

Interactive comment on “Towards better error statistics for atmospheric inversions of methane surface fluxes” by A. Berchet et al.

A. Berchet et al.

antoine.berchet@lsce.ipsl.fr

Received and published: 24 May 2013

We thank the referee for his/her comments that will help improve the presentation of our study. We noted the worries of the referee about the objective evaluation of our case study on real data. This evaluation is left for the further steps of our work. By then, we agree with the request of the referee to be more cautious in some formulations.

We have addressed below in detail all the comments. They are copied hereafter in italics with our detailed answers inserted in standard font where appropriate.

This is a very interesting piece of work, addressing a commonly recognized weakness in the use of inverse modeling for estimating sources and sinks of greenhouse gases.

C2727

It takes more than average effort to understand what was done, but it is worth the investment. Except for some inaccurate formulations, this has more to do with the complexity of the problem than the description of it by the authors. Below some suggestions are made to further improve the readability.

The most important concern I have is that while the proposed methods are interesting to read about, it is difficult to grasp where the joint information provided by the a priori fluxes, the transport model and the measurements end up in the solution. It is interesting to see the techniques applied to a real problem, but it makes it very hard to objectively evaluate their performance. After all we have very limited means of verifying the results of inversions.

The application to a real case indeed makes it difficult to separate the contribution of each piece of information. But the methods have been well described in their theoretical aspects previously and have been tested on toy models and on real cases in other fields (e.g., Desroziers et al. 2001; Reichle et al., 2008; Winiarek et al., 2010). Consequently, we decided to apply them directly to a real case in our field in order to evaluate their efficiency and feasibility. The computed matrices seem very consistent to what one could expect according to expert knowledge. A further evaluation of the errors made on the computed matrices should be carried out in further steps of our work (see the last answer of the section General Comments).

As explained below, some worrying results, in my opinion, should have received more attention. Else, the study would benefit from an extended discussion concerning the recommended use in inversions with large state vectors. With these topics addressed, the study should be acceptable for publication.

GENERAL COMMENTS

Differences between the techniques for estimating B and R are mostly discussed in terms of uncertainties rather than the impact of the actual solution. In the end only a single method is selected for discussing the estimated fluxes, where it would have been interesting to have a comparison between the three techniques. It would allow evaluating the relevance of the difference between the methodologies, which is important for application to larger inverse problems (what are the returns of larger computational investments?).

The inversions are very sensitive to the assigned error statistics. In part 4.1, we give some illustrations of the results from each methods. ML and DS methods compute some features not physically consistent: strongly negative fluxes or not quantified posterior correlations between regions contiguous to the lateral boundaries. Therefore, though the performance of the methods cannot be estimated through a comparison with independent data, we consider that the matrices that are statistically and algebraically the most consistent give the better results. The matrices given by ND methods are the best in that sense. In particular, in our problems, diagonal matrices are known to miss critical patterns of the error statistics. We then give inversion results only for ND matrices (as illustrations) and chose to focus mainly on the computed uncertainties rather than on the fluxes themselves.

C2729

Of course, the solution can only be judged in relation to its (posterior) uncertainty. There I was much surprised to read that the uncertainties in Figure 7 do not exceed 1%. Given that the estimated uncertainties are supposed to account for contributions from transport and representation uncertainty this is a very surprising outcome. It is hard to imagine that it is really possible to estimate the emissions at this level of accuracy.

The uncertainties on the optimized fluxes are crucial information to evaluate the results of an inversion. In Figure 7, getting uncertainties smaller than 1% can appear unrealistic. However, one must keep in mind that the regions of aggregation are very large. Because of the aggregation, the uncertainty on the total emissions in a big region is smaller than the uncertainties on the emissions in single pixels. Moreover, emissions in Europe during the period of interest are mostly anthropogenic and well documented. We can then expect that the uncertainties on emissions in single pixels are low, which leads to even lower uncertainties on the big aggregated regions.

In further steps of our work, a proper evaluation of the errors made during the computation of the uncertainty statistics should give some additional materials on the posterior errors on fluxes.

One necessary test, in my opinion, is if the other two methods yield solutions that are really within that uncertainty range. If not then how consistent are the posterior solutions and what does it say about the realism of the estimated R and Bs?

The high sensitivity of the results of an inversion to the statistics of errors is a known fact. Regarding the inverted fluxes, the three methods are not always consistent with

C2730

each other. The three methods do not rely on the same mathematical assumptions. In a sense, they do not describe the same world.

Hence, even if we use the same real data for the three methods, it is expected that the results are not statistically consistent with each other.

In the discussion some loose statements are made comparing the inverse methods, which estimate prior R and B uncertainties versus those that rely on expert judgment. The claim is made that the former approach gives better results, without providing any evidence. In my opinion such a comparison is essential to assess how important it is to invest in the optimization of R and B. The question is also how to measure the relative performances, given that we don't know the true solution. Without further explanation and demonstration it is not possible to judge value of these statements.

We tested our inversion with matrices built on expert knowledge, in particular diagonal matrices. The results were typically similar to those inferred with the matrices from ML and DS algorithms, especially in the way the lateral boundary conditions are dealt with (see Sect. 4.1). We then chose not to develop this point further.

Regarding the evaluation of the exact performance of the algorithms, this does not seem feasible in a real framework: an OSSE would be needed for the further steps of our work (see the last answer before the specific comments).

Currently, the ND method is favored over the other methods. To apply this method to large inverse problems, however, seems impossible. The question remains if the

C2731

cheaper methods are still preferable over the expert judgment approach.

Regarding its statistical properties, the ND method is favored over the others. The ND method is indeed not feasible in a framework with large dimensions. The ML method computation time also depends on the size of the state vector (at least linearly) and on the number of observation ($> O(n_{obs}^3)$). So it seems difficult to apply it to a problem with large state vectors.

One should see our work as a way of computing guidelines to build R and B in high dimension problems by projecting them in sub-spaces of lower dimensions.

Further assessment is needed of the signals in the data that determine the estimated uncertainties. The signals that have been identified are PBL height and the passage of frontal systems. They are interpreted as a confirmation that the methods do what they are supposed to do, which is reproduce known sources of transport model uncertainty. The question, however, is whether or not the results would have been any different for a perfect transport model. Another possibility is that the algorithm has the tendency to reduce the weight of any conditions that cause high variability (of which PBL dynamics and frontal passages are important examples). Part of the variability is signal, which the inversion is supposed to translate into fluxes. Reducing the weight of signal will therefore limit the performance of the inversion.

The algorithm reduces the weight of any situation that the model does not reproduce well. Transport models are known to miss high variability features. Enhanced errors in situations of strong temporal gradients are then very consistent with the physics of the problem.

C2732

It will probably require dedicated OSSEs to determine how successful the methods really are in separating flux uncertainty from model uncertainty. Until then, I believe more careful formulations are needed.

We agree that a dedicated OSSE would be a very natural continuation to our work. The transport model could be used to build a truth from arbitrary known fluxes with known uncertainties. We could then determine some true observations. A Monte-Carlo approach could be used to estimate the uncertainties on the computation of the error statistics. More precisely, one could perturb the observations by a random noise inferred from the known uncertainties of the true fluxes and compute the algorithm to infer a range of possible tuples (R,B).

SPECIFIC COMMENTS

Specific comments are answered when explicit explanations seemed necessary. Otherwise we will directly clarify the points in the modified version of the manuscript.

•P3742: *The state vector elements don't seem to add up.*

The state vector has a dimension of 99: 85 for the lateral boundary conditions, 1 for the initial conditions, 1 for the offset and 12 for the scaling factor on the emissions in aggregated regions.

•P3748, line 8: *What physical considerations?*

C2733

Errors dependent of PBL variability especially. And flux uncertainties consistent with inventories uncertainties.

•P3750, line 9: *Design on B. The underlying assumptions are not clear to me. What are d_b^a and d_b^0 in equation 9?*

•P3750, line 14-17: *Please explain in more detail (I'm getting lost here).*

This section was obviously too briefly described. We will give more details on the methods.

In the Eq. (9), only a single matrix **B** is compatible with the equality if we consider that **H** comes from a family of free vectors (that is physically very likely in most cases and is verified in our study). Then, any way to find a compatible **B** matrix actually gives the only one. We estimate a compatible **B** using the inverse of $\mathbf{H}\mathbf{H}^T$:

$$\mathbf{B} = (\mathbf{H}\mathbf{H}^T)^{-1}\mathbf{H}^T E[d_b^a (d_b^0)^T] \mathbf{H} (\mathbf{H}\mathbf{H}^T)^{-1}$$

The equation should have given the opportunity of implementing an iterative algorithm. But it actually diverges very quickly after a single step of improvement, probably for the same reasons as in the computation of **R**. That is why we stop at the first step and consider **B** as a rough estimation of the optimum.

•P3751, line 12: *What is meant by 'results'? The R matrix? What is to be expected in this case depends on the off diagonals (size and sign). It is hard to make a comparison, if matrices differ not only by off diagonals but also the diagonals. Therefore it is not clear to me that this outcome was expected.*

The sentence indeed should be reformulated. "Results" refer to the ND tuple. The order of the description maybe was not well chosen. The variances in \mathbf{R}_{ND} are smaller than in the other methods. In a first approximation, one will then expect a better constraint by the observations. But the correlations make it difficult to interpretate precisely the

C2734

covariances. We then analyze in more details the correlations in Sect. 3.3.

•P3751, line 25: *Is this cause hypothesized or has it really been tested?*

In the DS scheme, the subsets are very large and the separation into subparts of the observations are made on criteria relevant mostly for plain station. The PBL creates high uncertainties in mountain sites at rates that are not necessarily synchronized with the repartition we chose.

Consequently, the DS scheme does not recover consistent patterns of errors in mountain sites.

•P3752, line 6: *So for ML the observational and the background errors are smaller than for DS. How in that case can the overall χ^2 's be comparable (i.e. satisfy the χ^2 criterion)?*

The χ^2 test evaluates the balance between observation and flux errors. Then, smaller errors are not incompatible with a still valid χ^2 test.

•P3752, line 8: *what is meant by 'no clear behavior'?*

This should be reformulated in the new version of the manuscript. We cannot extract any generic patterns by comparing (R_{ND}, B_{ND}) to the other tuple. Some variances are smaller, while other are bigger than for the other methods with no obvious physical interpretation.

•Table 2: *I'm surprised by the low p values given the generally low r values, to make this clearer it would be helpful to include to number of data.*

We agree that the p -value are very small in most cases. But the calculations give these figures. We use from a few hundreds to a thousand occurrences to compute the correlations which are hence very well constrained.

C2735

•P3762, line 28: *To make use of nighttime data using increased uncertainties assumes that model errors associated with the simulation of the nighttime PBL are random, which is likely not the case.*

This is indeed the case. We should clarify the point and present more careful recommendations in the use of nighttime data.

TECHNICAL COMMENTS

•P3750, line 3: *'convergence' instead of 'divergence'?*

•P3751, line 21: *'as' instead of 'then'*

•P3752, line 19: *'as a' instead of 'by way of'*

•P3753, line 18: *interquartile 'range' instead of 'gap'*

•P3755, line 22: *remove 'issue'*

•P3760, line 1: *'as' instead of 'than'*

•P3760, line 22: *'assumptions' instead of 'hypotheses'*

•P3761, line 3: *'of' instead of 'to'*

Wording corrections will be taken into account in the next version of the manuscript.

C2736

References

- Desroziers, G. and Ivanov, S. (2001), Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q.J.R. Meteorol. Soc.*, 127: 1433-1452. doi: 10.1002/qj.49712757417
- Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008), An adaptive ensemble Kalman filter for soil moisture data assimilation, *Water Resour. Res.*, 44, W03423, doi:10.1029/2007WR006357.
- Winiarek, V., M. Bocquet, O. Saunier, and A. Mathieu (2012), Estimation of errors in the inverse modeling of accidental release of atmospheric pollutant: Application to the reconstruction of the cesium-137 and iodine-131 source terms from the Fukushima Daiichi power plant, *J. Geophys. Res.*, 117, D05122, doi:10.1029/2011JD016932.