Atmos. Chem. Phys. Discuss., 13, C1592–C1601, 2013
www.atmos-chem-phys-discuss.net/13/C1592/2013/

Atmospheric Chemistry and Physics Discussions

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: 23 April 2013

We thank I. G. Enting for his comments that will help improve the presentation of our study. We have addressed them all. They are copied hereafter in italics with our detailed answers inserted in standard font where appropriate.

General comments

This paper represents an important advance if the technical issues noted below can be addressed. In the 20 years since Bayesian inversions were described, the refinement of the statistical basis has been quite slow. My own view is that for regional inversions, it will probably be appropriate to go beyond the assumption of normally distributed

errors, (see for example the Cape Grim CO_2 data set plotted in Enting (2002)).

We agree that the Gaussian assumption is a critical point. This hypothesis could be discussed and inquired into through the use of anamorphosis or other known technics adaptable to a non Gaussian framework. This should be done in further steps of our work. However, our current results do not seem to be much affected by non-Gaussianity.

Context

A point that is emphasised by Enting (2002) and more recently by Enting et al. (2012) is that in terms of statistical analysis, the inverse problem should be seen as one of statistical estimation. In order to better integrate with the statistics community, and draw from the wider literature, the use of standard statistical notation and terminology is highly desirable. A particularly important aspect is the use of the 'hat' notation, to denote estimates as, for example, $\hat{x}$. (As an example of the importance of the distinction, things such as expressing the mean square error of an estimate as $\mathrm{E}[(\hat{x}-x)^2]$ *become much more complicated and/or obscure without such a distinction.) Similarly E[.] is termed the 'expectation' not (as is done in the paper) the 'expectancy'. As a minor point, since E is not a mathematical variable (unlike in* $E = mc^2$ *) an upright font should be used.*

The issue on the notations used in the manuscript is noted. Efforts will be made to make equations consistent with the notations of the statistics community in the revised manuscript.

ACPD

13, C1592–C1601, 2013

Comments on presentation

•The paper seems not to come to grips with an essential question: how much information about $(\mathbf{R}, \mathbf{B})$ *can be obtained from the observations.*

The referee has emphasized the necessity of quantifying the amount of information one can extract from the observations to build the covariance matrices of errors. Actually, information is used in parallel and not redundantly for covariance building and for the inversion itself. We cannot know exactly the balance between the information used for the matrix building and the information used for the inversion because we do not have direct access to the errors of the estimations of the hyper-parameters (error variances and covariances). This could be done by Monte Carlo perturbations on the prior information put into the system. As far as we know, this has not been done in past studies computing objectively hyper-parameters. Moreover, in a system using real data, it could be very time-consuming to get a reliable estimation of these errors considering the expected correlations of the errors on the hyper-parameters. Hence, a proper computation of a dfs is difficult.

However, in section 5.2, we estimated the dfs taken by the assimilation system (Cardinalli et al., 2004). We found a figure of around 50 to be compared to the 100 state parameters and 5000 observations. Then, many pieces of information could have been used for the covariance matrices building.

We agree with the necessity of evaluating the errors on the hyper-parameters (in particular the error correlations). But, for the current step, the computed hyper-parameters are very consistent with our physical knowledge and we comment them as such.

ACPD

13, C1592–C1601, 2013

- Desroziers: $(\mathbf{R}, \mathbf{B})$ are taken as diagonal and then the data/state vectors are stratified so that 41 (26+15) variances are to be estimated. Conceptually, this seems similar to the approach of Michalak et al. (2005): characterise $(\mathbf{R}, \mathbf{B})$ using a physically-based stratification using a small number of parameters (except that Michalak et al. (2005) do a joint estimation of state and error parameters).

The Desroziers' scheme indeed uses a stratification of the covariance matrices. The principle is basically the same as in Michalak et al. (2005). But Michalak et al. use a Gauss-Newton method to optimize the likelihood whereas we use a fixed point method based on Desroziers et al. (2001).

- Maximum likelihood: This case raises a range of questions. The implication is that although $(\mathbf{R}, \mathbf{B})$ are still taken as diagonal, all diagonal elements are being estimated independently, estimating more quantities than the number of data - this would seem to be insufficiently robust to be useful.

The referee has raised some issues on the consistency and the robustness of the algorithm. We agree that the limited number of pieces of information to infer the set of covariance matrices maximizing the likelihood and then to assimilate the observations is critical.

However, the algorithm explicitly computes a maximizing value of $\mathbf{R}+\mathbf{HBH}^T$. One can prove that $\mathbf{R}+\mathbf{HBH}^T$ is unique when assuming that $\mathbf{R}$ and $\mathbf{B}$ are diagonal. Moreover,

ACPD

13, C1592–C1601, 2013

again when $\mathbf{R}$ and $\mathbf{B}$ matrices are diagonal, this matrix $\mathbf{R} + \mathbf{HBH}^T$ can be expressed in a unique way by the related tuple ($\mathbf{R}$,$\mathbf{B}$). Then, only n_{obs} degrees of freedom are used in this section (2.2.3 Maximum of likelihood).

These points will be explained and/or clarified in the next version of the manuscript. This method relies on the independence of background and observation errors and on their Gaussian structure. These points must be discussed in the following steps of the work. An estimation of the errors on the estimation of the variances should also be computed in further steps.

- Observation space diagnostics: In this case, many more quantities are being estimated than the available date and again would seem to be insufficiently robust to be useful. This would imply a χ^2 test with a negative number of degrees of freedom. There is a need to explain more clearly what is being done. Other technical issues with the description are noted below.

We have not been clear enough about the purpose of this method. The dimension of the space of compatible tuples of covariance matrices is large and no unambiguous solution can be computed. The algorithm is only computed to give guidelines in possible correlation structures in observation and background errors. Given this aim, we use Desroziers' diagnostics as necessary conditions to be filled by the errors.

In principle, these diagnostics give the opportunity of computing a fixed point iterative algorithm. The relations on the expectation of the product of the innovation vectors described in Desroziers et al. (2005) can be developed in term of $\mathbf{R}$ and $\mathbf{B}$ using the fact that a product between observation and background errors will have a zero expectation (independence of the two spaces). The calculated expression, giving the expression

ACPD

13, C1592–C1601, 2013

of $\mathbf{R}_{k+1}$ as a function of $\mathbf{R}_k$, $\mathbf{B}_k$ and $\mathbf{H}$, could be used for iteratively computing $\mathbf{R}$. The same holds for $\mathbf{B}$. However, as reported by Desroziers et al. (2005), there is no mathematical evidence of a possible convergence of such an algorithm. We tested it on our set up and the algorithm indeed did not converge.

The Monte-Carlo method was tested and gave better results with significant improvements in the likelihood value. This is why we developed this point. More explicitly, the expression for iterative calculation is:

$$E\left[\left(\hat{\vec{y^0}} - \mathbf{H}\vec{x}(\hat{\vec{y^0}}, \mathbf{R}, \mathbf{B})\right)\left(\hat{\vec{y^0}} - \mathbf{H}\hat{\vec{x^b}}\right)\right] = \mathbf{R}_{k+1}$$

where $\hat{\vec{y^0}}$ is a realization of $\vec{y^0}$ of error $\mathbf{R}_k$, $\hat{\vec{x^b}}$ is a realisation of $\vec{x^b}$ of error $\mathbf{B}_k$ and $\vec{x}(\hat{\vec{y^0}}, \mathbf{R}, \mathbf{B})$ depends on $\mathbf{R}_k$, $\mathbf{B}_k$, $\hat{\vec{x^b}}$ and $\hat{\vec{y^0}}$

The algorithm does not end up with an explicit convergence point. However, the likelihood value is significantly improved compared to what the diagonal restriction led to.

Issues

•While I think this work is an important advance, I am unhappy to see it described as optimal. Indeed I think that such a description is meaningless in the absence of any specification of the criteria (e.g. minimising a specified objective function) against which is being optimised. (The title of the paper merely says 'better'.)

The two objective criteria used to qualify our system as optimal is the χ^2 test and the maximum of the likelihood. In that sense, the covariance matrices specified in

ACPD
13, C1592–C1601, 2013

the system are optimized, though not optimal as detailed above. The wording will be changed in the next version of the manuscript.

•What is going on here, in at least some of the cases, is the use of the observations to estimate the tuple (R, B) and then to use the same data to estimate the state as if (R, B) is known exactly. I think that this is technically unsound, although numerically it might not be important. The issue should at least be noted, even if actual tests (e.g. by Monte Carlo) are left to a later study.

As detailed earlier, the errors on the hyper-parameters should indeed be estimated, though the computation is left for later. This point will be more specifically emphasized in a later version of the manuscript.

•I am having great difficulty analysing the procedure associated with relations (8)

- a specific reference to equations in (Desroziers et al., 2005) would be helpful;
- as an iterative procedure, this only seems to make sense if the subscripts on the left hand sides of lines 2 and 3 of relation (8) are $k+1$ rather than k.
- Desroziers et al. (2005) state that they are solving a non-linear fixed point relation. It would be helpful of the authors could say what they are solving.

My guess is

ACPD
13, C1592–C1601, 2013

$$\mathrm{E}\left[\left(\vec{y^0} - \mathbf{H}\hat{x}(\vec{y^0}, \mathbf{R}, \mathbf{B})\right)\left(\vec{y^0} - \mathbf{H}\vec{x^b}\right)\right] = \mathbf{R}$$

where $\hat{x}(\vec{y^0}, \mathbf{R}, \mathbf{B})$ indicates that $\hat{x}$ (i.e. $\vec{x^a}$) depends on $(\vec{y^0}, \mathbf{R}, \mathbf{B})$ because of eqn (2).

- Desroziers et al. (2005) appear to be estimating the expectation from a sum over the observational data set. It is not clear to me that the Monte Carlo technique used here is a valid way of evaluating the expectation for the purposes of solving the non-linear fixed point relation. It seems that the expectations are being calculated over random variables from two different realisations from (approximately) the same distribution: firstly a fixed sample of observations and secondly samples from a Monte Carlo simulation. Any terms in (7) that represent products of such random variables should have zero expectation if the variables come from different realisations, but will in general have non-zero expectation if the random variables come from the same realisation. The analysis given here needs to be justified by a term by term expansion that captures these distinctions. (I may be able to comment more later in the discussion period if the authors are able to confirm that they are trying to solve a fixed point relation (and clarify relation (8)).

The exact procedure is described above. The remarks will be taken into account. More details on the equations will be provided for the fixed-point method.

Queries

•P 3745, L 1 Should 'maximizing' really be 'minimizing'?

Minimizing is the correct term as the cost function is written in equation (1)

•P 3745, L 2 Should $\mathrm{J}^a(\vec{x^a})$ *really be* $\mathrm{J}^b(\vec{x^a})$ *?*

ACPD

13, C1592–C1601, 2013

Interactive Comment

Will be transformed into J^b

•P 3749, eqn 4, The subscript n on the identity matrix should not be bold font.
The modification will be taken into account.

Wording

•P 3739 L 28, to inverse → to invert

•P 3741, L 2, life time → lifetime

•P 3746, L 13, converges to Eq (4) → converges to a tuple that satisfies Eq (4)

•P 3748, L 16, hence the ML algorithm → hence the constrained ML algorithm

•P 3749, L 5, do is an innovation vector, but in this context, do

•P 3749, L 12, 20, criterium → criterion

(criterium refers to bicycle races). Also p3750, L14.

•P 3750, L 10, constraints → constrains

Other

Finally there are a small number of places where minor changes might better reflect English idiom. Some suggestions are:

ACPD
13, C1592–C1601, 2013

•*P 3737, L 18, to close → for closing*

•*P 3739, L 7, takes benefit of → exploits*

•*P 3739, L 15-16, punctual → point*

•*P 3741, L 6, In all the study → Throughout the study*

•*P 3745, L 14, follows → satisfies*

•*P 3750, L 1, with → based on*

•*P 3763, L 14, influent → influentual Printer-friendly Version*

•*P 3764, L 22, implementation → inclusion Interactive Discussion*

•*P 3755, L 4, should be subdued → is unlikely to apply*

•*P 3762, L 17, Totalizing → Summing*

The wording and re-phrasing will be taken into account according to reviewer's suggestions.

Interactive comment on Atmos. Chem. Phys. Discuss., 13, 3735, 2013.