Geosci. Model Dev. Discuss., 7, C2227–C2233, 2014
www.geosci-model-dev-discuss.net/7/C2227/2014/


Geoscientific Model Development Discussions
Open Access

# *Interactive comment on* "Objectified quantification of uncertainties in Bayesian atmospheric inversions" *by* A. Berchet et al.

**M. Bocquet (Referee)**

bocquet@cerea.enpc.fr

Received and published: 10 November 2014

This a very ambitious paper, with three significant achievements: the co-called marginalisation method, the reduction of the number of control variables by defining a proper aggregation of these variables, and an application/validation with a far from trivial methane inversion OSSE.

This paper is interesting in many ways. It has the essential quality of advocating a little further the use of rigorous tools from inverse problems and Bayesian inference, hopefully making current atmospheric inversion studies more credible. For these reasons, I believe it has to be published after suitable revision.

The manuscript is very well structured. However, it now and then lacks clarity. Gentle



reminders throughout the manuscript are actually helpful in that respect. It is quite long, but that's acceptable since there are a lot of material and ideas in it. However, the manuscript suffers from time to time from ingenuity in the the naming of concepts. This absolutely needs to be fixed. For instance, I believe a vast majority of researchers across fields are more accustomed to the maximum likelihood estimation (maximum likelihood in short, or MLE) rather than "Maximum of Likelihood". What you call a node is actually known as mode (which coincides for the Gaussian distribution with the mean), i.e. the maximum of the pdf! Please change the many occurrences of these in the manuscript. There are other examples that I mention below.

Page 4785 is crucial in the exposition of the main ideas. Even though I believe it is essentially correct, there are several unclear, confusing and sometimes incorrect and badly-informed statements that need your attention and correction.

You claim that maximum likelihood estimation of $\mathbf{B}$, $\mathbf{R}$, actually of the hyper-parameters, is a rough approximation may not be correct. The reason why we chose this approach is because the pdf $p(\mathbf{y}|\mathbf{R},\mathbf{B})$ is actually very highly peaked onto a couple a matrices, and even more some onto a few hyper-parameters, when $\mathbf{R}$ and $\mathbf{B}$ are restricted to a few parameters. This is a marginalisation on the most likely hyper-parameters values. It takes the argument of the supremum value of a pdf, whereas you choose a mean estimate which we expect to be better if we can numerically afford it, or worse if you can't. It may be that the fewer the observations, the larger the number of hyper-parameters, the less correct keeping the argument of the supremum is. So what you propose is certainly correct and interesting but not always necessary. This has to be discussed.

A list of more of less important minor remarks follows.

p. 4870, l. 4: "deterministically" : "univocally" would be more to the point.

p. 4783, l. 5: "at a scale large enough for the turbulence to be negligible" why so? a tracer transported by a turbulent flow has a linear dependence on the source. So either your remark is incorrect or I missed your point.

p. 4783, l. 23: "apart from the technical issues in the implementation of the theory on computers" $\longrightarrow$ "apart from technical issues in the numerical implementation of the theory"

p. 4784, l. 4: "tuple" seems a little bit pedantic if you just meant "couple" (python language practitioners?)

p. 4785, l. 3: "Here, we assume no prior information...is then uniform". No! It is well known in Bayesian statistics that in the absence of extra information, the uniform distribution is often a bad choice. It is usually advised to resort to one of the so-called non-informative prior. You could have a look at Bocquet (2011), where an extension of the ensemble Kalman filter that efficiently accounts for sampling errors at almost no additional cost. In this paper, I actually marginalised on $\mathbf{B}$ just as you do, but choose for $p(\mathbf{B})$ the Jeffreys' distribution (usually called an hyper-prior). Bocquet (2011) is also of interest to you because the marginalisation also bears on $\mathbf{x}_b$, hence the bias.

p. 4785, l. 12: "There is no reason for the complete pdf to be a Gaussian itself": a nice example taken from Bocquet (2011): the marginalisation on $(\mathbf{x}_b, \mathbf{B})$ gives a multivariate T-distribution with large tails.

p. 4785, l. 13: "it cannot be described with only its mode and its covariance matrix": this is a confusing statement because for instance, in Bocquet (2011), as soon as we know it is a multivariate T-distribution with a specific parameter then a mode and the covariance matrix are enough to characterise the distribution. I would get rid of this statement.

p. 4785, l. 17-21: I did not get your point. Please clarify.

p. 4786, l. 3: Even though I like Michalak's paper very much (which needs to be cited in the paper anyway as you did), I believe Dee (1995) would be a more appropriate



reference here.

p. 4786, l. 23-24: "But, with such a direct algorithm, ...on the result". That is too strong a statement. There are good reasons (maybe not always justified) why MLE is actually very good in most application. See my main comment.

p. 4786, l. 23-24: That was also done by Koohkan et al. (2013) working on diagonal $\mathbf{B}$ and $\mathbf{R}$, in the estimation of VOC fluxes.

p. 4787, l. 17: What do you mean by "pseudo-Newtonian"? quasi-Newton (like the BFGS method), or something else?

p. 4787, l. 16-26: Even with the diagram the algorithm is still not clear enough to me. How do you sample the diagonal $\mathbf{R}$ and $\mathbf{B}$? (I guess I have understood but I can't be sure your readership will.) A very important detail: How many draws do you use in the Monte Carlo sampling?

p. 4790, l. 7-15: Please add numbers to the equations. The one for the error is awkward. Is that intentional?

p. 4789, 4790: There was a reason why Bocquet et al. (2011) did not choose any $\mathbf{P}$ or $\Pi$ (but $\Gamma_\omega$) for what you designate as $\Pi_\omega$. This is not truly a projector but the composition of a projection with an injection operator which might confuse the reader who really wants to go into the algebra.

p. 4791, l. 20-22: "For this reason, we decide to define ...": Is this "$\Lambda_\omega$" a linear operator? If no, does this invalidate the use of your equations?

p. 4793, l. 18: Koohkan and Bocquet (2012) is not the one you intended to cite here but Koohkan et al. (2012) (where a fixed optimal representation $\omega$ is computed for a fixed global network).

p. 4793, l. 28: Again check this very odd expression: "pseudo-Newtonian Maximum of Likelihood". Did you mean a maximum likelihood minimisation using a quasi-Newton method?

p. 4794, l. 1-9: This is very interesting. It may offer a measure of the optimality of the representation. It may not truly be "numerical artefacts". The optimality criteria used by Bocquet and co-authors are actually information theory-based and uses $\mathbf{KH}$.

p. 4794, l. 20: What is a "Fisher-like" distribution?

p. 4795, l. 15: "a Pseudo-Newtonian ascending algorithm": !? Do you mean "a quasi-Newton descent method"?

p. 4795, l. 23: "marginalize inversion" $\longrightarrow$ "marginalized inversion"

p. 4795, l. 24: "The main difference": with what? If you mean with the rest of the literature, I disagree, objective online estimation of error covariance matrices are already performed in atmospheric chemistry inversion and numerical weather forecast. The added value here is the computation of a Monte Carlo marginalisation, which has not been attempted in the field of atmospheric chemistry inversion (at least not to my knowledge).

p. 4796, l. 6-13: Representativeness errors are also embedded in the observation errors as seen from the data assimilation system. The fact this instrumental error is negligible does not change this fact. You could state this. Now, you can decide to set it to zero in the OSSE, the representativeness errors being very difficult to simulate in an OSSE.

p. 4798, l. 4: "Monte-Carlo tuples" $\longrightarrow$ "Monte Carlo draws" (note the absence of dash in English).

p. 4799, l. 12: "ays" $\longrightarrow$ "days"

p. 4799, l. 5: "obervation" $\longrightarrow$ "observation"

p. 4801, l. 7: The symbol that you use is usually not reserved for convolution. What does this operator correspond to exactly? I assume it is point-wise multiplication. If



that is correct, replace "convolution" with "point-wise multiplication".

p. 4803, l. 6-9: This is very similar to (Koohkan et al., 2012) where the footprints of FLEXPART are used to determine the representation.

p. 4803, l. 11: "non-hydrostatic": such attribute is mostly used to describe (often convective-scale) meteorological models, not CTMs. What do you mean by that?

p. 4810, l. 11-20: Another more consistent option is to marginalise over the biases, like what is done in Bocquet (2011) which results in some additional blurring of the ensemble mean.

p. 14810, l. 25: "We developed a new Bayesian method of inversion from the classical Bayesian framework": It is more fair to say that you extended the classical Bayesian framework. State-of-the-art geophysical estimation nowadays includes some objective covariance parameter (hyper-parameter) estimation, which is marginalising on the most likely hyper-parameters. You extend this by Monte Carlo computing corrections to the most likely hyper-parameters.

p. 1811, l. 6: "virtual truth": usually called "nature run" using OSSEs' terminology.

p. 4825, Figure 4: It is customary to show the frequency on the y-axis.

## References


M. Bocquet. Ensemble Kalman filtering without the intrinsic need for inflation. *Nonlin. Processes Geophys.*, 18:735–750, 2011. doi: 10.5194/npg-18-735-2011.

M. Bocquet, L. Wu, and F. Chevallier. Bayesian design of control space for optimal assimilation of observations. I: Consistent multiscale formalism. *Q. J. R. Meteorol. Soc.*, 137:1340–1356, 2011. doi: 10.1002/qj.837.

D. P. Dee. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, 123:1128–1145, 1995.

M. R. Koohkan and M. Bocquet. Accounting for representativeness errors in the inversion of atmospheric constituent emissions: Application to the retrieval of regional carbon monoxide fluxes. *Tellus B*, 64:19047, 2012. doi: 10.3402/tellusb.v64i0.19047.

M. R. Koohkan, M. Bocquet, L. Wu, and M. Krysta. Potential of the international monitoring system radionuclide network for inverse modelling. *Atmos. Env.*, 54:557–567, 2012. doi: 10.1016/j.atmosenv.2012.02.044.

M. R. Koohkan, M. Bocquet, Y. Roustan, Y. Kim, and C. Seigneur. Estimation of volatile organic compound emissions for Europe using data assimilation. *Atmos. Chem. Phys.*, 13:5887–5905, 2013. doi: 10.5194/acp-13-5887-2013.

---

Interactive comment on Geosci. Model Dev. Discuss., 7, 4777, 2014.