Atmos. Chem. Phys. Discuss., 15, C444–C453, 2015
www.atmos-chem-phys-discuss.net/15/C444/2015/

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Balancing aggregation and smoothing errors in inverse models" *by* A. J. Turner and D. J. Jacob

M. Bocquet (Referee)

bocquet@cerea.enpc.fr

Received and published: 27 February 2015

1 Main comments

The manuscript is about one of the most interesting theoretical problems of inverse modelling/data assimilation in geosciences, namely finding the control space resolution (possibly in the form of an adaptive grid) that would minimise the inversion errors. This optimal resolution could differ from the forward model resolution.

I appreciate the efforts of the authors to address this difficult issue and some of the numerical illustrations of the manuscript. However, the authors overlooked the findings that have been reported by Bocquet et al. (2011); Wu et al. (2011). From the results

of Bocquet et al. (2011); Wu et al. (2011), I believe that the optimal resolution as seen by the authors is the results of *suboptimal* choices. What is reported in the present manuscript is nevertheless interesting since those suboptimal choices could be made for the sake of numerical efficiency. It is problematic that the authors are (unintentionally) hiding what actually leads to the appearance of a minimum in the total error curve as a function of resolution. I do believe this paper will eventually become acceptable once those issues have been clarified and once the manuscript better discusses its relationship with Bocquet et al. (2011); Bocquet and Wu (2011); Wu et al. (2011).

1. Frankly, the notations are unfriendly. I understand the authors follow those of Rodgers (2000). Yet, they diverge a lot from standard data assimilation or inverse modelling notations that have been widely adopted in atmospheric chemistry data assimilation. For instance "a" usually refers to the analysis while the authors use it to refer to the prior, when "b" ("f" in a sequential context) is very often chosen. The gain is usually designated as $\mathbf{K}$, not $\mathbf{G}$; "$\mathbf{H}$" is much preferred to "$\mathbf{K}$" for the observation/Jacobian/source-receptor operator. That said, the choice of notations belongs to the authors. But, I guess that the present notations would significantly distract potential readers.

2. One of the results of Bocquet et al. (2011) is that with a proper choice of prolongation operator, one can reduce the *smoothing* error as much as possible, so that the total error (smoothing+aggregation) is actually a monotonically decreasing function of the resolution. If this is correct, there is no optimal resolution but the finest one (CTM's for instance), except from a numerical efficiency standpoint or if one introduces other sources of scale-dependant errors (such as model errors). The authors presumably obtain such (discrete) optimum because they make an arbitrary choice in the prolongation operator which restricts the transfer of information through scales. Mathematically speaking, this can be seen as an *artifact*. Had the authors made another implicit choice for the prolongation operator, they

ACPD
15, C444–C453, 2015

would have found a different result, possibly leading to the finest grid being optimal. If this correct, the authors should clearly acknowledge this and give a fair account of the findings of (Bocquet et al., 2011). I give mathematical details below.

2 Discussion on the total error

Using more standard notations, Equation (27) of the manuscript reads :

$$\begin{aligned}\mathbf{P}_\omega^a &= \underbrace{(\mathbf{I}-\mathbf{K}_\omega\mathbf{H}_\omega)\mathbf{B}_\omega(\mathbf{I}-\mathbf{K}_\omega\mathbf{H}_\omega)^{\mathsf{T}}}_{\text{smoothing error}} \\ &+\underbrace{\mathbf{K}_\omega(\mathbf{H}-\mathbf{H}_\omega\Gamma_\omega)\mathbf{B}(\mathbf{H}-\mathbf{H}_\omega\Gamma_\omega)^{\mathsf{T}}\mathbf{K}_\omega^{\mathsf{T}}}_{\text{aggregation error}} \\ &+ \underbrace{\mathbf{K}_\omega\mathbf{R}\mathbf{K}_\omega^{\mathsf{T}}}_{\text{observation error}} \end{aligned} \quad (1)$$

This decomposition agrees and is entirely consistent with what is derived in Bocquet et al. (2011); Wu et al. (2011) (I checked).

In the manuscript, the authors do not specify what $\mathbf{H}_\omega$, $\mathbf{K}_\omega$ are. Of course, such choices must be made for the method to be applied. Bocquet et al. (2011) discuss these choices. The main choice they work with is based on a prolongation operator that is consistent with the BLUE formalism. In the following, for the sake of simplicity, I assume that the background estimate (also known as first guess) is zero. The general case is summarised in Bocquet et al. (2015). Yet, the total error budget is not affected by this simplification. However this could be important, for instance when one has an inventory at the finest scale (which is frequent with greenhouse gases, and which means that the authors would need to properly address this issue too). A choice of a

ACPD
15, C444–C453, 2015

prolongation operator consistent with the BLUE formalism is

$$\mathbf{\Gamma}_\omega^\star \equiv \mathbf{B}\mathbf{\Gamma}_\omega^\mathsf{T}\left(\mathbf{\Gamma}_\omega\mathbf{B}\mathbf{\Gamma}_\omega^\mathsf{T}\right)^{-1}. \tag{2}$$

which leads to the definition of the projection operator

$$\mathbf{\Pi}_\omega \equiv \mathbf{B}\mathbf{\Gamma}_\omega^\mathsf{T}\left(\mathbf{\Gamma}_\omega\mathbf{B}\mathbf{\Gamma}_\omega^\mathsf{T}\right)^{-1}\mathbf{\Gamma}_\omega \tag{3}$$

which quantifies the lost information when mapping from and to the finest grid (the CTM grid), but passing through the coarse grid ω. The restriction and prolongation operators satisfy

$$\mathbf{\Gamma}_\omega\mathbf{\Gamma}_\omega^\star = \mathbf{I}_N\,, \qquad \mathbf{\Gamma}_\omega^\star\mathbf{\Gamma}_\omega = \mathbf{\Pi}_\omega\,. \tag{4}$$

The Jacobian matrix $\mathbf{H}$ becomes $\mathbf{H}_\omega = \mathbf{H}\mathbf{\Gamma}_\omega^\star$ in ω. The source-receptor/observation equation reads

$$\boldsymbol{\mu} = \mathbf{H}_\omega\mathbf{x}_\omega + \boldsymbol{\epsilon}_\omega = \mathbf{H}\mathbf{\Gamma}_\omega^\star\mathbf{\Gamma}_\omega\mathbf{x} + \boldsymbol{\epsilon}_\omega = \mathbf{H}\mathbf{\Pi}_\omega\mathbf{x} + \boldsymbol{\epsilon}_\omega\,. \tag{5}$$

With the consistent choice Eq. (2), it can be proven that the innovation statistics (defined in observation space) are scale independent, which is rather intuitive if information is properly transferred through the scales. This reads:

$$\mathbf{R}_\omega + \mathbf{H}_\omega\mathbf{B}_\omega\mathbf{H}_\omega^\mathsf{T} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T}\,. \tag{6}$$

Because of the properties of this prolongation operator, it was shown that Eq. (1) can eventually be written:

$$\mathbf{P}_\omega^a = \mathbf{\Gamma}_\omega\left(\mathbf{B} - \mathbf{B}\mathbf{H}^\mathsf{T}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^\mathsf{T}\right)^{-1}\mathbf{H}\mathbf{B}\right)\mathbf{\Gamma}_\omega^\mathsf{T} = \mathbf{\Gamma}_\omega\mathbf{P}^a\mathbf{\Gamma}_\omega^\mathsf{T}\,, \tag{7}$$

which is disappointingly simple. It relies on the fact that the representativeness (aggregation) error is properly accounted for in the retrieval performed in the ω grid.

ACPD
15, C444–C453, 2015

A measure of the total error is $E = \mathrm{Tr}(\mathbf{P}^a_\omega) = \mathrm{Tr}(\mathbf{P}^a\mathbf{\Gamma}^{\mathsf{T}}_\omega\mathbf{\Gamma}_\omega)$. If $\mathbf{\Gamma}^{\mathsf{T}}_\omega\mathbf{\Gamma}_\omega$ is proportional to a projection operator (it is likely to be so for $\mathbf{\Gamma}_\omega$ being a coarse-graining operator), then this mathematically implies that the total error is a monotonically decreasing function of resolution (or that it systematically grows with aggregation). See Bocquet et al. (2011), section 5.3 for a specific proof.

A directly related measure is the reduction of uncertainty proposed in Bocquet (2009): $\mathrm{DFS} = \mathrm{Tr}(\mathbf{I} - \mathbf{B}^{-1}_\omega\mathbf{P}^a_\omega)$. It is better because (i) it is invariant by reparametrisation of the control variables, (ii) it is directly related to the uncertainty reduction widely used in greenhouse gas inversions and to the number of degrees of freedom for the signal widely used in experimental design for geosciences data assimilation. In that case, $\mathrm{DFS} = \mathrm{Tr}(\mathbf{\Pi}_\omega\mathbf{\Omega})$, where $\mathbf{\Omega} = \mathbf{B}\mathbf{H}^{\mathsf{T}}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathsf{T}}\right)^{-1}\mathbf{H}$. $\mathbf{\Omega}$ and $\mathbf{\Pi}_\omega$ are $\mathbf{B}-$symmetric matrices. With respect to this scalar product, $\mathbf{\Pi}$ is an orthogonal projector and $\mathbf{\Omega}$ is positive definite. As a consequence, the DFS are a monotonically increasing function of any refinement applied to ω (see section 5.3 of Bocquet et al. (2011)).

When one cannot afford to build Eq. (2), as implicitly assumed in the authors' case, the analysis is suboptimal and the total error could well have a minimum. This has been mentioned in Bocquet et al. (2011); Wu et al. (2011), and even tested numerically in Wu et al. (2011). In that context, the optimal prolongation operator leads to a baseline total error, which is an important though simple result. This should be acknowledged.

The manuscript exemplifies what happens when the choice is not optimal (which is likely to happen quite often for us all in practice) and that is why I think it is very interesting, provided the authors acknowledge what the manuscript overlooks in the present form.

Still using Eq. (2), the error in observation space reads:

$$\begin{aligned}\mathbf{H}_\omega\mathbf{P}^a_\omega\mathbf{H}^{\mathsf{T}}_\omega &= \mathbf{H}_\omega\mathbf{\Gamma}_\omega\mathbf{P}^a\mathbf{\Gamma}^{\mathsf{T}}_\omega\mathbf{H}^{\mathsf{T}}_\omega = \mathbf{H}\mathbf{\Gamma}^\star_\omega\mathbf{\Gamma}_\omega\mathbf{P}^a\mathbf{\Gamma}^{\mathsf{T}}_\omega\left(\mathbf{\Gamma}^\star_\omega\right)^{\mathsf{T}}\mathbf{H}^{\mathsf{T}} \\ &= \mathbf{H}\mathbf{\Pi}_\omega\mathbf{P}^a\mathbf{\Pi}^{\mathsf{T}}_\omega\mathbf{H}^{\mathsf{T}} \end{aligned} \quad (8)$$

ACPD
15, C444–C453, 2015

$E = \mathrm{Tr}\left(\mathbf{P}^a \mathbf{\Pi}_\omega^\mathrm{T} \mathbf{H}^\mathrm{T} \mathbf{H} \mathbf{\Pi}_\omega\right)$ is very likely to behave similarly to the error in state space, for at least simple/regular enough $\mathbf{H}$. I could imagine that the mismatch between the satellite observations grid and the CTM grid could generate a non trivial dependence of E on the resolution. But I don't see why it would be pronounced. Besides, I guess a proper study of the problem would involve accounting for satellite observation error (non-diagonal $\mathbf{R}$).

3 Minor points or comments related to the major points

1. Title: We all know there is no such thing as an "inverse model". This is an abuse of language that I would personally avoid in a title. "Inverse modelling" is almost always preferred.

2. p. 1002, l. 4-6: "When the observation vector is large, such as with satellite data, selecting a suitable dimension for the state vector is a challenge". Selecting a suitable dimension for the state vector space is always a challenge, even, and perhaps even more so when the observation vector is small. Let me just mention one paper directly related to what you are discussing and where the observations are *in situ* and far less abundant than in a satellite retrieval context: Koohkan et al. (2012).

3. p. 1003, l. 6-7: Same remark as above.

4. p. 1003 l. 18-19: "and may not be able to depart from that knowledge". It all depends on the balance between the observation and background statistics. If the background is not informative enough, the solution may be highly oscillating. In a flux inversion context, the retrieved fluxes would increase around the observations sites, which is all but smoothing. In my humble opinion, the appellation is partially misleading. But I might not have understood its interpretation very clearly

ACPD
15, C444–C453, 2015

(from your manuscript or even Rogders' book), in spite of some experience with inverse modelling.

5. p. 1003 l. 18-19: "smoothing error" lacks a proper definition (although it is given later) and interpretation.

6. p. 28-29: "Numerical solutions using variational methods circumvent this problem but not inherently provide error characterisation as part of the solution": we know that this is not true. If that kind of statement was fine a few years ago, I believe it should be nowadays mitigated. Several researchers are using conjugate-gradient and quasi-Newton methods such as BFGS that inherently provide estimation of the posterior errors (for instance Bousserez et al., 2015).

7. p. 1004: the literature is incomplete. I believe you have to mention Wu et al. (2011), given it is very close to your objective and analysis and also related to greenhouse gas flux inversions.

8. p. 1006, l. 5-12: This is incomplete or partially incorrect. The Jacobian can also be computed using the model adjoint, requiring m runs. By the Sherman-Morisson-Woobury lemma, the matrix algebra will scale like m^3. Also, sequential updating by serial processing of observations usually (unless the scheme is sub-optimal) leads to the same numerical cost.

9. p. 1007, l. 11: "Probabilistic" is one word too many. Bocquet et al. (2011) additionally provide a probabilistic interpretation. But it can be seen as an entirely deterministic process just as the Best Linear Unbiased Estimator (BLUE) formalism. Please remove the word "probabilistic" which conveys the wrong idea in the context of this sentence.

10. p. 1007, l. 12: "However, construction of this prolongation operator is not a well-posed problem because the operator is not unique". Please rephrase the sentence. The construction as defined by Bocquet et al. (2011) is well-defined and

ACPD
15, C444–C453, 2015

well-posed. But in general the choice of the prolongation operator is not unique. Incidentally, you do make a choice for the operator without acknowledging it! That is why I disagree and think that your method might be less robust. But maybe you meant "more practical" rather than "more robust", did you? If you did intend "less robust", please justify your statement with precision.

11. p. 1007, Eq.(12): Please define $\mathbf{K}_\omega$ (the source-receptor matrix). That is where you put the definition of the prolongation operator under the carpet... This must be discussed.

12. p. 1008, l. 7-8: The introduction of the concept of ensemble is cumbersome (just as it is in Rodgers (2000) to be fair). It requires more justification. It appears as a *deus ex machina*.

13. p. 1008; l. 5: "A" for aggregation, and "a" for background. Really? Why not use "b" for background instead of “a”?

14. p. 1010, l. 10: Please define the gain $\mathbf{G}_\omega$ explicitly. As I explained, several choices can be made, one being more consistent. Without an explicit definition, you hide what is at the origin of the appearance of the fittest resolution.

15. p. 1011, l. 6-8: The sum of an increasing and decreasing function does not always possess a minimum.

16. Koohkan et al. (2012) discuss how to choose the optimal resolution and how it is impacted by the error balance (observation versus background, section 2.2). Since, ultimately, you end up making the same choice as all the papers I am referring to, that is to say choosing the resolution on a numerical cost basis, Koohkan et al. (2012)' discussion is relevant and perhaps a bit more precise than only adjustment with respect to the observation error only.

17. p. 1011, l. 12-18: Again, this discussion appears like a *deus ex machina*.

ACPD
15, C444–C453, 2015

18. p. 1012, l. 5: Actually the adaptive grid method based on tiling was introduced in Bocquet (2009). Moreover, it's worth mentioning that these grid are built to be optimal for the purpose of the inversion.

19. p. 1012, l. 8: Bocquet and Wu (2011) also use PCA coupled to the hierarchical grid to compute an optimal grid in a numerically efficient way yet capturing the variability of the prior. This should be acknowledged.

20. p. 1013, l. 15- 21: Rodgers (2000) also suggests projection over a specific function basis albeit in a different context.

21. p. 1016, l. 5: Could you please briefly discuss the numerical cost of the approach?

22. p. 1016: The application of the GMM methods is very interesting. From the methodological standpoint, I believe the fact that the control space is defined with a *probabilistic* mixture is quite novel in this context.

23. p. 1017: What about the time dimension? Do you apply aggregation in time? I assume you didn't, but you could have.

24. p. 1017: What if the background error covariance matrices were not diagonal? Could you discuss the issue a little? Apart from the numerical problem, we can see from Eq. (2) that properly transferring information through the scales is more tricky. If one chooses a pragmatical $\mathbf{\Gamma}^{\star}_{\omega}$ as you do (or as I could as well for a very high-dimensional application), it is possible that the resulting "optimal" resolution would be more pronounced.

25. p. 1017, l. 18: Please spell out SD (standard deviation?).

26. p. 1018, l. 8-13: Your result is not surprising. Because of the baseline results of Bocquet et al. (2011), I was expected that kind of results with a non-pronounced minimum (unless your implicit prolongation operator is badly chosen). Above all,

ACPD
15, C444–C453, 2015

you end up choosing the optimal resolution on a numerical efficiency criterion, just as we did (for not only practical but also theoretical reasons). This should be acknowledged.

27. p. 1018-1019: The conclusion should be amended.

References

Bocquet, M.: Towards optimal choices of control space representation for geophysical data assimilation, Mon. Wea. Rev., 137, 2331–2348, doi:10.1175/2009MWR2789.1, 2009.

Bocquet, M. and Wu, L.: Bayesian design of control space for optimal assimilation of observations. II: Asymptotics solution, Q. J. R. Meteorol. Soc., 137, 1357–1368, doi:10.1002/qj.841, 2011.

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

Bocquet, M., Wu, L., Chevallier, F., and Koohkan, M. R.: Sected topics in multiscale data assimilation, in: Advanced data assimilation for geosciences, edited by Blayo, É., Bocquet, M., Cosme, E., and Cugliandolo, L. F., pp. 415–431, Oxford University Press, Les Houches school of physics, 2015.

Bousserez, N., Henze, D. K., Perkins, A., Bowman, K. W., Lee, M., Liu, J., Deng, F., and Jones, D. B. A.: Improved analysis-error covariance matrix for high-dimensional variational inversions: application to source estimation using a 3D atmospheric transport model, Q. J. R. Meteorol. Soc., pp. n/a–n/a, doi:10.1002/qj.2495, 2015.

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

Rodgers, C. D.: Inverse methods for atmospheric sounding, vol. 2, World Scientific, Series on Atmospheric, Oceanic and Planetary Physics, 2000.

Wu, L., Bocquet, M., Lauvaux, T., Chevallier, F., Rayner, P., and Davis, K.: Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data, J. Geophys. Res., 116, D21 304, doi:10.1029/2011JD016198, 2011.

ACPD
15, C444–C453, 2015