Geosci. Model Dev. Discuss., 7, C3442–C3461, 2015
www.geosci-model-dev-discuss.net/7/C3442/2015/

Geoscientific
Model Development
Discussions
Open Access

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

A. Berchet et al.

antoine.berchet@lsce.ipsl.fr

Received and published: 27 February 2015

1 General comments

First we would like to sincerely thank the two reviewers for their efforts in reading and understanding through our relatively long manuscript. We thank them for their fruitful comments and review.

Following their recommendations, we substantially modified the structure of the text. We rearranged the sub-part order in the method section, added milestones and refined the general structure in order to significantly improve the readability of the manuscript. Now, the reader can follow our point from the subsection titles only.

As asked by Dr. Marc Bocquet, we clarified the terminology in the text and justified the use of the maximum likelihood estimation in a deeper way.

Referee $\#1$ proposed to drastically increase the readability of the manuscript. To do so, he/she proposed to remove part of the method section and to focus on the result analysis. In particular, it was suggested to elude the part where we develop the considerations for reducing the problem size. We think that this part is actually the key section of our work. We then preferred a deep clarification of the material, rather than removing this central part.

New sub-parts and important modifications are embedded in the new version of the manuscript attached to this document and are highlighted in red in the supplementary material.

The new structure of the document is now:

- 1. Introduction
- 2. Marginalized Bayesian inversion
  - 2.1. Context and motivation for the marginalization
    - 2.1.1. Bayesian inversion framework
    - 2.1.2. Ambivalent uncertainty set-up
    - 2.1.3. Possible uncertainty handling
  - 2.2. Marginalization of the inversion
    - 2.2.1. Theoretical formulation
    - 2.2.2. Monte Carlo sampling

    * 2.2.3. Processing the Monte Carlo posterior ensemble

- 3. Informed definition of the problem
  - 3.1. Principle for problem reduction
    * 3.1.1. Motivations and definition
    * 3.1.2. Mathematical formulation
  - 3.2. Representation choice
    * 3.2.1. Observation space sampling
    * 3.2.2. Observational constraints
    * 3.2.3. Flux aggregation
  - 3.3. Numerical artefacts
- 4. Validation experiments
  - 4.1. Required tests
    * 4.1.1. Method summary
    * 4.1.2. Test strategy
  - 4.2. OSSE evaluation
    * 4.2.1. Scoring system
    * 4.2.2. Posterior correlation processing

- 5. Set up of the OSSEs
  - 5.1. Virtual true state x^t
    * 5.1.1. State space components
    * 5.1.2. Generation of a perturbed reference state x^t
  - 5.2. Simulation of the observation operator H
    * 5.2.1. The Lagrangian model: FLEXPART
    * 5.2.2. The Eulerian model: CHIMERE
  - 5.3. Synthetic observations y^o
- 6. Results and discussion
  - 6.1. Robustness of the method
    * 6.1.1. Impact of the correlation processing
    * 6.1.2. Hot-spots and large-area emissions
  - 6.2. Spatial evaluation
  - 6.3. Limitations and benefits
    * 6.3.1. Promising computation of the uncertainties
    * 6.3.2. Subjective choices and biases
  - 7. Conclusions

2 Specific comments

Every single comment from the reviewers are answered here. Comments on lexical or formulation issue are only succinctly answered.

Comments by M. Bocquet (resp. referee $\#1$) are written in blue (resp. green).

1. p. 4870, l. 4: "deterministically" : "univocally" would be more to the point.
   Corrected.
2. Page 4778, line 11: computing ! performing
   Corrected in the text.
3. Page 4778, line 18: includes ! calculates
   Corrected in the text.
4. Page 4778, line 24: “real observation sites”. This is somewhat misleading, since the reader might think you also use real observations. I suggest: “with simulated observations on existing observation sites”
   We agree with this point. We replaced the expression by “with virtual observations on a realistic network in Eurasia”.
5. Page 4778, line 26: gas ! methane
   Modified.
6. Page 4779, line 11: reliably ! reliable
   Modified.
7. Page 4779, line 11: understanding ! understanding of
   Modified.
8. Page 4779, line 14: “inquire into the surface fluxes” ! “obtain information about the surface fluxes”
   Corrected.

9. Page 4779, line 16: the transport, the atmospheric chemistry, and the surface fluxes: “the” should be removed three times.
   Taken into account.
10. Page 4779, line 20: inferring back ! inferring
    Modified.
11. Page 4779, line 27: on the vertical column ! over the vertical column
    Modified.
12. from here on I mostly skip correcting small English grammar issues. Please use a native speaker to correct manuscript.
    We did our best to fix syntax and grammar issues in the body of the manuscript.
13. Page 4780, line 4: “The Bayesian: : :possible in order to: : :”. Rewrite.
    Rewritten in the new version.
14. Page 4780, line 15: “of the errors the transport model makes, “. Rewrite, e.g. “transport model error statistics”.
    Updated.
15. Page 4780, line 20: enlarge ! enlarges
    Corrected.
16. Page 4780, line 22 and on... to Page 4781, line 17: Most of this is the description of the method, and does not belong in the introduction. Refrain to a short description of Maximum Likelihood and maybe something about biases of non-continuous measurements.
    This paragraph is mainly supposed to give general guidelines for our method and to replace it in its theoretical context. Though lengthy, we consider this point necessary, so that the reader is aware of our approach.
17. Page 4781, line 14: do not prevent?? I think “do prevent”.
    Technical issues prevent continuous measurements. We generate virtual data when measurements are not prevented.

18. Page 4782, line 11: cause a variability ! cause variability
Corrected.
19. Page 4782, line 13: infer back ! infer
Corrected.
20. Page 4782, line 21: please define that x refer to the fluxes.
The state vector x does not only refer to the fluxes. For example, in the case of a limited domain, boundary concentrations also have to be included in this vector. It is now clarified in the text.
21. Page 4783, line 2 and 3: Is assimilated really what you mean here? I think converted is better.
Using “assimilated” in its common sense in a data assimilation framework may be clumsy. We corrected the use of terms.
22. 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.
Turbulence is indeed linear from a physical point of view. However, the way local turbulence is numerically dealt with by the transport model might lead in some cases to non-linearity. Large scales average the non-linearities which then vanish. The cited sentence is replaced by: “at scales large enough, so that the treatment of the local scale turbulence by the model does not generate numerical non-linearity”.
23. Page 4783, line 21: Also, here it is probably wise to state that matrix R refers to the uncertainties in both the observation and the projection of the fluxes to the measurement space (Hx). So, model representation errors are in R.
This is indeed the case. We clarify the statement in the new version of the manuscript.
24. p. 4783, l. 23: "apart from the technical issues in the implementation of the theory on computers" ! "apart from technical issues in the numerical

implementation of the theory"
Modified.
25. Page 4784, line 2: purely!pure
Corrected.
26. p. 4784, l. 4: "tuple" seems a little bit pedantic if you just meant "couple" (python language practitioners?)
Python language indeed might have influenced our wording. Modified in the body of the manuscript.
27. Page 4784, line 19: said tuple?? What do you mean by said? The tuple mentioned above?
The ambiguity is removed.
28. Page 4784, line 21: a local dependence? Unclear what is meant here. If I understand this well, it represents the posterior solution for a specific choice (R,B).
This is correct. We clarify this in the new manuscript.
29. 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 B just as you do, but choose for p(B) the Jeffreys' distribution (usually called an hyper-prior). Bocquet (2011) is also of interest to you because the marginalisation also bears on xb, hence the bias.
This part was indeed unclear and not mandatory in our approach description. In the updated version of the manuscript, section 2.1.3 describes the historical maximum likelihood before we introduce marginalization concepts.

The link to the marginalization is then more natural. Technical details on the maximum likelihood are transferred to the cited literature. We agree with the reviewer's statement on this point. However, we still rely on a uniform approximation of the hyperparameter distribution to drastically simplify the considerations to follow (as developed by Michalak et al., 2005).

30. 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 (xb;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.
The integrated sum of weighted normal distribution is indeed a multivariate T-distribution. We corrected our statement. The T-distribution information could help in reducing the size of the Monte Carlo ensemble in future steps of our work.
31. Page 4785, line 19: located to ! located at. It is unclear here, what is meant with “dummy tuple”. I understand that for each (R,B) you can calculate an x_a, P_a, but what do you mean by “dummy”.
The word “dummy” designate a variable which is used only in the calculus in is inside; for example, varying parameters inside an integral. We avoid this word in the new manuscript to avoid an misunderstanding.
32. Page 4785, line 21: Unclear what is a sample here: do you mean one particular (R,B)? Also, what is mean by “local” vectors x_a. Are there also non-local vectors? I see the symmetry with respect to x_a, but I do not see why the P_a drops out of the sampling procedure.
p. 4785, l. 17-21: I did not get your point. Please clarify.
The use of words is now clarified. Every vector $\vec{x}^a$ in the integral is a func-

tion of the variables (R,B) within the integral. The symmetric statement was indeed unclear. It is now replaced by:
“Each sample of the ensemble must take into account the spread of $\mathcal{N}(\widetilde{\vec{x}^a}, \widetilde{\mathbf{P}^a})$ in Eq. 3. To do so, we describe the *pdf* $p(\vec{x}|\vec{y}^o - \mathbf{H}\vec{x}^b, \vec{x}^b)$ not from the ensemble of posterior fluxes $\left(\widetilde{\vec{x}^a}\right)$, but from a perturbed ensemble of $\left(\widetilde{\vec{x}}\right)$, with each $\widetilde{\vec{x}}$ a random sample of $\mathcal{N}(\widetilde{\vec{x}^a}, \widetilde{\mathbf{P}^a})$.”
33. 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.
Actually, we should have clarified that we were considering cases with a very limited number of observations. This is now discussed in a separated part prior to the marginalization presentation, section 2.1.3.
34. Page 4786. The pdf of (R,B). After equation (2) it is stated that “we assume no prior information of the uncertainty matrices”. Here a chi-squared distribution is assumed. Please explain this better.
Our statement was misleading on this point. The χ^2 distribution is deduced from the integration of both unknown prior information on the hyperparameters (e.g., expected range of magnitude for variances) and of information in the observations. This is developed in the references we cite (Dee, 1995; Chapnik et al., 2004; Michalak et al., 2005; Winiarek et al., 2012). Section 2.1.3 now presents the maximum likelihood approach in a clearer way with a natural reference to the χ^2 distribution.
35. Page 4786, line 6: This part of the paper becomes very messy. You are in the middle of a “theory” section, in which the Monte-Carlo method is explained. Here this is mixed up with a “method” section, in which a practical choice is motivated (only diagonal matrices (R,B)). A whole section is now devoted to the effect of diagonal matrices on the inversion (“too optimistic a reduction of uncertainties on the fluxes” ! “a too optimistic reduction of the

uncertainties on the fluxes"), and the reduction of the state space, with a reference to section 3. I suggest separating the "method" and "theory" in a better way, i.e. to move the particular setting to section 3. Given the more practical application that follows, it seems logical to stop section 2 here and start section 3 with the Maximum Likelihood (figure 1).
We agree with the messy impression given by this part. As a consequence, we rethought the general structure of this part. It is now separated into subparts with distinct elements (theory, motivation, technical points, etc.). The detailed new structure of the manuscript is presented in introduction of this document.

36. Page 4786, line 23: infered ! inferred Here it is claimed that the Maximum Likelihood choice for (R,B) would overestimate the error reduction in the case of diagonal matrices (R,B). However, a valid question here is what is wrong with the Maximum Likelihood solution using full matrices (R,B), since this would provide a realistic solution to the inverse problem.
The maximum likelihood indeed provides a realistic solution to the inverse problem if we consider only the optimized fluxes and not the uncertainties on the fluxes. However, in frameworks with very scarce observations, the uncertainties on the computation of the maximum likelihood can have dramatic impact on the flux posterior uncertainties. This is the main motivation for our approach and it is now detailed in part 2.1.3.

37. p. 4787, l. 16-26: Even with the diagram the algorithm is still not clear enough to me. How do you sample the diagonal R and 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?
This is now clarified in Section 2.2.2. For each diagonal element of R or B, the pdf to be sampled is a χ^2 distribution described in Eq. 4. We use 60000 draws in our Monte Carlo. This is a critically high number if no size reduction is applied on the problem. However, the T-distribution representation of the

integrated pdf should help in reducing the number of draws in the further developments.

38. Page 4787, line 23: "each others" ! "each other"
Corrected.

39. Page 4788, line 10: reduce ! reduced; damp ! dampen; shall!should The discussion here links nice to the part on diagonal (R,B) matrices in section 2.
Modified.

40. Page 4788, line 15: "physical". I would remove this word.
Removed.

41. Page 4788, line 17: straighter ! straightforward
Corrected.

42. Page 4788, line 18: space ! spaces
Corrected.

43. Page 4788, line 26: "a number of pieces of data" ! "a number of data points"
Corrected.

44. Page 4789, line 20: "in order to inquire into the" ! "in order to study"
Ok.

45. Page 4789, line 26: said ! processed
Ok.

46. Page 4790, line 1: write ! derive
Modified.

47. Page 4790, eq. 5: The term E_w seems misprinted (or needs better introduction).

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

Is that intentional? The equation is rewritten with sub-indexes for clarity and the awkward rendering is corrected.

49. p. 4789, 4790: There was a reason why Bocquet et al. (2011) did not choose any P or Π (but Γ) for what you designate as Π. 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.
We did not realize the possible implication of the choice of Π. Considering your statement, we now uniformly use Γ.
50. p. 4791, l. 20-22: "For this reason, we decide to define ...": Is this "Λ" a linear operator? If no, does this invalidate the use of your equations?
The definition of Λ was indeed ambiguous. We now replace it by: "which, for each day and observation site, selects the component of the observation vector when the daily minimum of concentrations within a planetary boundary layer higher than 500 m is observed"
Then, Λ literally picks values from a designated hour of the day (corresponding in general to the moment when the PBL height is maximal). Thus, this operator is indeed linear, which is critical for the simplicity of our formulae.
51. Page 4793, line 4: depicts ! represents
Modified.
52. 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 ! is computed for a fixed global network).
Thank you for noticing this mistake in computing the references.
53. 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?
54. p. 4795, l. 15: "a Pseudo-Newtonian ascending algorithm": !? Do you mean

"a quasi-Newton descent method"?
This is what we mean. We correct the wording when referring to Newton method.
55. 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 KH.
Thank you for this new insight. We succinctly include it in section 3.3.
56. p. 4794, l. 20: What is a "Fisher-like" distribution?
Actually, we were thinking about a Fischer-Snedecor distribution. This is clarified in the text.
57. p. 4795, l. 23: "marginalize inversion" ôĂĂĂ! "marginalized inversion"
Corrected.
58. 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).
This over-statement is now replaced by:
"The main difference with most other atmospheric inversions resides into the objective and automatic computation of the influence of ill-specified error statistics, in contrast with the traditional assigning of frozen error matrices based on expert knowledge and with the more recent online computations of error hyperparameters"
59. 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.
We indeed prefer to put representativeness errors to zero as they are hard to catch with OSSEs. This is clarified in new sub-part 4.1.2.

60. Page 4797, line 24: the z_abs is not really clear to me. It will lead to an asymmetric distribution of errors, at least for positive emissions. This is also illustrated in figure 5, which shows large values of z_abs (all due to overestimates???).
z_abs is the absolute score. It depicts the distance (in Tg) between the true emissions and the retrieved ones. We chose this score as the final physical point we are interested in is to recover realistic absolute emissions at the regional scale.
61. p. 4798, l. 4: "Monte-Carlo tuples" ! "Monte Carlo draws" (note the absence of dash in English).
Corrected. The dash for Monte Carlo has been removed throughout the entire text.
62. Page 4798, line 10: But each inversion that provides a posterior error covariance matrix can be used to calculate these scales, am I right? So also the ML method without marginalization?
Actually, only analytical inversions explicitly provide posterior error covariance matrices. Analytical inversions rarely rely on an objective choice of the representation, hence they suffer from high aggregation errors. Variational inversions require further computation to deduce posterior errors. These computation mostly relies on Monte Carlo sampling on very small ensembles (as each variational inversion is critically resource demanding). The posterior errors in a variational framework are then characterized with high uncertainties in most cases.
The maximum likelihood could provide a good approximation of the covariance matrix, though with under-estimated variances. We then prefer using

the final ensemble to compute the posterior uncertainties and correlations.

63. Page 4798, line 24: couple ! couples
Modified.
64. Page 4798, bottom: the role of BCs remains rather vague. What criteria are used to flag?
The only criterion for flagging a group is the existence or not of any contribution from LBC. This is done for every possible grouping pattern, i.e. for each correlation threshold.
65. p. 4799, l. 5: "obervation" ! "observation"
Corrected.
66. Page 4799, line 12: ays ! days
Corrected.
67. Page 4799, line 14: oxydation ! oxidation
Corrected.
68. Page 4800, line 24: punctual!localized Unclear also if wildfire emissions are included now or not. I assume they are not optimized. So why mention these emissions?
We tried to include fire emissions in our system. But the hot spot filtering actually filters out all the fire contributions. At the end, no fire emissions are integrated, but they are nevertheless taken into account as they contribute in the reduction of information in the system (which is unavoidable as fires influence the observation network we use). This is now briefly clarified in section 5.1.1.
69. 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".
We indeed use point-wise multiplication. This is now clarified.

70. Page 4801: Here, it should be mentioned on which timescale the emissions are allowed to vary. I guess 10 days, like the LBCs, but I could not find it. The pseaudo-data or obtained using an inversion with real data. Now I wonder why (i) not to use real data in the framework (ii) in what respect the simulated data reflect already large biases in the system. Anyhow, this is one of the part of the paper that needs rethinking. It is hard to understand what is (i) a raw inversion (ii) how well the inversion is capable to reproduce the data. We never see any simulated of measured time-series in the paper.
LBC are perturbed on a 10-basis, as well as wetland emissions. Anthropogenic emissions are perturbed on a monthly basis. This is clarified in section 5.1.2.
We do not use real data in the OSSEs as we would not have been able to control the true emissions and the representativeness errors.
We use real data to infer the correction factors in order to get potentially realistic variations in the emissions. The raw inversion is an expert-knowledge based inversion, i.e. with a unique frozen couple of uncertainty matrices (R, B).
As the observation and state spaces are modified during the computation of our system, it is not any more relevant to represent observation time series. This is the main reason why we define general scoring to evaluate our inversion system. This is made possible by our control on the natural run and on the 'true' emissions in OSSEs.
71. 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?
Actually, we use the term non-hydrostatic as the constraining wind fields do not need to rely on the hydrostatic hypothesis.
72. Page 4803: top. Here biases are discussed. However, hardly any conclusions are drawn concerning biases. This makes the paper lengthy and

messy.
Biases are critical in the inversion framework. We decided to neglect them in the theoretical framework we developed. However, as they represent a critical point of data assimilations, we preferred mentioning them here and there (sections 2.1.2., 2.2.2., 5.2., 5.2.1., 6.1.1., 6.1.2., 6.3.2.). These references to biases are now better integrated into the body of the manuscript.
73. Page 4805: r_max could be determined by one region, e.g. a large correlation between emissions in one region and a neighboring region. Should r_max not reflect an average correlation between state vector elements?
r_max is chosen before post-processing the Monte Carlo ensemble. Scores are computed for a prescribed r_max. At the end, we decide to chose r_max = 0.5 as this value balances the discussed effects in the scores.
74. Page 4806: discussion is very detailed (e.g. bias vs. filtering). I loose the view on the most important aspects of the paper.
We added sub-sections and the links in the discussion so it is easier to follow.
75. Page 4807, line 4: dominates ! dominate
Corrected.
76. Page 4808, line 19: the closest to the observation network: unclear sentence.
The sentence is now clarified.
77. Page 4809: I like the comparison with the “frozen” error matrices. I miss however the connection with the earlier statement that this leads to an “underestimation of errors” when diagonal matrices are used. Anyhow, you should specify here whether the matrices are diagonal and the grid is reduced. A comparison with the classical inversions should employ the classical correlations I guess, and also present the ML solution.
Fig. 1 shows the ML solution compared to the Monte Carlo one. In this fig-

ure, we see the interest of our method in terms of proper estimation of error magnitude. The post processing with correlation thresholds r_max logically leads to the same kind of score profiles. But as the scores are significantly higher, it proves the importance of using optimized matrices (based on the ML). This is developed in section 6.3.1.

78. 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.
    This should be tested in the next steps of this work. However, for the moment, Computation costs may prevent us from doing it.

79. p. 4810, 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.
    We clarified our contribution to the field in the conclusion section.

80. p. 4811, l. 6: "virtual truth": usually called "nature run" using OSSEs' terminology.
    We now use this terminology here and in other places in the text for better consistency with the community.

81. Page 4825: component ! components
    Corrected.

82. Page 4826: plain! solid
    Modified.

References

Chapnik, B., Desroziers, G., Rabier, F., and Talagrand, O.: Properties and first application of an error-statistics tuning method in variational assimilation, Q. J. R. Meteorol. Soc., 130, 2253–2275, doi:10.1256/qj.03.26, http://onlinelibrary.wiley.com/doi/10.1256/qj.03.26/abstract, 2004.

Dee, D. P.: On-line estimation of error covariance parameters for atmospheric data assimilation, Mon. Weather Rev., 123, 1128–1145, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.53.7382&rep=rep1&type=pdf, 1995.

Michalak, A. M., Hirsch, A., Bruhwiler, L., Gurney, K. R., Peters, W., and Tans, P. P.: Maximum likelihood estimation of covariance parameters for Bayesian atmospheric trace gas surface flux inversions, Journal of Geophysical Research: Atmospheres, 110, n/a–n/a, doi:10.1029/2005JD005970, http://onlinelibrary.wiley.com/doi/10.1029/2005JD005970/abstract, 2005.

Winiarek, V., Bocquet, M., Saunier, O., and Mathieu, A.: 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, D05 122, doi:10.1029/2011JD016932, http://onlinelibrary.wiley.com/doi/10.1029/2011JD016932/abstract, 2012.

Please also note the supplement to this comment:
http://www.geosci-model-dev-discuss.net/7/C3442/2015/gmdd-7-C3442-2015-supplement.pdf

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