to?**

  In the revised manuscript, we do not scale the EDGAR inventory. Many atmospheric CH_4 observation sites near wetlands are also located far from large anthropogenic emissions. As a result, any effort to scale the EDGAR inventory at these sites could be error-prone. Instead, we present the inventory as is.

- **Page 9364: left figure contains wrong information 'Observation site (Fig. 4)'. Units in the right figure conflicts with caption.**

  The figure should be correct as is.

- **Page 9368: Label each curve in Fig. 5 as in Fig. 4.**

  We have updated the figure accordingly.

4 Supplement

- **Page 2, lines 56-57: 'In this equation, zsynthetic (n x 1) represents the synthetic observations generated for an observation site'. What is the number of sites (k) and does n vary for each of the k sites?**

  We have added more detail to the supplement on the observation sites.

  The study includes 15 total tower-based observation sites and 17 regular aircraft-based observation sites. Refer to Miller et al. (2013), Miller et al. (2014a), or the NOAA program websites (http://www.esrl.noaa.gov/gmd/ccgg/insitu/, http://www.esrl.noaa.gov/gmd/ccgg/aircraft/) for more detail individual observation sites. The observations at each tower site are either daily flasks (most US sites) or daily averages of continuous data (Canadian sites). A few sites, including the tower sites in Oklahoma and in Maine only have weekly flask measurements. The frequency of the regular aircraft observations varies depending upon the site (see http://www.esrl.noaa.gov/gmd/ccgg/aircraft/).

- **Page 2, lines 71-72: 'The other WETCHIMP models, in contrast, predict much higher fluxes (Fig. 4)'. Fig. 4 only showed concentrations or mole fractions, change 'much higher fluxes' to 'much higher concentrations'.**

  We have changed the text accordingly.

- **Page 2, 80-82: 'The magnitude and spatial/temporal structure of these errors were estimated in Miller et al. (2013) for the US and Miller et al. (2014b) for Canada.' The referenced works were for different prior fluxes. Authors need to show why new error estimates are not needed.**

  There are two covariance matrices used in the paper (Fig. S3). The first matrix, $\mathbf{R}$ describes model-data mismatch errors – errors in the measurements, atmospheric transport, and errors due to the finite spatial or temporal resolution of WRF-STILT. These errors should be invariant to the choice of prior model.

  The second matrix, $\mathbf{Q}$, describes the residuals between the true fluxes (denoted s) and prior model estimate ($\mathbf{X}\beta$). The diagonal terms quantify the variance of these residuals and the off-diagonal terms quantify the spatial and temporal covariances in these residuals. The variances and covariances in $\mathbf{Q}$ can change depending upon the choice of prior model.

  In the synthetic data studies, we construct a statistical model that is representative of a prototypical real data inverse model. Similarly, we want to use values for $\mathbf{Q}$ and $\mathbf{R}$ that are representative of what one would likely encounter in a real-data setup. Miller et al. (2013) and Miller et al. (2014a) constructed real data inverse models over the US and Canada, respectively. Those studies used a model selection framework to find prior models that show optimal fit against available observations. In each study, the authors then estimated the elements of $\mathbf{Q}$ and $\mathbf{R}$ using that prior model. The resulting estimates of $\mathbf{Q}$ are representative of prior models that shows optimal agreement with atmospheric observations. We use these values of $\mathbf{Q}$ and $\mathbf{R}$ in the synthetic data studies. We also use these pro-

totypical variances and covariances to generate random errors in the synthetic data (see Section 2.3).

  We have revised the treatment of the covariance matrix parameters in the real data experiments (Section 2.4). Previously, we had used covariance matrix parameters from Miller et al. (2013) and Miller et al. (2014a) in these experiments. In the revised manuscript, we re-estimate the covariance matrix parameters for each prior model. We estimate these parameters using Restricted Maximum Likelihood (RML) (Corbeil and Searle, 1976; Kitanidis, 1995; Michalak et al., 2004; Gourdji et al., 2012), the same procedure used in Miller et al. (2013) and Miller et al. (2014a).

  We have added an explanation of these concepts in the supplement.

- **Page 3: Correct the difference in units in the Figure and caption.**

  The figure should be correct as is. We could not find any difference or discrepancy in the units.

References

Corbeil, R. R. and Searle, S. R.: Restricted maximum likelihood (REML) estimation of variance components in the mixed model, Technometrics, 18, pp. 31–38, 1976.

Fang, Y. and Michalak, A. M.: Atmospheric observations inform CO_2 flux responses to enviroclimatic drivers, Global Biogeochem. Cy., p. 2014GB005034, doi:10.1002/2014GB005034, 2015.

Fang, Y., Michalak, A. M., Shiga, Y. P., and Yadav, V.: Using atmospheric observations to evaluate the spatiotemporal variability of CO_2 fluxes simulated by terrestrial biospheric models, Biogeosciences, 11, 6985–6997, doi:10.5194/bg-11-6985-2014, 2014.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluszkiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C., and Michalak, A. M.: North American CO_2 exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, Biogeosciences, 9, 457–475, doi:10.5194/bg-9-457-2012, 2012.

Kitanidis, P.: Quasi-linear geostatistical theory for inversing, Water Resour. Res., 31, 2411–2419, doi:10.1029/95WR01945, 1995.

Melton, J. R., Wania, R., Hodson, E. L., Poulter, B., Ringeval, B., Spahni, R., Bohn, T., Avis, C. A., Beerling, D. J., Chen, G., Eliseev, A. V., Denisov, S. N., Hopcroft, P. O., Lettenmaier, D. P., Riley, W. J., Singarayer, J. S., Subin, Z. M., Tian, H., Zürcher, S., Brovkin, V., van Bodegom, P. M., Kleinen, T., Yu, Z. C., and Kaplan, J. O.: Present state of global wetland extent and wetland methane modelling: conclusions from a model inter-comparison project (WETCHIMP), Biogeosciences, 10, 753–788, doi:10.5194/bg-10-753-2013, 2013.

Michalak, A., Bruhwiler, L., and Tans, P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res.-Atmos., 109, doi:10.1029/2003JD004422, 2004.

Miller, S. M., Wofsy, S. C., Michalak, A. M., Kort, E. A., Andrews, A. E., Biraud, S. C., Dlugokencky, E. J., Eluszkiewicz, J., Fischer, M. L., Janssens-Maenhout, G., Miller, B. R., Miller, J. B., Montzka, S. A., Nehrkorn, T., and Sweeney, C.: Anthropogenic emissions of methane in the United States, P. Natl. Acad. Sci. USA, 110, 20 018–20 022, doi:10.1073/pnas.1314392110, 2013.

Miller, S. M., Worthy, D. E. J., Michalak, A. M., Wofsy, S. C., Kort, E. A., Havice, T. C., Andrews, A. E., Dlugokencky, E. J., Kaplan, J. O., Levi, P. J., Tian, H., and Zhang, B.: Observational constraints on the distribution, seasonality, and environmental predictors of North American boreal methane emissions, Global Biogeochem. Cy., 28, 146–160, doi:10.1002/2013GB004580, 2014a.

Miller, S. M., Michalak, A. M., and Levi, P. J.: Atmospheric inverse modeling with known physical bounds: an example from trace gas emissions, Geoscientific Model Development, 7, 303–315, doi:10.5194/gmd-7-303-2014, 2014b.

Mueller, K. L., Yadav, V., Curtis, P. S., Vogel, C., and Michalak, A. M.: Attributing the variability of eddy-covariance CO_2 flux measurements across temporal scales using geostatistical regression for a mixed northern hardwood forest, Global Biogeochem. Cy., 24, doi:10.1029/2009GB003642, 2010.

Pickett-Heaps, C. A., Jacob, D. J., Wecht, K. J., Kort, E. A., Wofsy, S. C., Diskin, G. S., Worthy, D. E. J., Kaplan, J. O., Bey, I., and Drevet, J.: Magnitude and seasonality of wetland methane emissions from the Hudson Bay Lowlands (Canada), Atmos. Chem. Phys., 11, 3773–3779, doi:10.5194/acp-11-3773-2011, 2011.

Ramsey, F. and Schafer, D.: The Statistical Sleuth: A Course in Methods of Data Analysis, Cengage Learning, Boston, 3 edn., 2012.

Schwarz, G.: Estimating the dimension of a model, Ann. Stat., 6, 461–464, doi:10.1214/aos/1176344136, 1978.

Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L., and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO_2 measurements, Geophys. Res. Lett., 41, 4381–4388, doi:10.1002/2014GL059684, 2014.

Turner, A. J., Jacob, D. J., Wecht, K. J., Maasakkers, J. D., Lundgren, E., Andrews, A. E., Biraud, S. C., Boesch, H., Bowman, K. W., Deutscher, N. M., Dubey, M. K., Griffith, D. W. T., Hase, F., Kuze, A., Notholt, J., Ohyama, H., Parker, R., Payne, V. H., Sussmann, R., Sweeney, C., Velazco, V. A., Warneke, T., Wennberg, P. O., and Wunch, D.: Estimating global and North American methane emissions with high spatial resolution using GOSAT satellite data, Atmos. Chem. Phys., 15, 7049–7069, doi:10.5194/acp-15-7049-2015, 2015.

Wecht, K. J., Jacob, D. J., Frankenberg, C., Jiang, Z., and Blake, D. R.: Mapping of North American methane emissions with high spatial resolution by inversion of SCIAMACHY satellite data, J. Geophys. Res.-Atmos., 119, 7741–7756, doi:10.1002/2014JD021551, 2014.

Yadav, V., Mueller, K. L., Dragoni, D., and Michalak, A. M.: A geostatistical synthesis study of factors affecting gross primary productivity in various ecosystems of North America, Biogeosciences, 7, 2655–2671, doi:10.5194/bg-7-2655-2010, 2010.

Interactive comment on Biogeosciences Discuss., 12, 9341, 2015.