Final answer to the reviews of the manuscript
"Objectified quantification of uncertainties in Bayesian atmospheric inversions"

Berchet A.[1]**, Pison I.**[1]**, Chevallier F.**[1]**, Bousquet P.**[1]**, Bonne J.-L.**[1] **and Paris J.-D.**[1]

[1]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, IPSL, Gif-sur-Yvette, France.

Correspondence to: A. Berchet (aberchet@lsce.ipsl.fr)

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.

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 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.

5

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 hyperprior). 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 $\mathbf{x}^a$ in the integral is a function 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 $\mathcal{N}(\widetilde{\mathbf{x}^a}, \widetilde{\mathbf{P}^a})$

9

in Eq. 3. To do so, we describe the pdf $p(\mathbf{x}|\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}, \mathbf{x}^{\mathrm{b}})$ not from the ensemble of posterior fluxes $\left(\widetilde{\mathbf{x}^{\mathrm{a}}}\right)$, but from a perturbed ensemble of $(\widetilde{\mathbf{x}})$, with each $\widetilde{\mathbf{x}}$ a random sample of $\mathcal{N}(\widetilde{\mathbf{x}^{\mathrm{a}}}, \widetilde{\mathbf{P}^{\mathrm{a}}})$."

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 (**??**Michalak et al., 2005; **?**). 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 sub-parts 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.

11

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.

13

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" ! "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 emis-

14

sions 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.

15

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

16

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.

17

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 figure, 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 assimila-
    tion, 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.

Objectified quantification of uncertainties in Bayesian atmospheric inversions

Berchet A.[1,*], Pison I.[1], Chevallier F.[1], Bousquet P.[1], Bonne J.-L.[1] and Paris J.-D.[1]

[1]Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, IPSL,
Gif-sur-Yvette, France.
[*]Now at Laboratory for Air Pollution/Environmental Technology, Swiss Federal Laboratories for
Materials Science and Technology, Empa, Dübendorf, Switzerland.

Correspondence to: A. Berchet (antoine.berchet@empa.ch)

Abstract. Classical Bayesian atmospheric inversions process atmospheric observations and prior emissions, the two being connected by an observation operator picturing mainly the atmospheric transport. These inversions rely on prescribed errors in the observations, the prior emissions and the observation operator. When data pieces are sparse, inversion results are very sensitive to the prescribed error distributions, which are not accurately known. The classical Bayesian framework experiences difficulties in quantifying the impact of mis-specified error distributions on the optimized fluxes. In order to cope with this issue, we rely on recent research results to enhance the classical Bayesian inversion framework through a marginalization on a large set of plausible errors that can be prescribed in the system. The marginalization consists in computing inversions for all possible error distributions weighted by the probability of occurrence of the error distributions. The posterior distribution of the fluxes calculated by the marginalization is not explicitly describable. As a consequence, we carry out a Monte Carlo sampling based on an approximation of the probability of occurrence of the error distributions. This approximation is deduced from the well-tested method of the maximum likelihood estimation. Thus, the marginalized inversion relies on an automatic objectified diagnosis of the error statistics, without any prior knowledge about the matrices. It robustly accounts for the uncertainties on the error distributions, contrary to what is classically done with frozen expert-knowledge error statistics. Some expert knowledge is still used in the method for the choice of an emission aggregation pattern and of a sampling protocol in order to reduce the computation cost. The relevance and the robustness of the method is tested on a case study: the inversion of methane surface fluxes at the meso-scale with virtual observations on a realistic network in Eurasia. Observing System Simulation Experiments are carried out with different transport patterns, flux distributions and total prior amounts of emitted methane. The method proves to consistently reproduce

the known 'truth' in most cases, with satisfactory tolerance intervals. Additionally, the method explicitly provides influence scores and posterior correlation matrices. An in-depth interpretation of the inversion results is then possible. The more objective quantification of the influence of the observations on the fluxes proposed here allows us to evaluate the impact of the observation network on the characterization of the surface fluxes. The explicit correlations between emission aggregates reveal the mis-separated regions, hence the typical temporal and spatial scales the inversion can analyse. These scales are consistent with the chosen aggregation patterns.

1 Introduction

Characterizing the global biogeochemical cycles of greenhouse gases requires a reliable understanding of the exchanges at the surface-atmosphere interface. The description of these exchanges must encompass the absolute amounts of gas released to and removed from the atmosphere at the surface interface, the spatial distribution and the temporal variability of the fluxes, and the determination of the underlying physical processes of emissions and sinks. Such an integral depiction is still missing for most greenhouse gases (Ciais et al., 2013). One of the possible approaches to inquire into the surface fluxes is the analysis of the atmospheric signal. The drivers of the spatial and temporal variability of the atmospheric composition are atmospheric transport, chemistry and surface fluxes. Therefore, monitoring the atmospheric composition and using a representation of the atmospheric transport and chemistry with Global Circulation Models (GCMs) or Chemistry-Transport Models (CTMs) can help in inferring information on the fluxes (Bousquet et al., 2006; Bergamaschi et al., 2010). This approach, called atmospheric inversion, suffers from two practical issues in its implementation. First, the atmospheric composition is still laconically documented, though the number of global monitoring projects with extensive surface observation networks and satellite platforms has been increasing in the last decades (e.g., Dlugokencky et al., 1994, 2009). Indeed, the satellite platforms have a global coverage but the observed atmospheric composition is integrated over the vertical column, while the surface sites can provide continuous observations but only at fixed point locations. Second, the atmosphere behaves as an integrator and the air masses are mixed ambivalently through the transport (Enting et al., 1993). Thus, the inverse problem of tracking back the fluxes from the variability of the atmospheric composition cannot be solved univocally. The Bayesian formalism allows statistical analyses of the atmospheric signal, so that one can identify confidence intervals of fluxes compatible with the atmospheric composition (Tarantola, 1987).

Bayesian inversions have been extensively used at the global scale, providing insights on the greenhouse gas budgets (e.g., Gurney et al., 2002; Kirschke et al., 2013; Bergamaschi et al., 2013). However, non compatible discrepancies in the results appear between the possible configurations of atmospheric inversion systems (Peylin et al., 2013). The various configurations include the choice of the atmospheric transport, its spatial and temporal resolutions, the meteorological driving fields, the

2

type and density of the observations, etc. In the Bayesian formalism, some assumptions also have to be made on the transport model error statistics, on the errors made when comparing a discretized model to observations (Geels et al., 2007) and on the confidence we have on the prior maps and time profiles of emissions (Enting, 2002). All these choices are based on technical considerations and on the expert perception of the problem to solve. Comparing results based on different choices that are physically adequate, but subjective, is difficult, especially to track inconsistencies, which enlarge the range of flux estimates.

In the following, we focus on the development of an enhanced Bayesian method that objectifies the assumptions on the statistics of the errors and that takes the unavoidable uncertainties generated by our lack of knowledge on these error statistics into account. In this approach, the confidence ranges of the optimized surface fluxes are computed by a Monte Carlo marginalization on all the possible error statistics, which is more general than the usual Bayesian approach deducing posterior uncertainties from a single error statistic combination only. The weight function for the marginalization is inferred from an already-tested maximum likelihood approach (e.g., Dee, 1995; Michalak et al., 2005), processing the pieces of information carried by the differences between the measurements and the prior simulated concentrations. The potential and consistency of the method is tested through Observing System Simulation Experiments (OSSEs) on a realistic configuration of atmospheric inversion.

The case study is the quantification of methane fluxes in the Siberian Lowlands with a network of surface observation sites that have been operated for a few years by the Japanese National Institute for Environmental Studies (Sasakawa et al., 2010) and the German Max Planck Institute (Winderlich et al., 2010). The characterization of the region is challenging, with co-located massive methane emissions from anthropogenic activity (oil and gas extraction) and from wetlands in summer. Moreover, the wetland emissions have a very high temporal variability (due to their sensitivity to the water table depth and to the temperature; e.g., Macdonald et al., 1998; Hargreaves and Fowler, 1998). Their quantification is then difficult. In order to catch the influence of the sampling bias due to non-regularly distributed observation sites and non-continuous measurements, we produce virtual observations from a known 'truth' at locations where real observations are carried out and at dates when the logistical issues do not prevent the acquisition of measurements. We then check the capability of our method to reproduce consistent flux variability and distribution with seven degraded inversion configurations (perturbed transport, flat flux distributions, etc.).

In Sect. 2, we describe the theoretical framework of our method of marginalization. The enhancements on the general theoretical framework need a cautious definition of the problem to be implementable in term of computational costs and memory limits. In Sect. 3, guidelines for a suitable definition of the problem are developed. The whole structure of the method is summarized in Sect. 4.1. In Sect. 4, we present the particular set up of the OSSE carried out for proving the robustness of the

3

method. The specific Siberian configuration we test our method on is detailed in Sect. 5. The OSSE
are evaluated along defined objective statistical scores in Sect. 6.

2 Marginalized Bayesian inversion

We first describe the motivations for using a marginalized inversion in Sect. 2.1. In Sect. 2.2, we
describe the marginalization itself and the Monte Carlo approach chosen in order to compute it.

2.1 Context and motivation for the marginalization

2.1.1 Bayesian inversion framework

The surface-atmosphere fluxes, through transport, cause a variability in the atmospheric mixing ra-
tios of the species we are interested in. The atmospheric inversion relies on the processing of the
atmospheric variability in order to infer the surface-atmosphere fluxes. Since the atmosphere is dif-
fusive and irreversibly mixes air masses from different origins, it is physically impossible to infer
univocal information on the fluxes from the integrated atmospheric signal alone (Tarantola, 1987;
Enting, 2002). We then pursue a thorough characterization of the pdf of the state of the system $\mathbf{x}$
(e.g., the spatial and temporal distribution of the surface fluxes, but also background concentrations
and baselines in some cases), assuming some prior knowledge on the system and having some ob-
servations of the atmospheric physical variables related to our problem. That is to say, we want to
calculate the pdf $p(\mathbf{x}|\mathbf{y}^{\mathrm{o}} - H(\mathbf{x}^{\mathrm{b}}), \mathbf{x}^{\mathrm{b}})$ for all possible states $\mathbf{x}$; $\mathbf{y}^{\mathrm{o}}$ is a vector gathering all the avail-
able observations, $\mathbf{x}^{\mathrm{b}}$ is the background vector including the prior knowledge on the state of the
system and H is the observation operator converting the information in the state vector to the obser-
vation space. Typically, H embraces the atmospheric transport and the discretization of the physical
problem. In the scope of applications of the atmospheric inversions, the observation vector $\mathbf{y}^{\mathrm{o}}$ gath-
ers measurements of dry air mole fraction. As for the observation operator, it is computed with a
model which simulates mixing ratios. As we are interested in trace gases, we will consider that the
dry air mole fractions can be treated as mixing ratios. In all the following, we also consider that H is
linear; hence, H is represented by its Jacobian matrix $\mathbf{H}$ and $H(\mathbf{x}^{\mathrm{b}}) = \mathbf{H}\mathbf{x}^{\mathrm{b}}$. This approximation is
valid for all non reactive atmospheric species at scales large enough, so that the treatment of the lo-
cal scale turbulence by the model does not generate numerical non-linearity. When the atmospheric
chemistry must be taken into account (for instance with methane), either the window of inversion
must be short compared with the typical lifetime in the atmosphere for the linear assumption to be
valid, or the concentration fields of the reactant species (e.g., OH radicals for methane) must be
accurately known.

In general, the characterization of the pdf is built within the Bayesian formalism with the as-
sumption that all the involved pdfs are normal distributions (Enting et al., 1993). The pdfs are
then explicitly described through their mode and their matrix of covariance. In this case, the pdf

$p(\mathbf{x}|\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}, \mathbf{x}^{\mathrm{b}}) \propto \mathcal{N}(\mathbf{x}^{\mathrm{a}}, \mathbf{P}^{\mathrm{a}})$ is defined by its mode, $\mathbf{x}^{\mathrm{a}}$, the posterior state, and its matrix of covariance, $\mathbf{P}^{\mathrm{a}}$. In addition to the linear assumption, we also consider that the uncertainties are unbi-

130   ased. That is to say: $p(\mathbf{x} - \mathbf{x}^{\mathrm{b}}) \propto \mathcal{N}(\mathbf{0}, \mathbf{B})$ and $p(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{t}}) \propto \mathcal{N}(\mathbf{0}, \mathbf{R})$ where $\mathbf{x}^{\mathrm{t}}$ is the true state of the system. The uncertainty matrix $\mathbf{B}$ (resp. $\mathbf{R}$) encompasses the uncertainties on the background $\mathbf{x}^{\mathrm{b}}$ (resp. on the measurements and on the model, including representation errors, i.e. the errors made when approximating the real world by a numerical gridded model). Under these assumptions, we can explicitly write the posterior vector and the posterior matrix of covariance:

135
$$p(\mathbf{x}|\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}, \mathbf{x}^{\mathrm{b}}) \propto \mathcal{N}(\mathbf{x}^{\mathrm{a}}, \mathbf{P}^{\mathrm{a}}) : \left\{ \begin{array}{ll} \mathbf{x}^{\mathrm{a}} & = \mathbf{x}^{\mathrm{b}} + \mathbf{K}(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{b}}) \\ \mathbf{P}^{\mathrm{a}} & = \mathbf{B} - \mathbf{K}\mathbf{H}\mathbf{B} \end{array} \right. \tag{1}$$

with $\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}})^{-1}$ the Kalman gain matrix.

2.1.2 Ambivalent uncertainty set-up

Atmospheric inversion is straightforward (apart from technical issues in the numerical implementation of the theory) as long as the uncertainty matrices $\mathbf{R}$ and $\mathbf{B}$ are defined.

140   Some of their components can be calculated unambiguously, such as measurement errors in matrix $\mathbf{R}$. Other errors are derived, in most cases, following expert knowledge on, e.g., the behaviour of the atmospheric transport and of the surface fluxes. This expert knowledge is acquired, for example, through extensive studies on the sensitivity of the transport model to its parametrization and forcing inputs (e.g., Denning et al., 1999; Ahmadov et al., 2007; Lauvaux et al., 2009; Locatelli et al., 2013),

145   or by comparing prior fluxes to measured local fluxes (e.g., Chevallier et al., 2006). Some studies also rely on pure physical considerations (e.g., Bergamaschi et al., 2005, 2010).

However, the complex and unpredictable structure of the uncertainties is hard to reproduce accurately from the expert knowledge alone and an ill-designed couple of uncertainty matrices $(\mathbf{R}, \mathbf{B})$ can have a dramatic impact on the inversion results (e.g., Berchet et al., 2013; Cressot et al., 2014).

150   The discrepancies between the possible configurations of inversion can also reveal some biases, η, in the models: in that case $p(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{t}}) \propto \mathcal{N}(\eta, \mathbf{R})$ instead of $p(\mathbf{y}^{\mathrm{o}} - \mathbf{H}\mathbf{x}^{\mathrm{t}}) \propto \mathcal{N}(\mathbf{0}, \mathbf{R})$, which would require a different handling of Eq. 1. For example, the horizontal wind fields can be biased or the vertical mixing in the planetary boundary layer systematically erroneous. That makes it difficult to compare simulated concentrations in the boundary layer to measurements (e.g., Peylin et al., 2002;

155   Dee, 2005; Geels et al., 2007; Williams et al., 2013; Lauvaux and Davis, 2014). Biases can have critical impacts on inversion results and must be inquired into independently (e.g., Bocquet, 2011). Nevertheless, for our study, we decide to neglect the biases in the inversion. We discuss in Sect. 6.3 the potential impacts of biases that are not significant in our specific application. We then focus only on the mis-specification of the uncertainty matrices $\mathbf{R}$ and $\mathbf{B}$.

2.1.3 Possible uncertainty handling

In order to address the uncertainty issue in atmospheric inversions, efforts are carried out towards objectifying the way the error statistics are chosen (e.g., Schwinger and Elbern, 2010; Winiarek et al., 2012; Berchet et al., 2013). These efforts focus on specific algebraic properties of the uncertainty matrices (e.g., Desroziers and Ivanov, 2001; Desroziers et al., 2005) or more generally on understanding the likelihood of the prior innovation vector, $\mathbf{y}^o - \mathbf{Hx}^b$, as a function of the uncertainty matrices (Dee, 1995). Under Gaussian assumptions, the likelihood of the innovation vector can be written:

$$p(\mathbf{y}^o - \mathbf{Hx}^b | \mathbf{R}, \mathbf{B}, \mathbf{x}^b) = \frac{e^{-\frac{1}{2}(\mathbf{y}^o - \mathbf{Hx}^b)^T(\mathbf{R} + \mathbf{HBH}^T)^{-1}(\mathbf{y}^o - \mathbf{Hx}^b)}}{\sqrt{(2\pi)^d |\mathbf{R} + \mathbf{HBH}^T|}} \tag{2}$$

with d the dimension of the observation space and $|\cdot|$ the determinant operator.

In the likelihood framework, the couple of uncertainty matrices $(\mathbf{R}, \mathbf{B})$ that maximizes Eq. 2 is considered as optimal and will be hereafter referred to as the maximum likelihood. This maximum likelihood optimally balances the observation and prior state error variances and covariances according to the prior innovation vector $\mathbf{y}^o - \mathbf{Hx}^b$ (Chapnik et al., 2004). A direct algorithm computing the maximum likelihood (applied to atmospheric inversion in, e.g., Winiarek et al., 2012; Berchet et al., 2013) is then supposed to provide a good approximation of the couple of optimal matrices $(\mathbf{R}_{max}, \mathbf{B}_{max})$ which can be used forward in the inversion (Eq. 1). In order to dampen the computation cost of the maximum likelihood estimation, most studies just maximize the likelihood on hyperparameters (e.g., correlation lengths), describing the couple of matrices $(\mathbf{R}, \mathbf{B})$ in a more simple way.

Though general, the estimation of the innovation vector maximum likelihood relies on strong assumptions, it can suffer from strong numerical errors and it is not necessarily univocal. More explicitly, as showed by previous works, the *pdf* of the uncertainty matrices $p(\mathbf{R}, \mathbf{B})$ behaves as a χ^2 distribution with d degrees of freedom, d being the dimension of the observation space. Thus, the likelihood is highly dominated by the mode of $p(\mathbf{R}, \mathbf{B})$, co-located with the maximum likelihood. However, the peaked likelihood argument may be too rough in some cases. As the number of observations decreases compared to the number of state dimensions, this optimal case becomes less univocal. In the frameworks where observations are too scarce, the maximum likelihood may lead to flawed results. To assess the validity of the peak assumption, estimations of the Hessian matrix of the likelihood at its maximum have been used (e.g., Michalak et al., 2005; Wu et al., 2013). Hessian matrices give the magnitude of the uncertainties on the computation of the uncertainty matrices. Nevertheless, to our knowledge, no atmospheric inversion account for the impact of the Hessian matrix of the likelihood on the inversion results.

In addition, even when the *pdf* $p(\mathbf{R}, \mathbf{B})$ is intensely peaked at its maximum, the inferred inversion results from Eq. 1 with a direct maximum likelihood algorithm would erroneously under-estimated uncertainties on the result (see Fig. 1 and, e.g., Berchet et al., 2013). Indeed, at the maximum like-

lihood, all the pieces of information in the system are considered perfectly usable by the inversion which then gives too optimistic posterior uncertainties in this case.

2.2 Marginalization of the inversion

2.2.1 Theoretical formulation

200 Here, we compute the *pdf* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ by a marginalization on the uncertainty matrices to comprehensively account for the uncertainties in the characterization of the uncertainties and to quantify the impact of ill-specified uncertainty matrices. In statistics, marginalizing a probability density function (*pdf*) $p(\mathbf{x})$ consists in rewriting it as a sum of conditional probabilities $p(\mathbf{x}|\mathbf{z})$ weighted by $p(\mathbf{z})$.

205 Thus, the complete *pdf* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ classically described by Eq. 1 is separated into a sum of the contribution of each possible couple of covariance matrices $(\mathbf{R}, \mathbf{B})$ weighted by the probability of occurrence of the couple $(\mathbf{R}, \mathbf{B})$:

$$
\begin{aligned}
& p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \\
& = \int_{(\mathbf{R},\mathbf{B})} p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b, \mathbf{R}, \mathbf{B}) \\
& \qquad \times \quad p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \qquad \mathrm{d}(\mathbf{R}, \mathbf{B}) \\
& \propto \int_{(\mathbf{R},\mathbf{B})} \mathcal{N}(\widetilde{\mathbf{x}^a}, \widetilde{\mathbf{P}^a}) \\
& \qquad \times \quad p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b) \qquad \mathrm{d}(\mathbf{R}, \mathbf{B})
\end{aligned}
\tag{3}
$$

In Eq. 3, $\widetilde{(.)}$ depicts a dependency to the couple $(\mathbf{R}, \mathbf{B})$. The complete *pdf* $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ 210 then has the shape of an infinite sum of weighted normal distributions. This infinite sum could be described as a multi-variate T-distribution (Bocquet, 2011).

The general expression of Eq. 3 encompasses the classical case with only one couple of matrices $(\mathbf{R}, \mathbf{B})$ which considers $p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ as a Dirac-like distribution (centered at the maximum likelihood or at any expert-based couple of uncertainty matrices). More generally, $p(\mathbf{R}, \mathbf{B}|\mathbf{y}^o - 215 \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$ is not so well known as discussed in Sect. 2.1.3 above.

2.2.2 Monte Carlo sampling

Hereafter, a direct Monte Carlo characterization of Eq. 3 is carried out to deduce $p(\mathbf{x}|\mathbf{y}^o - \mathbf{H}\mathbf{x}^b, \mathbf{x}^b)$.

The Monte Carlo ensemble is to be defined along the *pdf* $p(\mathbf{R}, \mathbf{B})$, but the exact distribution 220 of the error statistics is intricate. In all the following, we then approximate the *pdf* $p(\mathbf{R}, \mathbf{B})$ by a multi-variate χ^2 distribution with d (the number of observations) degrees of freedom, centered at the maximum likelihood of the prior innovation vector (following Dee, 1995). The Monte Carlo

marginalization is consequently a direct extension of the maximum likelihood estimation now classically used in the atmospheric inversion framework.

225   The maximum likelihood can be estimated first by a quasi-Newtonian descent method. However, descent methods have high computation costs and thus require a reduced number of hyperparameters (variances, correlation lengths, etc.) to describe the full uncertainty matrices. From here, we decide to reduce the distribution of the matrices $(\mathbf{R}, \mathbf{B})$ to the subspace of the diagonal positive matrices. Using a subspace of the possible error statistics can dampen the generality of the method. In particular, error correlations will be excluded with diagonal uncertainty matrices. Correlations can be used in some frameworks to detect the biases in the system (Berchet et al., 2013). But, more importantly, correlations of observation or background errors can indicate redundant pieces of information in the inversion system. For instance, an inversion computed with no observation correlation tries to use too much information and is expected to give too optimistic a reduction of uncertainties on the fluxes. Nevertheless, in Sect. 3, we reduce the observation and state spaces in order to numerically compute the Monte Carlo marginalization. The reduction of the observation and state spaces indirectly depicts correlations in the full-resolution system. In this configuration, the correlation issue is then attenuated and the diagonal assumption is valid.

At the end, for each diagonal term of the uncertainty matrices $(\mathbf{R}, \mathbf{B})$, we prescribe a χ^2 distribution with d (i.e. the dimension of the observation space) degrees of freedom, rescaled so that its average equals the associated term in the computed maximum likelihood couple $(\mathbf{R}_{\max}, \mathbf{B}_{\max})$. That is to say, for each diagonal element $\mathbf{r}^{i,i}$ of the matrix $\mathbf{R}$ (equivalently of the matrix $\mathbf{B}$):

$$p\left(\frac{\mathbf{r}^{i,i}}{\mathbf{r}^{i,i}_{\max}} \times d\right) \propto \chi^2(d) \tag{4}$$

as the mean of the χ^2 distribution with d degrees of freedom, $\chi^2(d)$, is d.

245   The χ^2 distributions are then sampled on a large ensemble – the Monte Carlo approach stabilizes after a few tens of thousands draws in our case study – to characterize the final output *pdf* $p(\mathbf{x}|\mathbf{y}^\mathrm{o} - \mathbf{H}\mathbf{x}^\mathrm{b}, \mathbf{x}^\mathrm{b})$. Each samples of the ensemble must take into account the spread of $\mathcal{N}(\widetilde{\mathbf{x}^\mathrm{a}}, \widetilde{\mathbf{P}^\mathrm{a}})$ in Eq. 3. To do so, we describe the *pdf* $p(\mathbf{x}|\mathbf{y}^\mathrm{o} - \mathbf{H}\mathbf{x}^\mathrm{b}, \mathbf{x}^\mathrm{b})$ not from the ensemble of posterior fluxes $(\widetilde{\mathbf{x}^\mathrm{a}})$, but from a perturbed ensemble of $(\widetilde{\mathbf{x}})$, with each $\widetilde{\mathbf{x}}$ a random sample of $\mathcal{N}(\widetilde{\mathbf{x}^\mathrm{a}}, \widetilde{\mathbf{P}^\mathrm{a}})$.

2.2.3   Processing the Monte Carlo posterior ensemble

In Fig. 1, we draw an example of the distribution of the Monte Carlo posterior vector ensemble along one component of the state space. The black curve depicts the posterior distribution inferred from the maximum likelihood, with under-estimated spread compared to the Monte Carlo distribution. On the opposite, as illustrated by the green curve, a Normal distribution with the same mode and the same standard deviation gives a misleading flat shape. As for a Gaussian, we then define the symmetric tolerance interval, so that 68.27% of the samples are in the range (the hatched portion of the histogram in Fig. 1). This interval is equivalent to the Gaussian $\pm\sigma$ interval, with σ the

standard deviation. One shall remind that the computed tolerance interval does not depict a Normal distribution. A Normal distribution with the same tolerance interval (the red curve in Fig. 1) is still misleadingly flat. In all the following, we will write the tolerance interval TI_{68}, $[\mathbf{x}^{\text{low}}, \mathbf{x}^{\text{high}}]$.

To summarize (as represented in the block diagram of Fig. 2), the maximum likelihood is first estimated using a quasi-Newtonian algorithm, similarly to what has been done in the literature (e.g., Winiarek et al., 2012; Berchet et al., 2013). We deduce from this maximum likelihood a plausible distribution of the uncertainty matrices $(\mathbf{R}, \mathbf{B})$. Through a Monte Carlo sampling of uncertainty matrices $(\mathbf{R}, \mathbf{B})$ along the deduced distribution, we compute an ensemble of possible posterior vectors $(\widetilde{\mathbf{x}}^{\text{a}}_{(\mathbf{R}, \mathbf{B})})$. We can then define the tolerance intervals TI_{68} and a posterior covariance matrix filled by the covariances of the ensembles of state components with each other.

Posterior covariance matrices are not always easy to compute in the atmospheric inversion framework. Here, the posterior covariance matrix is computed explicitely and objectively. The explicit definition of this matrix can give valuable information on the ability of the inversion to separate co-located emissions and emissions at different periods and locations. This capacity is used for the evaluation of the OSSEs in Sect. 4.2 and 6.

3 Informed definition of the problem

The general approach defined in Sect. 2 applies a Monte Carlo method on tens of thousands individual inversions after the completion of a maximum likelihood algorithm. This procedure requires extensive amounts of memory and computation power that can't be afforded in most real cases. For instance, the explicit computation of $\mathbf{H}$ with a Chemistry-Transport Model (CTM) closely depends on the dimension of the state space: every column of the observation operator needs one model simulation (Bousquet et al., 1999). Additionally, each step of the algorithm to compute the maximum likelihood of the prior innovation vector and each step of the Monte Carlo method relies on matrix products, matrix determinants and matrix inverses. At first sight, all these operations are as many technical issues in high dimension problems.

As a consequence, the application of the theoretically simple framework developed in Sect. 2 relies closely on an informed definition of the problem. The dimensions of the observation and state spaces should be reduced to dampen the numerical obstacles, but one shall keep resolutions physically relevant for the system we are analysing. By synthesising the recent literature on the subject, we show in the following that approximations can be reasonably applied to the full-resolution problem while not impacting the quality of the marginalized inversion results. Applying the Monte Carlo marginalized inversion is then technically feasible in a problem defined with a reduced dimension from the full-dimension problem.

9

3.1 Principle for problem reduction

3.1.1 Motivations and definition

In the observation space, more and more surface observation sites nowadays provide quasi-continuous measurements (at least a few data points per minute in the data set we use; Sasakawa et al., 2010; Winderlich et al., 2010). For long windows of inversion at the regional scale (of a few weeks or months), such a frequency of acquisition generates a number of data points technically impossible to assimilate all together in our framework. Concerning the fluxes, one shall aim at a characterization of the fluxes on very fine pixels and at a high temporal resolution. As the window of inversion lengthens and the domain widens, the number of flux unknowns grows dramatically.

In the inversion framework, the most straightforward way of minimizing the dimension of a problem is to reduce the dimensions of the observation and state spaces. Aggregating components of the state space and sampling observations are classically used for this purpose. In most studies, the reduction of the problem is carried out arbitrarily. However, aggregation can generate large errors (Kaminski et al., 2001; Bocquet et al., 2011), which would mitigate the benefits of the Monte Carlo marginalized approach compared to more classical ones applied in other atmospheric inversion studies with no aggregation (e.g., variational inversions; Courtier et al., 1994; Bergamaschi et al., 2005; Pison et al., 2009). Here, we propose a more objective way to do so following recent literature.

Using the formalism from Bocquet et al. (2011), we aim at defining a representation ω that encompasses the horizontal and temporal resolution of the fluxes, the choice of the regions of aggregation and the temporal sampling of the observations. The representation ω is defined through two operators $\mathbf{\Gamma}_\omega$ and $\mathbf{\Lambda}_\omega$, which projects respectively the full-resolution state and observation space to smaller ones. After the state space "projection" with $\mathbf{\Gamma}_\omega$, the inversion applies corrections on regions of aggregation with fixed emission distributions, instead of on single pixels. The adjoint of this operator, $\mathbf{\Gamma}_\omega^{\mathrm{T}}$, then redistributes total emissions on finer scales with the same fixed emission distribution. The choice of $\mathbf{\Gamma}_\omega$ impacts both the state vector $\mathbf{x}$ and the observation operator $\mathbf{H}$. The observation sampling $\mathbf{\Lambda}_\omega$ can consist in averaging or picking one value per time step (chosen accordingly to the physical resolution inquired into). For instance, one can decide to average the observations by day in order to study the synoptic variability of the atmosphere, related to the fluxes at the meso-scale. The observation sampling applies to both the observation vector $\mathbf{y}^{\mathrm{o}}$ and the observation operator $\mathbf{H}$. The observation operator $\mathbf{H}$ computes the contribution from single sources to single observations. The adjoint of the observation sampling, $\mathbf{\Lambda}_\omega^{\mathrm{T}}$, will then redistribute an average or a sample for each chosen time steps along this time step. The redistribution will follow the raw observed temporal profile within the processed time step.

10

3.1.2 Mathematical formulation

325   At first glance, choosing the aggregation pattern and the sampling protocol can be considered as two independent physical problems. However, as they both influence the dimension of the observation operator $\mathbf{H}$, they cannot be fixed separately. More explicitly, we can derive a formula, which links $\mathbf{\Gamma}_\omega$ and $\mathbf{\Lambda}_\omega$. Indeed, our final objective is to compute total posterior fluxes for each aggregated region that are as close as possible to the posterior fluxes from a full-resolution inversion aggregated afterwards.

330   That is to say, we want to confine the norm of $\mathbf{x}_\omega^a - \mathbf{\Gamma}_\omega \mathbf{x}_t^a$ with $\mathbf{x}_\omega^a$ the posterior state vector resolved in the representation ω and $\mathbf{x}_t^a$ the posterior state vector computed with a full-resolution representation of the problem. Algebraic manipulations lead to:

$$\mathbf{x}_\omega^a - \mathbf{\Gamma}_\omega \mathbf{x}_t^a = \mathbf{\Gamma}_\omega \mathbf{B} \mathcal{E}_\omega (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \tag{5}$$

where:

335
$$\mathcal{E}_\omega = \mathcal{P}_\omega \mathbf{H}^T \mathbf{\Lambda}_\omega^T \mathcal{S}_\omega^{-1} \mathbf{\Lambda}_\omega - \mathbf{H}^T \mathcal{S}^{-1}, \tag{6a}$$

$$\mathcal{S} = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T, \tag{6b}$$

$$\mathcal{S}_\omega = \mathbf{\Lambda}_\omega \left\{ \mathbf{R} + \mathbf{H}(\mathcal{A}_\omega + \mathcal{P}_\omega \mathbf{B} \mathcal{P}_\omega)\mathbf{H}^T \right\} \mathbf{\Lambda}_\omega^T, \tag{6c}$$

$$\mathcal{P}_\omega = (\mathbf{\Gamma}_\omega)^T \mathbf{\Gamma}_\omega, \tag{6d}$$

$$\mathcal{A}_\omega = (\mathbf{I} - \mathcal{P}_\omega)\mathbf{x}_t \mathbf{x}_t^T (\mathbf{I} - \mathcal{P}_\omega), \tag{6e}$$

340
$$\mathbf{x}_t = \text{the true state of the system}, \tag{6f}$$

$$\mathbf{I} = \text{the identity matrix}. \tag{6g}$$

In Eq. 5, $\mathbf{R}$ and $\mathbf{B}$ are the full-resolution matrices of the error statistics.

For the aggregation errors to be limited, $\mathcal{E}_\omega$ (Eq. 6a) must tend towards $\mathbf{0}$. Then, $\mathcal{S}$ (Eq. 6b) and $\mathcal{S}_\omega$ (Eq. 6c) must be as close as possible to each other and the impact of $\mathcal{P}_\omega$ (Eq. 6d) and of the

345   sandwich product with $\mathbf{\Lambda}_\omega$, $\mathbf{\Lambda}_\omega^T(\cdot)\mathbf{\Lambda}_\omega$, must be as small as possible. $\mathbf{\Gamma}_\omega^T$ extrapolates the fluxes from the aggregated regions to a finer resolution following an a priori repartition. The matrix $\mathcal{P}_\omega$ then redistributes the fluxes over a region with respect to the prior repartition, but keeping the same total emissions on the region.

In Sect. 3.2 below, we explain how to reduce these terms. The exact estimation of Eq. 5 is com-

350   plicated and requires extensive numerical resources (e.g., Wu et al., 2011). In the following, we use physical considerations towards minimizing Eq. 5. The errors that are intrinsic to the aggregation process and that are unavoidable are controlled so that the benefit from the general marginalization is not wasted. We show in Sect. 6.3 that the physical considerations for choosing the representation ω in our case do not depreciate the inversion results compared to what would have been obtained

355   with the exact resolution of Eq. 5.

Considering the computer resources we use, all the principles we define are applied in order to limit the size of the observation space (resp. the state space) to a dimension of roughly 2000 (resp.

11

1500). For instance, in the meso-scale Eurasian case study described in Sect. 5, these considerations lead to the aggregation patterns displayed in Fig. 2 and Fig. 6. With this problem dimensions, the ensemble used in the Monte Carlo sampling consists of 60000 draws.

When the observation and the state space aggregation are chosen, the operator $\mathbf{H}$ can be computed with the so-called 'response functions', based on forward simulations of the transport for each state component (Bousquet et al., 1999).

3.2 Representation choice

3.2.1 Observation space sampling

The sandwich product with $\mathbf{\Lambda}_\omega$, $\mathbf{\Lambda}_\omega^{\mathrm{T}}(\cdot)\mathbf{\Lambda}_\omega$, aggregates the errors in the observation space and diffuses them back within each aggregate along a prescribed temporal profile. For example, it can typically compute the average error per day; then it allocates for each sub-daily dimension an error proportional to the contribution of the related component of $\mathbf{y}^{\mathrm{o}}$ to the daily mean. However, a daily averaging would be dominated by the outliers (e.g, singular spikes or night-time observations when the emissions remain confined close to the surface due to weak vertical mixing) that are generally associated to very high observation errors (due to fine scale mis-representations of the transport and erroneous night vertical mixing). For this reason, we decide to define $\mathbf{\Lambda}_\omega$ as the sampling operator, which, for each day and observation site, picks the component of the observation vector when the daily minimum of concentrations within a planetary boundary layer higher than 500 m is observed. Below this threshold, the vertical mixing by the model is known to be possibly erroneous (e.g., Berchet et al., 2013). The daily resolution is chosen in order to keep a representation of the transport relevant to the meso-scale expectations on flux characterization. Higher time resolution would not improve the inversion efficiency due to strong within-day temporal correlations of errors (Berchet et al., 2013).

3.2.2 Observational constraints

One can notice that far from the observational constraints, the atmospheric dispersion (depicted by the sandwich product with $\mathbf{H}$, $\mathbf{H}(\cdot)\mathbf{H}^{\mathrm{T}}$) makes the potential errors negligible compared to the errors generated in the areas close to the stations. Indeed, gathering two close hot spots of emissions thousands of km away from the observation sites is not problematic since the air masses coming from the two spots will be well mixed. On the opposite, two hot spots that are as distant from each other as the first two, but close to an observation site, will generate plume-like air masses with a very high sensitivity to the errors of mixing and transport in the model. We use an estimation of the observation network footprints (approximating $\mathbf{H}^{\mathrm{T}}$) in order to fix the typical regions constrained by the network and avoid unfortunate grouping. At this step, approximate footprints are preferred to the heavy computation of the complete $\mathbf{H}^{\mathrm{T}}$ and are sufficient for our physical considerations. Within

the constrained regions, we use a small spatial resolution for the fluxes and the transport and fine aggregation patterns; outside of them, we choose a coarse resolution and large aggregation patterns. These guidelines for using footprints prior to an inversion can be applied more systematically, as what is done in Thompson and Stohl (2014). An illustration of aggregation patterns in our case study can be looked at in Fig. 6.

3.2.3 Flux aggregation

Some terms in Eq. 5 are directly related to the aggregation of the fluxes. The term $\mathbf{H}\mathcal{A}_\omega\mathbf{H}^\mathrm{T}$ in Eq. 6c depicts the aggregation errors coming from the uncertain distribution and temporal profile of the fluxes within each aggregation region, then transported to the observation sites. It must be close to $\mathbf{0}$. In our application below, this is particularly important for hot spots of emissions the location of which is poorly known. The term $\mathbf{H}\mathcal{P}_\omega\mathbf{B}\mathcal{P}_\omega\mathbf{H}^\mathrm{T}$ in Eq. 6c must be as close as possible to $\mathbf{HBH}^\mathrm{T}$. The factors of divergence between these two terms come from the areas that are not well constrained by the observations. If, within a region of aggregation, a part is upwind the observation sites, while the other is not seen, then the aggregation errors will scatter on the unseen part of the region. The main sources of errors can then be separated into two different types: 1) the resolution/representation type, and 2) the constraint type.

The type-1 errors are mainly related to the resolution of the observation operator. The models consider that the fluxes and the simulated atmospheric mixing ratios are uniform on a sub-grid basis and neglects sub-grid processes. This discretization contributes to type-1 errors, as 'representation' errors (Tolk et al., 2008). Additionally, the distribution within each aggregation region is fixed and sub-region re-scaling are forbidden. The fine resolution close to the observation network is bound to confine type-1 errors (e.g., Wu et al., 2011). Additionally, the representation error is critical for co-located emissions, especially when the typical temporal and spatial scales of these emissions are different. For instance, grouping hot spots from oil extraction emissions with widespread wetland emissions that quickly vary in time is hazardous. We then aggregate the emissions along their typical time and space scale, hence according to the underlying physical process. An in-depth analysis of the footprints and the small patterns of aggregation close to the observation sites should limit the type-2 constraint errors. Area under high observational constraints should not be grouped with under-constrained areas.

The resolution and aggregation choices can be computed objectively, but at a very high cost and only within a framework of prescribed frozen error matrices (Bocquet, 2009; Wu et al., 2011; Koohkan et al., 2013). For our purpose, we cannot afford such computation costs and rely on heuristic choices: small resolution and aggregation patterns close to the observation sites, aggregation by type of emission, separation of constrained/under-constrained areas by analysing the footprints. These non-optimal subjective choices may damp the efficiency of our method and must be carried

13

out cautiously. Nevertheless, in our case, afterwards checking shows that our choices did not have critical impact on the inversion results in our case.

3.3 Numerical artefacts

430 In addition to the need of defining a well-sized problem, smart adaptations can be applied to the computation of the method in order to enhance the efficiency of the algorithm. We face several sources of numerical artefacts in the computation of the method. In the quasi-Newtonian maximum likelihood algorithm, numerical artefacts are generated by the under-constrained regions. After a few steps, the computed gradient of the likelihood is dominated by these regions and the algorithm stays

435 stationary. This issue could be partly related to the under-optimality of the chosen representation ω as suggested by the optimality criteria described in Bocquet et al. (2011). The stagnation of the maximum likelihood algorithm could then be used to detect too small regions of aggregation.

The under-constrained regions perturbing the maximum likelihood algorithm can be diagnosed using the diagonal terms of the influence matrix $\mathbf{KH}$ (with $\mathbf{K}$ defined in Eq. 1 and following Cardi-

440 nali et al., 2004). This matrix represents the sensitivity of the inversion to elementary changes in the observations. Strong observation constraints are related to high sensitivity. After stagnation, the regions with a diagnosed $\mathbf{KH} < 0.5$ are flagged out and the algorithm is carried on. This way, only the sufficiently constrained components of the state vector are processed until the algorithm converges. A third to half of the regions are flagged out this way in our case study.

445 The detection of the mis-representation of hot-spot plumes should also be enhanced. Despite the minimum daily sampling and the fine resolution close to the observation network, the plume issue can still generate strong temporal and spatial mismatches. For example, a point source can influence a station in the real world, but not in the model because it has been mis-located, and conversely. This creates significant differences between the simulated and the observed concentrations. The maxi-

450 mum likelihood algorithm attributes such mismatches to prior errors and/or observation errors. High diagnosed errors in the maximum likelihood algorithm are then a criterion for plausible mismatches. We know such plumes must be flagged out from the inversion to avoid irrelevant high influence from very local sources hard to represent. Since we notice that the observation and prior computed errors seems to follow a Fischer-Snedecor distribution, we choose to flag out the observations that

455 are within the 95% tail of the distribution.

4 Validation experiments

In Sect. 2, we described our modified atmospheric inversion by marginalization. In Sect. 3, we proposed some essential rules to follow in order to properly define the problem, so that the rather simple theoretical framework is not hindered by finite numerical resources. The marginalization method has

460 to be validated along objective criteria. In the following, we summarize the general structure of the

14

method in order to identify the critical points to test in the method (Sect. 4.1). We deduce from these points some OSSEs to carry out. In Sect. 4.2, we define the scores according to which the method will be evaluated.

4.1   Required tests

4.1.1   Method summary

465

The method described in Sect. 2 and 3 is condensed in the block diagram in Fig. 2. The marginalized inversion takes the same input as any other atmospheric inversion: some atmospheric measurements and prior maps of fluxes with specified resolution and temporal profiles. In Sect. 3, we gave recommendations on the processing of the 'raw' inputs, so we get an observation vector $\mathbf{y}^o$,

470   a prior state vector $\mathbf{x}^b$ and an observation operator $\mathbf{H}$ that are small enough to be computable by the method. These highlights are mainly the sampling of the observations per day (in accordance with our objective of characterizing meso-scale fluxes in our case study) and the aggregation of the fluxes by regions (based on physical considerations and footprint analysis). The maximum likelihood algorithm processes $\mathbf{y}^o$, $\mathbf{x}^b$ and $\mathbf{H}$ in order to find a couple of optimal diagonal error matrices

475   $(\mathbf{R}_{max}, \mathbf{B}_{max})$. This maximum likelihood is found by a quasi-Newtonian descent method. We then infer from $(\mathbf{R}_{max}, \mathbf{B}_{max})$ the approximate χ^2 shape of the distribution of all the possible error matrices $(\mathbf{R}, \mathbf{B})$. We carry out a Monte Carlo sampling on these distributions of errors and get an ensemble of posterior state vectors $(\widehat{\mathbf{x}^a})$. The processing of this ensemble provides the final output of the method: a tolerance interval TI_{68} of the posterior state and the posterior correlations between the components

480   of the state space. The method also allows the explicit computation of the influence matrix $\mathbf{K}_{max}\mathbf{H}$ in order to analyse the constrained regions of emissions only.

To summarize, the marginalized inversion processes two vectors and one operator: $\mathbf{y}^o$, $\mathbf{x}^b$, and $\mathbf{H}$, as any other atmospheric inversion. The main difference with most other atmospheric inversions resides into the objective and automatic computation of the influence of ill-specified error statistics,

485   in contrast with the traditional assigning of frozen error matrices based on expert knowledge and with the more recent online computations of error hyperparameters. Thus, we do not have to inquire into the sensitivity of our method to the prescribed spatial correlations of flux errors, or to the error variances. Such a sensitivity is transposed to the choice of the aggregation patterns and the sampling protocol, as defined in Sect. 3.1. The chosen configuration of aggregation and the sampling protocol

490   are checked afterwards to be relevant in our case study. OSSEs are then to be carried out to evaluate the sensitivity of the method to $\mathbf{y}^o$, $\mathbf{x}^b$, $\mathbf{H}$.

4.1.2   Test strategy

We assume that, in our case, the method is not sensitive to errors in $\mathbf{y}^o$. Indeed, in all the following, we consider that the measurement errors are negligible compared to transport errors; this is true for

surface sites that fulfil the World Meteorological Organisation strict recommendations for accuracy and precision (WMO/GAW, 2011). This approximation does not hold for satellite total columns measurements, for which the transport errors are smoothed over the vertical atmospheric column and the instrument errors are larger. In addition, representativeness errors may also impact $\mathbf{y}^o$. OSSEs should account for these errors. However, OSSEs may face difficulties in explicitly highlighting these errors. Therefore, we do not perturb $\mathbf{y}^o$ in order to represent the instrumental uncertainties and representativeness errors in the OSSEs.

The OSSEs are then based on perturbations of $\mathbf{x}^b$ and $\mathbf{H}$. The discrepancies between the background $\mathbf{x}^b$ and the 'truth' $\mathbf{x}^t$ are of two types: 1) the erroneous distribution and temporal profile of the fluxes within aggregation regions, and 2) incorrect total emissions by region. For example, in Eurasia, the maps of the distribution of the wetlands differ drastically from one database to another (Frey and Smith, 2007). Apart from the distribution, the amount of gas emitted by each process is uncertain, due to mis-parametrizations or, for anthropogenic emissions, mis-specified activity maps (e.g., Rypdal and Winiwarter, 2001). The transport $\mathbf{H}$ differs from the 'true' transport mainly because of the resolution of the model, the parametrization of sub-grid processes (such as vertical turbulent mixing in the planetary boundary layer or deep convection), and the meteorological forcing fields (which are not necessarily optimized for transport applications).

The main sources of errors in the inversion are then: 1) a wrong representation of the transport (highly dependent of the transport model used, its resolution, its parametrization and the exactitude of forcing wind fields), 2) an erroneous distribution of the fluxes within aggregation regions (each inventory and database has different statistical methods and parameters to reproduce surface fluxes), and 3) incorrect total emissions by regions. In order to evaluate the impact of each point on the inversion result, we carry out OSSEs with perfect synthetic observations from a nature run (i.e. with 'true' emissions and 'true' transport, as defined in the set-up in Sect. 5). We test the ability of the marginalized inversion to reproduce the 'true' fluxes or, at least, to consistently include the 'truth' within the tolerance intervals. There are eight possible combination of correct or perturbed phases of the 3 parameters. The 'all true' combination is not relevant: then $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b = 0$ and the maximum likelihood algorithm is stationary. Seven combinations remain, detailed in Tab. 1. We run the marginalized inversion for the seven OSSEs and evaluate them along the scores defined in Sect. 4.2 below.

4.2 OSSE evaluation

4.2.1 Scoring system

We expect an atmospheric inversion to provide reliable ranges of uncertainties for surface fluxes. That is to say, for as many components of the state vector $\mathbf{x}_i$ as possible, the 'truth' $\mathbf{x}_i^t$ should be within the tolerance interval TI_{68}, $[\mathbf{x}_i^{\text{low}}, \mathbf{x}_i^{\text{high}}]$ (defined in Sect. 2). In order to evaluate the ability of

16

530 producing consistent fluxes, we define a relative score $\mathbf{z}_{\text{rel}}$ for each component of the state vector: $(\mathbf{z}_{\text{rel}})_i = 2\frac{|\mathbf{x}_i^{\text{a}} - \mathbf{x}_i^{\text{t}}|}{\mathbf{x}_i^{\text{high}} - \mathbf{x}_i^{\text{low}}}$. Hereafter, all the scores will be expressed in % for better readability. Statistically, $\mathbf{z}_{\text{rel}}$ has no upper bound. Relative scores bigger than 100% are not statistically inconsistent, but, for the method to be validated, we expect that the proportion of state components with $\mathbf{z}_{\text{rel}} < 100\%$ is dominant.

535 Furthermore, the atmospheric inversion is supposed to reveal pieces of information to the understanding of the system. Then, we also expect that a correct relative score below 100% is not reached by specifying huge tolerance intervals. To evaluate the ability of the marginalization of getting close to the reality, i.e. providing valuable information on the state of the system, we define an absolute score $\mathbf{z}_{\text{abs}}$: $(\mathbf{z}_{\text{abs}})_i = \left|\frac{\mathbf{x}_i^{\text{a}}}{\mathbf{x}_i^{\text{t}}} - 1\right|$. The smaller the absolute score, the more accurate the marginalized

540 inversion.

An inversion also must be able to evaluate the observation constraints on the regions. An objective estimator of the constraints on the regions is the influence matrix $\mathbf{KH}$ defined in Sect. 3. The Kalman gain matrix depends on the couple $(\mathbf{R}, \mathbf{B})$. Amongst all the Monte Carlo draws, we compute the influence matrix $\mathbf{K}_{\text{max}}\mathbf{H}$ for the couple associated to the maximum likelihood. The diagonal terms of this matrix range from 0 to 1. They give for all components of the state space the constraint given

545 by the observations. We then define the influence score: $(\mathbf{z}_{\text{infl}})_i = (\mathbf{K}_{\text{max}}\mathbf{H})_i$. The closest to 100% these terms, the more constraints the inversion provides. We can then deduce the typical range of influence of the observation sites and detect the blind spots of the used network.

For each component i of the state space, we then have defined 3 indicators:

550
$$
\begin{cases}
(\mathbf{z}_{\text{rel}})_i & = & 2\frac{|\mathbf{x}_i^{\text{a}} - \mathbf{x}_i^{\text{t}}|}{\mathbf{x}_i^{\text{high}} - \mathbf{x}_i^{\text{low}}} \\
(\mathbf{z}_{\text{abs}})_i & = & \left|\frac{\mathbf{x}_i^{\text{a}}}{\mathbf{x}_i^{\text{t}}} - 1\right| \\
(\mathbf{z}_{\text{infl}})_i & = & (\mathbf{K}_{\text{max}}\mathbf{H})_i
\end{cases}
\tag{7}
$$

4.2.2 Posterior correlation processing

Another point most inversions do not compute explicitly and objectively is the typical temporal and spatial scales the inversion can differentiate in the fluxes, considering the atmospheric transport and the density of the observations. Our marginalized inversion gives access to an explicit matrix of cor-

555 relations as defined in Sect. 3. Strong positive and negative correlations between two components of the state space indicate that the inversion cannot separate the contributions from the two components. For example, air masses observed at a station and coming from two regions upwind the station will have a mixed atmospheric signal difficult to analyse. Co-located emissions are also not necessarily differentiated in the atmospheric signal. Moreover, in a regional framework, when a model of lim-

560 ited area is coupled to lateral boundary conditions (LBC), the inversion must explicitly alert on the regions that cannot be separated from the boundary conditions, i.e. from the baseline signal.

In the case of strong correlations between two components of the state space in the posterior co-variance matrix, we consider that it is not relevant to try to infer specific information for the two

separate components. Then, we group the state space components according to their posterior correlations. We define a threshold of correlation r_{max} and associate couples of regions (i, j) within groups such that $|r_{i,j}| > r_{max}$. If we prescribe $r_{max} = 0$, all the regions will be grouped; on the opposite, if $r_{max} = 1$, no group will be formed. The optimal correlation threshold is not evident. We test the grouping for all possible r_{max}. We flag out from the processing of the results all the groups, which include some contributions from the LBC. Thus, from this post processing, we only keep the regions that are clearly constrained by the observation sites, with no interference from the LBC. Moreover, we can infer the spatial and temporal scale that the inversion can resolve from the grouping patterns.

In Tab. 1, the three scores defined in Eq. 7 are averaged on the whole domain of interest for the optimal correlation threshold r_{max} (as discussed in Sect. 6.1).

5 Set up of the OSSEs

We compute the OSSEs that we described in Sect. 4 in a realistic meso-scale case. We focus on a domain spanning over Eurasia, from Scandinavia to Korea. At this scale, the air masses residence time is typically of days to a few weeks. This time scale is small compared to the lifetime of methane of 8-10 years in the atmosphere (mainly due to oxidation by OH radicals; Dentener et al., 2003). Hence, the observation operator can be consider linear. We apply the method on a region characterized by significant fluxes, with collocation of different sources with different emission time-scales: Siberia. We describe the region of interest and the chosen 'truth' for the experiments in Sect. 5.1. We use two transport models in order to simulate atmospheric transport. The technical details on these models are summarized in Sect. 5.2. In Sect. 5.3, we explain how we choose and compute the synthetic observations for our experiments.

5.1 Virtual true state x^t

5.1.1 State space components

In the region of interest, the emissions of methane are dominated by wetland, anthropogenic (here, mainly related to the oil and gas industry) and wildfire emissions. In Fig. 3, the distributions of the wetlands and of the oil and gas industry in the region are displayed. Anthropogenic emissions of methane in the region are mainly hot spots related to the intense oil and gas industry in the Siberian Lowlands and to the leaks in the distribution system in population centres in the South part of Siberia. Wetland emissions are mainly confined in the lower part of Siberia in the West Siberian plain, half of which is lower than 100 meters above sea level.

The spatial distribution of the associated fluxes is deduced from: 1) EDGAR database v4.2 (http://edgar.jrc.ec.europa.eu) for year 2008 for anthropogenic emissions, 2) LPX-Bern v1.2 process model at a monthly scale for wetland emissions (Spahni et al., 2011), 3) GFED database at daily scale for wildfires (Giglio et al., 2009). The EDGAR inventory uses economic activity maps by sectors and

18

convolves them with emission factors estimated in laboratories or with statistical studies (Olivier et al., 2005). LPX-Bern is an update of process model LPJ-Bern (Spahni et al., 2011). It includes a dynamical simulation of inundated wetland areas, dynamic nitrogen cycle, and dynamic evolution of peatlands (Spahni et al., 2013; Ringeval et al., 2014). The model uses CRU TS 3.21 input data (temperature, precipitation rates, cloud cover, wetdays) and observed atmospheric CO_2 for each year for plant fertilization. GFED v4 is built from burnt area satellite product (MCD64A1). CH_4 emissions at monthly and daily scales are deduced from the burnt areas using the Carnegie-Ames-Stanford-Approach (CASA model; Potter et al., 1993) and emission factors (van der Werf et al., 2010). Wildfire emissions can be very strong and are punctual in time and space; they are then difficult to analyse by the inversion. Wildfires are included as inputs to the marginalized inversion, but are automatically filtered out during the computation. In all the following, we evaluate the OSSEs only in terms of anthropogenic and wetland emissions.

In addition, at the meso-scale, we use a CTM (see Sect. 5.2.2) with a limited area domain. Initial and lateral boundary conditions (IC and LBC) are then also to be optimized in the system to correct the atmospheric inflow in the domain. Lateral concentrations are deduced from simulations at the global scale by the general circulation model LMDz with the assimilation of surface observations outside the domain of interest (Bousquet et al., 2006). We aggregate the LBC along 4 horizontal components and 2 vertical ones (1013-600 hPa and 600-300 hPa).

5.1.2   Generation of a perturbed reference state $\mathbf{x}^\mathrm{t}$

The EDGAR fluxes are given at the yearly scale and the LPX fluxes are calculated at a monthly scale. Additionally, LPX monthly fluxes exhibit smoothed patterns while wetland emissions can vary drastically from a point to another. We want the nature run for OSSEs to reproduce the potential spatial and temporal variability of the emissions. To do so, we intensify the spatial and temporal contrasts from the databases to the nature run. We then compute the 'true' state vector $\mathbf{x}^\mathrm{t}$ by perturbing EDGAR emissions on a monthly basis and LPX on a 10-day basis. That is to say: $\mathbf{x}^\mathrm{t} = \alpha \otimes \mathbf{x}^\mathrm{data}$, with the vector α depicting the scaling factors by state space component, $\otimes$ the point-wise multiplication operator and $\mathbf{x}^\mathrm{data}$ the emissions from the databases. The perturbations in α from original EDGAR and LPX databases applied to get the 'truth' are scaling factors up to 10. These scaling factors could have been chosen randomly, but we prefer inferring them with a raw expert-knowledge-based inversion using real data. The purpose of using real data for computing $\mathbf{x}^\mathrm{t}$ is to generate potentially realistic variations within the state space.

For both anthropogenic and wetland emissions, the scaling factors can significantly differ from a period of inversion to another. We can then evaluate the ability of the marginalized inversion to catch quick variations. The distribution of the scaling factors α is shown in Fig. 4. These factors are plausible, knowing the uncertainties on the wetland emissions and gas leakage (e.g., Reshetnikov et al., 2000). Such target scaling factors are at the edge of the validity of the Gaussian assumption

19

and of the positivity of methane fluxes. The ability of the marginalization to recover such correction

635 factors will prove its robustness.

As for anthropogenic and wetland emissions, we apply the scaling factors α on the components of $\mathbf{x}^t$ related to LBC by periods of 10 days.

The OSSEs rely on $\mathbf{x}^b$ perturbed from $\mathbf{x}^t$, or not. We apply two types of perturbations as summarized in Tab. 1. In OSSE 1, 4, 5 and 7, we only implement prior fluxes with different total emissions

640 on the regions of aggregation. We take the emissions of the raw inventories as background to test the ability of recovering 'true' fluxes from realistic background fluxes without assigning frozen prior errors. In OSSE 2, 4, 6 and 7, the distribution of the prior fluxes is modified from the 'truth'. We choose all flat flux distributions for each region of aggregation as prior fluxes.

5.2 Simulation of the observation operator $\mathbf{H}$

645 The observation operator $\mathbf{H}$ is deduced from simulations of atmospheric transport We use two different transport models in order to evaluate the impact of the transport on the inversion. We define $\mathbf{H}_{\text{FLEXPART}}$ with the Lagrangian dispersion model FLEXPART and $\mathbf{H}_{\text{CHIMERE}}$ with the Eulerian Chemistry-Transport Model CHIMERE. Any transport model can be considered at some point biased compared with the reality. Confronting the results from FLEXPART to those from CHIMERE

650 will allow us to test the robustness of our method to the biases.

5.2.1 The Lagrangian model: FLEXPART

With the Lagrangian dispersion model FLEXPART (Stohl et al., 2005), we can compute the footprints of the observations, hence $\mathbf{H}_{\text{FLEXPART}}^{\text{T}}$. We use FLEXPART version 8.2.3 to compute numerous back-trajectories of virtual particles from the observation sites. The model is forced by the European

655 Centre for Medium-range Weather Forecast (ECMWF) ERA-Interim data at an horizontal resolution of $1° \times 1°$, with 60 vertical levels and 3 hour temporal resolution (Uppala et al., 2005). Virtual particles are released in a 3D box centred around each observation site with a 10-day lifetime backwards in time. The footprints are computed on a $0.5° \times 0.5°$ horizontal grid, following the method of Lin et al. (2003), taking the boundary layer height at each particle location into account. The footprints

660 only have to be convolved with the emission fields in order to get simulated concentrations at the observation sites. The method for computing the footprints considers that only the particles within the boundary layer are influenced by surface emissions and that the boundary layer is well-enough mixed to be considered as uniform. The uniform vertical mixing of the mixing layer can generate a bias on the surface simulated concentrations. Such a bias is critical in the classical inversion frame-

665 work and consequently in the one we describe since all the uncertainties are considered unbiased.

FLEXPART can easily compute an estimation of the adjoint of the full-resolution observation operator before choosing the representation ω. Hence, despite the expectable biases, we use this model to estimate the footprints of the network and deduce the aggregation patterns needed to compute

20

$\mathbf{H}_{\text{CHIMERE}}$. This same model FLEXPART may also be used to compute explicitly and rigorously the representation ω according to objective criteria (Koohkan and Bocquet, 2012).

5.2.2 The Eulerian model: CHIMERE

Using the Eulerian meso-scale chemistry transport model CHIMERE (Vautard et al., 2001; Menut et al., 2013) constrained by non-hydrostatic meteorological fields, we explicitly define the observation operator $\mathbf{H}_{\text{CHIMERE}}$ by computing the forward atmospheric transport from the emission aggregated regions (defined according to Sect. 3 criteria) to the observation sites. This model was developed in a framework of air quality simulations (Schmidt et al., 2001; Pison et al., 2007), but is also used for greenhouse gas studies (Broquet et al., 2011; Berchet et al., 2013). We use a quasi-regular horizontal grid zoomed near the observation sites after the considerations of Sect. 3. The domain of interest is of limited area and spans over the mainland of the Eurasian continent (see Fig. 3). The average side length of the grid cells near the stations is 25 km, while it spans over 150 km away from the observation sites. The 3D-domain roughly embraces all the troposphere, from the surface to 300 hPa (~ 9000 m), with 29 layers geometrically spaced. The model time step varies dynamically from 4 to 6 min depending on the maximum wind speed in the domain. The model is an off-line model which needs meteorological fields as forcing. The forcing fields are deduced from interpolated meteorological fields from ECMWF with a horizontal resolution of $0.5° \times 0.5°$ every 3 h.

5.3 Synthetic observations $\mathbf{y}^{\text{o}}$

We compute the nature run, i.e., the synthetic observations, from the 'true' state vector, with the CTM CHIMERE. That is to say, in all the following, we consider that: $\mathbf{y}^{\text{o}} = \mathbf{H}_{\text{CHIMERE}}\mathbf{x}^{\text{t}}$. The site and date of available observations are chosen according to the operated observation sites in the region. Thirteen Eurasian surface sites have been selected. These sites are maintained by NIES (Tsukuba, Japan; Sasakawa et al., 2010), IAO (Tomsk, Russian Federation), MPI (Iena, Germany; Winderlich et al., 2010), NOAA-ESRL (Boulder, United States of America; Dlugokencky et al., 2009), and KMA (Seoul, Korea). The description of the sites is given in Tab. 2. The observation sites do not carry out measurements all the year round due to logistical issues and instrument dysfunctions. In order to reproduce this sampling bias, we generate synthetic observations only when real measurements are available from January to December 2010.

6 Results and discussion

After the description of the set-up in Sect. 5, we now have a 'true' state $\mathbf{x}^{\text{t}}$ and some reference observations $\mathbf{y}^{\text{o}}$. We also have two observation operators $\mathbf{H}_{\text{CHIMERE}}$ and $\mathbf{H}_{\text{FLEXPART}}$ and several possible prior fluxes $\mathbf{x}^{\text{b}}$ as inputs for the marginalized inversion developed in Sect. 2. In order to evaluate

21

the method, we now carry out the OSSEs described in Tab. 1 following the complete procedure in Fig. 2. In Sect. 6.1, we examine the average robustness of the method. Then, in Sect. 6.2, we explore the spatial efficiency of the marginalized inversion in our case study. In Sect. 6.3, we discuss the enhancement provided by our method compared to the classical Bayesian framework, despite some limitations.

6.1 Robustness of the method

6.1.1 Impact of the correlation processing

The marginalization should consistently reproduce the nature run in the OSSEs or, at least, it should detect its inability in characterizing the fluxes from the given atmospheric constraints. As detailed in Sect. 4.2, the aggregation regions may have strong posterior correlations after the marginalized inversions. This denotes the difficulties that the inversion encounters in separating some emissions. The aggregation regions can be grouped along correlation thresholds r_{max} arbitrarily chosen in order to explicitly take the emission dipoles into account. In Fig. 5, we plot the profiles of the scores defined in Sect. 4.2 along the possible correlation thresholds r_{max} for grouping the regions. Specifying a correlation threshold r_{max} allows identifying the typical temporal and spatial scales that the inversion can separate. In the case of a limited domain CTM, the influence of the LBC and of the inside fluxes can be mis-separated. The correlations take this issue into account and the correlation threshold specifies the tolerance to such mis-separations.

For all OSSEs, the influence score $\mathbf{z}_{infl}$ increases with r_{max}. In the correlation processing after the computation of the marginalized inversion, the threshold r_{max} depicts the degree of tolerance to mis-separation between inside fluxes and LBC. The higher the threshold of tolerance r_{max}, the fewer inside fluxes are considered unseparable from the LBC. Hence, fewer aggregation regions are eliminated from the inversion and more fluxes are corrected by the inversion. As the number of constraints increases, we notice that the absolute and relative scores, $\mathbf{z}_{abs}$ and $\mathbf{z}_{rel}$, also tend to increase with r_{max}. That is to say, if we only try to get average information on big under-resolved regions, the posterior fluxes can be expected to be closer to the 'truth'. On the opposite, if we try to process too much spatial information from the inversion, we must expect more discrepancies with the 'truth'.

In particular, in Fig. 5, one can notice some outlier peaks for low r_{max}. For low r_{max}, very few regions are computed in the inversion. The peaks are created by the regions that are not any more considered as mis-separated from the LBC when r_{max} increases. For some OSSEs, these newly computed regions have very wrong scores and dominate upon the other few computed regions. For this reason, one should be very careful in the chosen correlation threshold. In order to avoid the score instability, the optimum threshold should be chosen higher than 0.4. Above 0.5, in our meso-scale case study, as described above, the inversion is limited by the temporal and spacial variability of

22

the fluxes to optimize and by the transport biases. Then, it can't reach the requirement of consistent reproduced fluxes.

One should find a balance between the physical scales one want to separate and the consistency of the results. In Tab. 1, we summarize the scores of every OSSE for a chosen correlation threshold with respect to result consistency.

6.1.2 Hot-spots and large-area emissions

Both in Tab. 1 and Fig. 5, looking at a given correlation threshold r_{max}, one would expect influence, relative and absolute scores that get more wrong when the inversion conditions degrades.

The fossil fuel influence score follows this trend: the more perturbed the transport and the prior fluxes are, the more state space components are considered un-invertible. The hot-spot regions of emissions are broadly filtered out and the remaining regions can be well characterized by the inversion even with wrong distribution and transport patterns. Some effects in the degrading conditions of the inversion can though compensate each other. For example, the absolute scores of OSSEs 5 and 7 are better than the one of OSSEs 3 and 6.

The situation for wetland emissions is different. These emissions are smoother than oil and gas emissions and are then not excluded because of wrong transport or distributions. For this reason, the influence score does not exhibit a clear trend with degrading inversion conditions. For wetland regions, transport seems to be the predominant factor of errors. OSSEs 3, 5, 6 and 7 do not consistently reproduce the 'truth' with relative scores higher than 100% when $r_{max} \geq 0.4$. These discrepancies can be attributed to the very high variability prescribed in the 'true' wetland emissions. An erroneous transport will fail in detecting brutal changes of emissions at the synoptic scale. The wetland emissions should then be grouped temporally and spatially in order to average the point releases of methane.

The erroneous tolerance intervals can also be attributed to the biased transport in FLEXPART compared with CHIMERE. Since we filtered out most of the plumes with spatial and temporal mismatches with the observations, the horizontal biases in the transport are confined. Concerning the vertical bias, a wrong simulated vertical mixing in the planetary boundary will affect all the fluxes. This bias will then have an impact on the atmospheric concentrations that is relatively smoothed, uniform and constant. Therefore, an accurate detection of such a bias is very difficult. Any inversion relies on the unbiased assumption of the errors. The inversion will attribute the biases to the flux for wetland regions, impacting the result of the inversion. As other inversions, despite the marginalization, it appears that the results on wetland regions may be sensitive to vertical transport biases in the models (see discussion in Sect. 6.3.2).

Thus, the marginalized inversion seems to be sensitive to transport biases and to fluxes varying too quickly, as any other inversions. Nevertheless, a post-processing is made possible by the explicit and objective computation of the posterior covariances and of the influence matrix. This post-processing

23

proves that the atmospheric inversion is not able to inquire into very fine scales in our case study. The correlation grouping of indifferentiable regions allows an accurate analysis of the best possible signal detectable by the inversion. In the following, we take a correlation threshold of 0.5 as a good balance between sufficient constraints on the system and consistent posterior fluxes.

6.2 Spatial evaluation

We have chosen a threshold of correlation grouping the regions so that the averaged scores on the whole domain of interest are optimal. The scores are not uniformly distributed. In Fig. 6, the distributions of the three scores are displayed for fossil fuel regions and wetlands for OSSE 1 (transport and distribution of the fluxes same as the 'truth', perturbed masses by regions; see Tab. 1). We choose the 'easiest' OSSE configuration in order to evaluate the behaviour of the marginalized inversion in the best configuration possible, thus getting the upper bound for the expectable quality of the results. Any more realistic set-up likely gives less good results. In the figure, the scores are projected on the aggregation grid built on the considerations in Sect. 3. Most of the observation sites are located in the centre of the domain (see Fig. 3). Then, the influence score is on average better close to the core of the network for the wetlands. For the fossil fuel regions, the influence score is relatively high also upwind the monitoring network (dominant winds blow West to East in the region). Additionally to the network density, the inversion suffers from mis-separation of side regions and LBC. For this reason, side regions tend to be less constrained than centre ones. However, one can notice in both wetland and fossil fuel maps that some centre regions are significantly less constrained than the core of the domain on average. These are regions of very high and dense emissions close to the observation sites (< 500 km). The air masses coming from these regions to the observation sites are plume shaped air masses. The inversion has troubles in assimilating single plumes. In Sect. 3, filters have been implemented in order to detect these problematic regions. The marginalized inversion effectively filtered out these regions.

The absolute and relative scores also show unexpected patterns. Scandinavia and China regions own some of the best absolute and relative scores. These two side regions are filtered out most of the time because of strong correlations with the LBC components of the state space (confirmed by their low influence score). Consequently, when not filtered out, these regions are very well and unambiguously constrained, so the good relative and absolute scores. For the rest of the domain, the scores are mostly the better, the closer to the observation network.

6.3 Limitations and benefits

6.3.1 Promising computation of the uncertainties

The marginalized inversion provides an objectified quantification of the errors in the inversion system. With the Monte Carlo approach we implemented, we are able to consistently take the sources of

24

uncertainties in the inversion process into account, especially those from the prescribed error covariance matrices. As evaluated through OSSEs, the method proved to consistently catch 'true' fluxes on average in the particular Siberian set-up. Moreover, the Siberian set-up is a difficult case study

810 for atmospheric inversions, with co-located intense fluxes that vary at temporal and spatial scales smaller than the meso-scale. The processing of hot-spots, critical in most inversion configurations, is consistently managed, through filters on the plume-shaped air masses. An in-depth analysis of the temporal variability of the fluxes is carried out in a sister publication with the Siberian set-up and real observations (Berchet et al., 2014). Additionally, as a comparison, we carried out the same OSSEs

815 on the same particular Siberian set-up, but with expert-knowledge frozen error matrices (diagonal matrices with the same representation ω as for the other OSSEs). The correlation profiles and the spatial structures of the scores with the expert-knowledge matrices are not shown because the general patterns are very similar to what is described for the marginalized inversion. Though similar in patterns, the values of the scores are significantly depreciated from the marginalized inversion to the

820 expert-knowledge one. The expert-knowledge relative and absolute scores are several times bigger than the ones from the marginalized inversion, thus statistically incompatible with the 'truth'.

The marginalized inversion explicitly and objectively computes the posterior covariance matrix and the influence matrix. The physical interpretation of the inversion results are then enhanced by a clear analysis of the observation constraints to the fluxes. The processing of the posterior corre-

825 lations makes the detection of the dipoles and of indistinguishable regions possible. The influence of the lateral boundary conditions, specific to the meso-scale and to the use of limited area CTMs, is estimated. Thus, the regions upwind the observation sites and mixed with lateral air masses can be excluded from the inversion. From the correlations, the grouping of regions gives an estimate of the typical spatial and temporal scale the method can compute. In our case, with few and distant

830 observation sites, the groups of regions cover very large areas. Thus, a grid-point high resolution inversion would not have given deep insights into the fluxes we are looking at. The reduced problem approach described in Sect. 3 is then relevant when computed cautiously.

6.3.2 Subjective choices and biases

Despite all these benefits compared with the classical Bayesian framework, our method still has

835 limitations. The technical implementation of the method needs extensive computation power and memory requirements. For this reason, we have to drastically reduce the size of the problem to solve. The size reduction relies on rigorous considerations that are difficult to formulate analytically. Therefore we applied heuristic principles in order to choose the aggregation patterns of the observations and the fluxes. This subjective procedure can modify the results of the inversion and must

840 be carried out very cautiously. The way we group the regions after the marginalized inversion in order to physically interpret the results is also subjective. We choose a correlation threshold of 0.5 in order to counter-balance the need of useful constraints from the inversion and the requirements

of consistently reproducing the 'true' fluxes. Other thresholds could have been chosen and the typical distinguishable temporal and spatial scales would slightly differ from one threshold to another. But, in any chosen correlation threshold, we notice that most aggregation regions are grouped within bigger ensembles, suggesting that the chosen aggregation patterns are small enough to have reduced impact on the inversion post-processed results.

The marginalized inversion suffers from transport biases as any other inversion. However, the maximum likelihood algorithm considers the biases as random errors and includes them into the error matrix $\mathbf{R}_{max}$. The biases are then taken into account in the marginalized inversion, though as random errors. Biases can be represented, or at least detected, with non-diagonal matrices as suggested by Berchet et al. (2013), but a non-diagonal framework would make the computation of the marginalized inversion critically complicated. Despite the implicit inclusion of the biases as random error in $\mathbf{R}_{max}$, we reduced the impact of the horizontal transport biases through filters on the plume-shaped air masses. The vertical biases are smoother and more difficult to detect. This issue must be inquired into in further works. Biases can be studied through marginalizations on the input vectors (e.g., Bocquet, 2011). Coupled marginalizations on the input vectors and on the error statistics would provide a more complete view on atmospheric inversion uncertainties.

7 Conclusions

At the meso-scale, inconsistencies between inversion configurations appear in the classical Bayesian framework. One of the main sources of inconsistencies is the specification of the error matrices and the non inclusion of the tenacious uncertainties on these matrices. Synthesizing the recent literature, we developed an updated Bayesian method of inversion from the classical Bayesian framework based on a marginalization on the error matrices and on an objectified specification of the probability density function of the error matrices. This new method makes the comprehensive inclusion of the impact of ill-specified uncertainty matrices possible for the first time in atmospheric inversion to our knowledge. In principle, this method needs very high computation power and memory resources. To avoid technical limitations, we reduce the size of the problem by aggregating the fluxes by region, following objective principles for reducing aggregation errors. We test this method through OSSEs on methane in a domain of interest spanning over Eurasia with significant emissions of different types and different time and space scales. The OSSEs are based on synthetic observations generated from a nature run. We evaluate the consistency and robustness of the method on OSSEs with inversion configurations from the more favourable to the most disadvantageous one (perturbed atmospheric transport, flat flux distribution and wrong total masses). The method produces very consistent and satisfactory results. In most cases, the tolerance intervals given by the inversion include the 'true' fluxes and the results remain close to the 'truth'. The method also provides an explicit computation of the constraints on the regions and allows flagging out regions critically mis-separated from the

lateral boundary condition. We hence have developed a robust and objectified method able to consistently catch 'true' greenhouse gas emissions at the meso-scale and to explicitly group the regions that are physically un-distinguishable with the atmospheric signal only. In addition, we developed a method that explicitly produces posterior tolerance intervals on the optimal distinguishable time and space flux scales and that computes the observation network influence on the fluxes.

The robustness of our method on the Siberian case with a biased transport proves that it can be generically applied to other meso-scale frameworks. The high spatial and temporal variability of the fluxes in Siberia ensures the possibility of using the system in 'easier' inversion set-up. Actual observations from the sites we used for the validation of the method are exploited in further steps of our work in order to quantify the 'real' methane fluxes in the Siberian Lowlands (Berchet et al., 2014).

References

900 Ahmadov, R., Gerbig, C., Kretschmer, R., Koerner, S., Neininger, B., Dolman, A. J., and Sarrat, C.: Mesoscale covariance of transport and CO$_2$ fluxes: Evidence from observations and simulations using the WRF-VPRM coupled atmosphere-biosphere model, J. Geophys. Res., 112, D22 107, doi:10.1029/2007JD008552, http://onlinelibrary.wiley.com.biblioplanets.gate.inist.fr/doi/10.1029/2007JD008552/abstract, 2007.

Amante, C. and Eakins, B. W.: ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis,
905 US Department of Commerce, National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, National Geophysical Data Center, Marine Geology and Geophysics Division, 2009.

Berchet, A., Pison, I., Chevallier, F., Bousquet, P., Conil, S., Geever, M., Laurila, T., Lavrič, J., Lopez, M., Moncrieff, J., Necki, J., Ramonet, M., Schmidt, M., Steinbacher, M., and Tarniewicz, J.: Towards better
910 error statistics for atmospheric inversions of methane surface fluxes, Atmos. Chem. Phys., 13, 7115–7132, doi:10.5194/acp-13-7115-2013, http://www.atmos-chem-phys.net/13/7115/2013/, 2013.

Berchet, A., Pison, I., Chevallier, F., Paris, J.-D., Bousquet, P., Bonne, J.-L., Arshinov, M. Y., Belan, B. D., Cressot, C., Davydov, D. K., Dlugokencky, E. J., Fofonov, A. V., Galanin, A., Lavrič, J., Machida, T., Parker, R., Sasakawa, M., Spahni, R., Stocker, B. D., and Winderlich, J.: Natural and anthropogenic
915 methane fluxes in Eurasia: a meso-scale quantification by generalized atmospheric inversion, Biogeosciences Discuss., 11, 14 587–14 637, doi:10.5194/bgd-11-14587-2014, http://www.biogeosciences-discuss.net/11/14587/2014/, 2014.

Bergamaschi, P., Krol, M., Dentener, F., Vermeulen, A., Meinhardt, F., Graul, R., Ramonet, M., Peters, W., and Dlugokencky, E. J.: Inverse modelling of national and European CH4 emissions using the at-
920 mospheric zoom model TM5, Atmos. Chem. Phys., 5, 2431–2460, doi:10.5194/acp-5-2431-2005, http://www.atmos-chem-phys.net/5/2431/2005/, 2005.

Bergamaschi, P., Krol, M., Meirink, J. F., Dentener, F., Segers, A., van Aardenne, J., Monni, S., Vermeulen, A. T., Schmidt, M., Ramonet, M., Yver, C., Meinhardt, F., Nisbet, E. G., Fisher, R. E., O'Doherty, S., and Dlugokencky, E. J.: Inverse modeling of European CH$_4$ emissions 2001-2006, J. Geo-
925 phys. Res., 115, doi:10.1029/2010JD014180, http://onlinelibrary.wiley.com.biblioplanets.gate.inist.fr/doi/10.1029/2010JD014180/abstract, 2010.

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A., Dlugokencky, E., Wofsy, S. C., Kort, E. A., Sweeney, C., Schuck, T., Brenninkmeijer, C., Chen, H., Beck, V., and Gerbig, C.: Atmospheric CH4 in the first decade of the 21st century: Inverse modeling analysis using SCIA-
930 MACHY satellite retrievals and NOAA surface measurements, J. Geophys. Res. Atmos., 118, 7350–7369, doi:10.1002/jgrd.50480, http://onlinelibrary.wiley.com/doi/10.1002/jgrd.50480/abstract, 2013.

Bocquet, M.: Toward Optimal Choices of Control Space Representation for Geophysical Data Assimilation, Monthly Weather Review, 137, 2331–2348, doi:10.1175/2009MWR2789.1, zotero://attachment/324/, 2009.

Bocquet, M.: Ensemble Kalman filtering without the intrinsic need for inflation, Nonlin. Processes Geophys.,
935 18, 735–750, doi:10.5194/npg-18-735-2011, http://www.nonlin-processes-geophys.net/18/735/2011/, 2011.

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

Bousquet, P., Ciais, P., Peylin, P., Ramonet, M., and Monfray, P.: Inverse modeling of annual atmospheric
CO$_2$ sources and sinks: 1. Method and control inversion, J. Geophys. Res., 104, 26 161–26 178,
doi:10.1029/1999JD900342, http://onlinelibrary.wiley.com/doi/10.1029/1999JD900342/abstract, 1999.

Bousquet, P., Ciais, P., Miller, J. B., Dlugokencky, E. J., Hauglustaine, D. A., Prigent, C., Werf, G. R. V. d.,
Peylin, P., Brunke, E.-G., Carouge, C., Langenfelds, R. L., Lathière, J., Papa, F., Ramonet, M., Schmidt, M.,
Steele, L. P., Tyler, S. C., and White, J.: Contribution of anthropogenic and natural sources to atmospheric
methane variability, Nature, 443, 439–443, doi:10.1038/nature05132, http://www.nature.com/nature/journal/
v443/n7110/full/nature05132.html, 2006.

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen,
A. T., and Ciais, P.: A European summertime CO$_2$ biogenic flux inversion at mesoscale from con-
tinuous in situ mixing ratio measurements, Journal of Geophysical Research: Atmospheres, 116, n/a–n/a,
doi:10.1029/2011JD016202, http://onlinelibrary.wiley.com/doi/10.1029/2011JD016202/abstract, 2011.

Cardinali, C., Pezzulli, S., and Andersson, E.: Influence-matrix diagnostic of a data assimilation system, Q.
J. R. Meteorol. Soc., 130, 2767–2786, doi:10.1256/qj.03.205, http://onlinelibrary.wiley.com/doi/10.1256/qj.
03.205/abstract, 2004.

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.

Chevallier, F., Viovy, N., Reichstein, M., and Ciais, P.: On the assignment of prior errors in Bayesian inversions
of CO$_2$ surface fluxes, Geophys. Res. Lett., 33, L13 802, doi:10.1029/2006GL026496, http://www.agu.
org/pubs/crossref/2006/2006GL026496.shtml, 2006.

Ciais, P., Sabine, C., Govindasamy, B., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway,
J., and Heimann, M.: Chapter 6: Carbon and Other Biogeochemical Cycles, IPCC (2013) Climate Change,
pp. 2–3, 2013.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using
an incremental approach, Q. J. R. Meteorol. Soc., 120, 1367–1387, doi:10.1002/qj.49712051912, http://
onlinelibrary.wiley.com/doi/10.1002/qj.49712051912/abstract, 1994.

Cressot, C., Chevallier, F., Bousquet, P., Crevoisier, C., Dlugokencky, E. J., Fortems-Cheiney, A., Frankenberg,
C., Parker, R., Pison, I., Scheepmaker, R. A., Montzka, S. A., Krummel, P. B., Steele, L. P., and Langenfelds,
R. L.: On the consistency between global and regional methane emissions inferred from SCIAMACHY,
TANSO-FTS, IASI and surface measurements, Atmos. Chem. Phys., 14, 577–592, doi:10.5194/acp-14-577-
2014, http://www.atmos-chem-phys.net/14/577/2014/, 2014.

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.

Dee, D. P.: Bias and data assimilation, Quarterly Journal of the Royal Meteorological Society, 131, 3323–3343,
doi:10.1256/qj.05.137, http://onlinelibrary.wiley.com/doi/10.1256/qj.05.137/abstract, 2005.

Denning, A. S., Holzer, M., Gurney, K. R., Heimann, M., Law, R. M., Rayner, P. J., Fung, I. Y., Fan, S.-
M., Taguchi, S., Friedlingstein, P., Balkanski, Y., Taylor, J., Maiss, M., and Levin, I.: Three-dimensional

transport and concentration of SF6, Tellus B, 51, 266–297, doi:10.1034/j.1600-0889.1999.00012.x, http://onlinelibrary.wiley.com/doi/10.1034/j.1600-0889.1999.00012.x/abstract, 1999.

Dentener, F., Peters, W., Krol, M., Van Weele, M., Bergamaschi, P., and Lelieveld, J.: Interannual variability and trend of CH$_4$ lifetime as a measure for OH changes in the 1979-1993 time period, J. Geophys. Res., 108, 4442, doi:10.1029/2002JD002916, http://www.agu.org/pubs/crossref/2003/2002JD002916.shtml, 2003.

Desroziers, G. and Ivanov, S.: 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, http://onlinelibrary.wiley.com/doi/10.1002/qj.49712757417/abstract, 2001.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Q. J. R. Meteorol. Soc., 131, 3385–3396, doi:10.1256/qj.05.108, http://onlinelibrary.wiley.com/doi/10.1256/qj.05.108/abstract, 2005.

Dlugokencky, E. J., Steele, L. P., Lang, P. M., and Masarie, K. A.: The growth rate and distribution of atmospheric methane, Journal of Geophysical Research: Atmospheres, 99, 17 021–17 043, doi:10.1029/94JD01245, http://onlinelibrary.wiley.com.biblioplanets.gate.inist.fr/doi/10.1029/94JD01245/abstract, 1994.

Dlugokencky, E. J., Bruhwiler, L., White, J. W. C., Emmons, L. K., Novelli, P. C., Montzka, S. A., Masarie, K. A., Lang, P. M., Crotwell, A. M., Miller, J. B., and Gatti, L. V.: Observational constraints on recent increases in the atmospheric CH$_4$ burden, Geophys. Res. Lett., 36, L18 803, doi:10.1029/2009GL039780, http://onlinelibrary.wiley.com/doi/10.1029/2009GL039780/abstract, 2009.

Enting, I. G.: Inverse problems in atmospheric constituent transport, Cambridge University Press, http://books.google.fr/books?hl=fr&lr=&id=b61nA7D1XXUC&oi=fnd&pg=PP1&dq=enting+inverse+problem+atmospheric+constituent+transport+&ots=vc8hwiQ_YK&sig=u5npCEtl4WdBNLyRDn9V2EfD1SQ, 2002.

Enting, I. G., Trudinger, C. M., Francey, R. J., and Granek, H.: Synthesis inversion of atmospheric CO$_2$ using the GISS tracer transport model, Tech. Rep. 29, Division of Atmospheric Research Technical Paper, CSIRO, Australia, http://www.csa.com/partners/viewrecord.php?requester=gs&collection=TRD&recid=N9414095AH, 1993.

Frey, K. E. and Smith, L. C.: How well do we know northern land cover? Comparison of four global vegetation and wetland products with a new ground-truth database for West Siberia, Global Biogeochemical Cycles, 21, doi:10.1029/2006GB002706, http://onlinelibrary.wiley.com/doi/10.1029/2006GB002706/abstract, 2007.

Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P., Vermeulen, A. T., Dargaville, R., Aalto, T., Brandt, J., Christensen, J. H., Frohn, L. M., Haszpra, L., Karstens, U., Rödenbeck, C., Ramonet, M., Carboni, G., and Santaguida, R.: Comparing atmospheric transport models for future regional inversions over Europe - Part 1: mapping the atmospheric CO$_2$ signals, Atmos. Chem. Phys., 7, 3461–3479, doi:10.5194/acp-7-3461-2007, http://www.atmos-chem-phys.net/7/3461/2007/, 2007.

Giglio, L., Loboda, T., Roy, D. P., Quayle, B., and Justice, C. O.: An active-fire based burned area mapping algorithm for the MODIS sensor, Remote Sensing of Environment, 113, 408–420, doi:10.1016/j.rse.2008.10.006, http://www.sciencedirect.com/science/article/pii/S0034425708003180, 2009.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi,

T., and Yuen, C.-W.: Towards robust regional estimates of CO$_2$ sources and sinks using atmospheric transport models, Nature, 415, 626–630, doi:10.1038/415626a, http://www.nature.com/nature/journal/v415/n6872/abs/415626a.html, 2002.

Hargreaves, K. J. and Fowler, D.: Quantifying the effects of water table and soil temperature on the emission of methane from peat wetland at the field scale, Atmospheric Environment, 32, 3275–3282, doi:10.1016/S1352-2310(98)00082-X, http://www.sciencedirect.com/science/article/pii/S135223109800082X, 1998.

Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, J. Geophys. Res., 106, 4703–4715, doi:10.1029/2000JD900581, http://onlinelibrary.wiley.com/doi/10.1029/2000JD900581/abstract, 2001.

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith, P., Castaldi, S., Chevallier, F., Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J.-F., Langenfelds, R. L., Le Quéré, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter, B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I. J., Spahni, R., Steele, L. P., Strode, S. A., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E., and Zeng, G.: Three decades of global methane sources and sinks, Nature Geosci, 6, 813–823, doi:10.1038/ngeo1955, http://www.nature.com/ngeo/journal/v6/n10/full/ngeo1955.html, 2013.

Koohkan, M. R. and Bocquet, M.: Accounting for representativeness errors in the inversion of atmospheric constituent emissions: application to the retrieval of regional carbon monoxide fluxes, Tellus B, 64, doi:10.3402/tellusb.v64i0.19047, http://www.tellusb.net/index.php/tellusb/article/view/19047/0, 2012.

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

Lauvaux, T. and Davis, K. J.: Planetary boundary layer errors in mesoscale inversions of column-integrated CO$_2$ measurements, Journal of Geophysical Research: Atmospheres, 119, 490–508, doi:10.1002/2013JD020175, http://onlinelibrary.wiley.com/doi/10.1002/2013JD020175/abstract, 2014.

Lauvaux, T., Pannekoucke, O., Sarrat, C., Chevallier, F., Ciais, P., Noilhan, J., Rayner, P. J., et al.: Structure of the transport uncertainty in mesoscale inversions of CO$_2$ sources and sinks using ensemble model simulations, Biogeosciences, 6, 1089–1102, http://content.imamu.edu.sa/Scholars/it/net/bg-6-1089-2009.pdf, 2009.

Lehner, B. and Döll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, Journal of Hydrology, 296, 1–22, doi:10.1016/j.jhydrol.2004.03.028, http://www.sciencedirect.com/science/article/pii/S0022169404001404, 2004.

Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the upstream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, Journal of Geophysical Research: Atmospheres, 108, doi:10.1029/2002JD003161, http://onlinelibrary.wiley.com/doi/10.1029/2002JD003161/abstract, 2003.

Locatelli, R., Bousquet, P., Chevallier, F., Fortems-Cheney, A., Szopa, S., Saunois, M., Agusti-Panareda, A., Bergmann, D., Bian, H., Cameron-Smith, P., Chipperfield, M. P., Gloor, E., Houweling, S., Kawa, S. R., Krol, M., Patra, P. K., Prinn, R. G., Rigby, M., Saito, R., and Wilson, C.: Impact of transport model errors on the

global and regional methane emissions estimated by inverse modelling, Atmospheric Chemistry and Physics Discussions, 13, 10 961–11 021, doi:10.5194/acpd-13-10961-2013, http://www.atmos-chem-phys-discuss. net/13/10961/2013/acpd-13-10961-2013.html, 2013.

Macdonald, J. A., Fowler, D., Hargreaves, K. J., Skiba, U., Leith, I. D., and Murray, M. B.: Methane emission rates from a northern wetland; response to temperature, water table and transport, Atmospheric Environment, 32, 3219–3227, doi:10.1016/S1352-2310(97)00464-0, http://www.sciencedirect.com/science/article/ pii/S1352231097004640, 1998.

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., and Vivanco, M. G.: CHIMERE 2013: a model for regional atmospheric composition modelling, Geosci. Model Dev., 6, 981–1028, doi:10.5194/gmd-6-981-2013, http://www.geosci-model-dev.net/6/981/2013/, 2013.

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.

Olivier, J. G. J., Van Aardenne, J. A., Dentener, F. J., Pagliari, V., Ganzeveld, L. N., and Peters, J. A. H. W.: Recent trends in global greenhouse gas emissions:regional trends 1970–2000 and spatial distributionof key sources in 2000, Environmental Sciences, 2, 81–99, doi:10.1080/15693430500400345, http://dx.doi.org/10. 1080/15693430500400345, 2005.

Peylin, P., Baker, D., Sarmiento, J., Ciais, P., and Bousquet, P.: Influence of transport uncertainty on annual mean and seasonal inversions of atmospheric CO$_2$ data, J. Geophys. Res., 107, 4385, doi:10.1029/2001JD000857, http://www.agu.org/pubs/crossref/2002/2001JD000857.shtml, 2002.

Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO$_2$ inversions, Biogeosciences, 10, 6699–6720, doi:10.5194/bg-10-6699-2013, http://www.biogeosciences.net/10/6699/2013/, 2013.

Pison, I., Menut, L., and Bergametti, G.: Inverse modeling of surface NOx anthropogenic emission fluxes in the Paris area during the Air Pollution Over Paris Region (ESQUIF) campaign, J. Geophys. Res., 112, D24 302, doi:10.1029/2007JD008871, http://www.agu.org/pubs/crossref/2007/2007JD008871.shtml, 2007.

Pison, I., Bousquet, P., Chevallier, F., Szopa, S., and Hauglustaine, D.: Multi-species inversion of CH4, CO and H2 emissions from surface measurements, Atmos. Chem. Phys., 9, 5281–5297, doi:10.5194/acp-9-5281-2009, http://www.atmos-chem-phys.net/9/5281/2009/, 2009.

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., and Klooster, S. A.: Terrestrial ecosystem production: A process model based on global satellite and surface data, Global Biogeochem. Cycles, 7, 811–841, doi:10.1029/93GB02725, http://onlinelibrary.wiley.com.biblioplanets.gate. inist.fr/doi/10.1029/93GB02725/abstract, 1993.

Reshetnikov, A. I., Paramonova, N. N., and Shashkov, A. A.: An evaluation of historical methane emissions from the Soviet gas industry, Journal of Geophysical Research: Atmospheres, 105, 3517–3529, doi:10.1029/1999JD900761, http://onlinelibrary.wiley.com/doi/10.1029/1999JD900761/abstract, 2000.

Ringeval, B., Houweling, S., van Bodegom, P. M., Spahni, R., van Beek, R., Joos, F., and Röckmann, T.: Methane emissions from floodplains in the Amazon Basin: challenges in developing a process-based model for global applications, Biogeosciences, 11, 1519–1558, doi:10.5194/bg-11-1519-2014, http://www.biogeosciences.net/11/1519/2014/, 2014.

Rypdal, K. and Winiwarter, W.: Uncertainties in greenhouse gas emission inventories - evaluation, comparability and implications, Environmental Science & Policy, 4, 107–116, doi:10.1016/S1462-9011(00)00113-1, http://www.sciencedirect.com/science/article/pii/S1462901100001131, 2001.

Sasakawa, M., Shimoyama, K., Machida, T., Tsuda, N., Suto, H., Arshinov, M., Davydov, D., Fofonov, A., Krasnov, O., Saeki, T., Koyama, Y., and Maksyutov, S.: Continuous measurements of methane from a tower network over Siberia, Tellus B, 62, 403–416, doi:10.1111/j.1600-0889.2010.00494.x, http://onlinelibrary.wiley.com/doi/10.1111/j.1600-0889.2010.00494.x/abstract, 2010.

Schmidt, H., Derognat, C., Vautard, R., and Beekmann, M.: A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe, Atmos. Environ., 35, 6277–6297, http://www.sciencedirect.com/science/article/pii/S1352231001004514, 2001.

Schwinger, J. and Elbern, H.: Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: A posteriori validation of error statistics in observation space, Journal of Geophysical Research: Atmospheres, 115, n/a–n/a, doi:10.1029/2009JD013115, http://onlinelibrary.wiley.com/doi/10.1029/2009JD013115/abstract, 2010.

Spahni, R., Wania, R., Neef, L., van Weele, M., Pison, I., Bousquet, P., Frankenberg, C., Foster, P. N., Joos, F., Prentice, I. C., and van Velthoven, P.: Constraining global methane emissions and uptake by ecosystems, Biogeosciences, 8, 1643–1665, doi:10.5194/bg-8-1643-2011, http://www.biogeosciences.net/8/1643/2011/, 2011.

Spahni, R., Joos, F., Stocker, B. D., Steinacher, M., and Yu, Z. C.: Transient simulations of the carbon and nitrogen dynamics in northern peatlands: from the Last Glacial Maximum to the 21st century, Clim. Past, 9, 1287–1308, doi:10.5194/cp-9-1287-2013, http://www.clim-past.net/9/1287/2013/, 2013.

Stohl, A., Forster, C., Frank, A., Seibert, P., Wotawa, G., et al.: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2, Atmos. Chem. Phys., 5, 2461–2474, http://hal-insu.archives-ouvertes.fr/hal-00301615/, 2005.

Tarantola, A.: Inverse Problem Theory, Elsevier, New York, 1987.

Thompson, R. L. and Stohl, A.: FLEXINVERT: an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid, Geosci. Model Dev. Discuss., 7, 3751–3801, doi:10.5194/gmdd-7-3751-2014, http://www.geosci-model-dev-discuss.net/7/3751/2014/, 2014.

Tolk, L. F., Meesters, A. G. C. A., Dolman, A. J., and Peters, W.: Modelling representation errors of atmospheric CO_2 mixing ratios at a regional scale, Atmos. Chem. Phys., 8, 6587–6596, doi:10.5194/acp-8-6587-2008, http://www.atmos-chem-phys.net/8/6587/2008/, 2008.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. a. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A.,

Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q.J.R. Meteorol. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, http://onlinelibrary. wiley.com/doi/10.1256/qj.04.176/abstract, 2005.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10, 11 707–11 735, doi:10.5194/acp-10-11707-2010, http://www.atmos-chem-phys.net/10/11707/2010/, 2010.

Vautard, R., Beekmann, M., Roux, J., and Gombert, D.: Validation of a hybrid forecasting system for the ozone concentrations over the Paris area, Atmospheric Environment, 35, 2449–2461, doi:10.1016/S1352-2310(00)00466-0, http://www.sciencedirect.com/science/article/pii/S1352231000004660, 2001.

Williams, I. N., Riley, W. J., Torn, M. S., Biraud, S. C., and Fischer, M. L.: Biases in regional carbon budgets from covariation of surface fluxes and weather in transport model inversions, Atmospheric Chemistry and Physics Discussions, 13, 19 051–19 083, doi:10.5194/acpd-13-19051-2013, http://www. atmos-chem-phys-discuss.net/13/19051/2013/acpd-13-19051-2013.html, 2013.

Winderlich, J., Chen, H., Gerbig, C., Seifert, T., Kolle, O., Lavrič, J. V., Kaiser, C., Höfer, A., and Heimann, M.: Continuous low-maintenance CO_2/CH_4/H_2O measurements at the Zotino Tall Tower Observatory (ZOTTO) in Central Siberia, Atmospheric Measurement Techniques, 3, 1113–1128, doi:10.5194/amt-3-1113-2010, http://www.atmos-meas-tech.net/3/1113/2010/amt-3-1113-2010.html, 2010.

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.

WMO/GAW: 16th WMO/IAEA Meeting on Carbon Dioxide, Other Greenhouse Gases, and Related Measurement Techniques (GGMT-2011), Tech. Rep. 206, World Meteorological Organization Global Atmospheric Watch, Wellington, New Zealand, http://www.wmo.int/pages/prog/arep/gaw/documents/Final_GAW_206_web.pdf, 2011.

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, http://cerea.enpc.fr/HomePages/wu/file/wu11carbon.pdf, 2011.

Wu, L., Bocquet, M., Chevallier, F., Lauvaux, T., and Davis, K.: Hyperparameter estimation for uncertainty quantification in mesoscale carbon dioxide inversions, Tellus B, 65, doi:10.3402/tellusb.v65i0.20894, http://www.tellusb.net/index.php/tellusb/article/view/20894, 2013.

Table 1: OSSEs summary. Three parameters of the inversion (sub-total masses emitted per regions, emission distribution and transport) can be perturbed compared with the 'truth'. The seven possible combinations are depicted by $=$ and $\neq$ signs for each parameter and each OSSE. Every OSSE is evaluated along the scores defined in Sect. 4.2. The scores are given in % for the best correlation threshold for grouping the state space components as presented in Sect. 4.2. The influence score must be as close to 100% as possible. The other two scores must be as small as possible. The regions are grouped along a correlation criterion r_{max} (see Sect. 4.2); we present the scores only for r_{max} with the best results. For OSSE 7, the scores are zeros for the fossil fuel regions because most of these regions were filtered out. The few remaining ones are very well constrained.

	OSSE 1		OSSE 2		OSSE 3		OSSE 4		OSSE 5		OSSE 6		OSSE 7	
Inversion inputs:														
x sub-totals	$\neq$		$=$		$=$		$\neq$		$\neq$		$=$		$\neq$	
x distributions	$=$		$\neq$		$=$		$\neq$		$=$		$\neq$		$\neq$	
H	$=$		$=$		$\neq$		$=$		$\neq$		$\neq$		$\neq$	
Optimal r_{max}	0.5		0.5		0.5		0.5		0.6		0.5		0.4	
Scores:	ff	wet	ff	wet	ff	wet	ff	wet	ff	wet	ff	wet	ff	wet
Relative score	79	94	16	27	40	84	3	66	30	117	20	93	0	112
Absolute score	9	16	2	11	36	24	1	27	18	40	37	30	0	15
Influence	63	56	39	37	45	30	37	28	46	58	32	32	13	33

Table 2: Eurasian site characteristics (Sect. 5.3). The altitudes of the sites are given as m above sea level (a.s.l.) and the inlet height is in m above ground level (a.g.l.).

| Station | ID | Location | | | Inlet |
| | | Lon | Lat | Alt | height |
		(° E)	(° N)	(m a.s.l)	(m a.g.l.)
Azovo	AZV	73.03	54.71	100	50
Berezorechka	BRZ	84.33	56.15	150	80
Demyanskoe	DEM	70.87	59.79	75	63
Igrim	IGR	64.42	63.19	25	47
Karasevoe	KRS	82.42	58.25	50	67
Noyabrsk	NOY	75.78	63.43	100	43
Pallas	PAL	24.12	67.97	560	5
Shangdianzi	SDZ	117.12	40.65	287	0
Tae-ahn Peninsula	TAP	126.12	36.72	20	0
Ulaan Uul	UUM	11.08	44.45	914	0
Vaganovo	VGN	62.32	54.50	200	85
Yakutsk	YAK	129.36	62.09	210	77
Zotino	ZOT	89.35	60.80	104	301
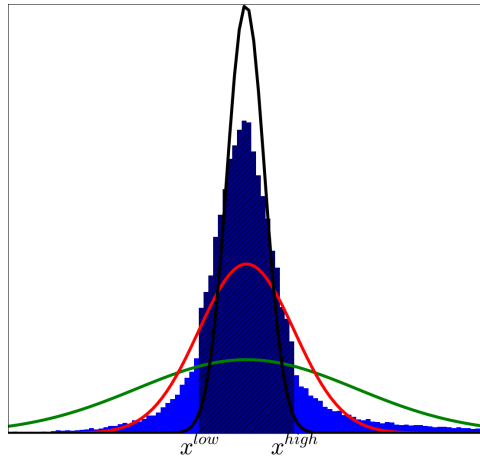
Figure 1: Distribution of one component of the Monte Carlo posterior ensemble. The histogram displays the raw posterior distribution. The dark hatched part of the histogram depicts the proportion of the ensemble within the tolerance interval TI_{68}, $[\mathbf{x}^{\mathrm{low}}, \mathbf{x}^{\mathrm{high}}]$ (as defined in Sect. 2.2). The red curve represents the Normal distribution with the same mode and tolerance interval; the green one stands for a Normal distribution with the same mode and the same standard deviation; the black one is the posterior distribution computed with the maximum likelihood couple of uncertainty matrices, presenting under-estimated skewness compared with the Monte Carlo distribution.

37
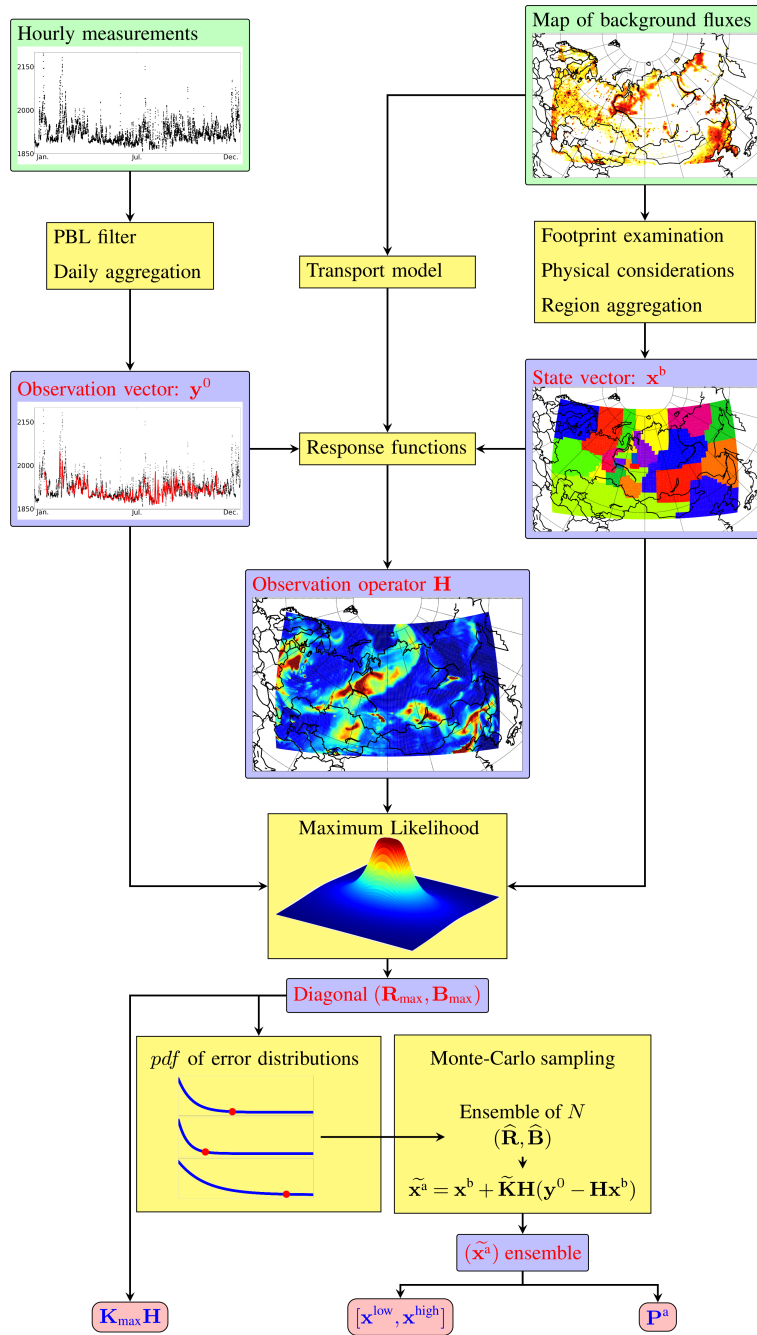
Figure 2: Block diagram of the method. Green boxes represent the raw inputs of the system. The blue ones are intermediary results and red ones the outputs to be interpreted. The yellow ones depict the algorithms to compute. Details in Sect. 2 and 3. Insights for output analyses are given in Sect. 4.2.
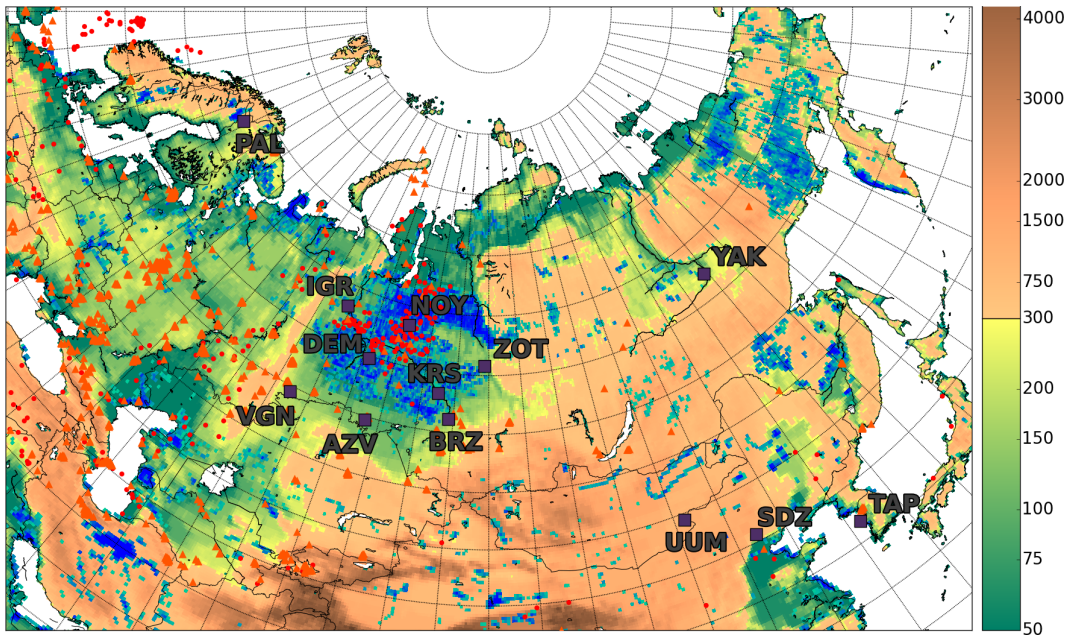
Figure 3: Topographic map of the domain of interest. The colorbar shows the altitude above sea level (from ETOPO1 database; Amante and Eakins, 2009). Red dots (resp. orange triangle) depicts hot spots of CH$_4$ emissions (based on EDGAR v4.2 inventory; see Sect. 5.1) related to oil welling and refineries (resp. gas extraction and leaks during distribution in population centers). Purple squares highlight the observation site localization. Blueish shaded areas represent average inundated regions, wetlands and peatlands (from the Global Lakes and Wetlands Database; Lehner and Döll, 2004)
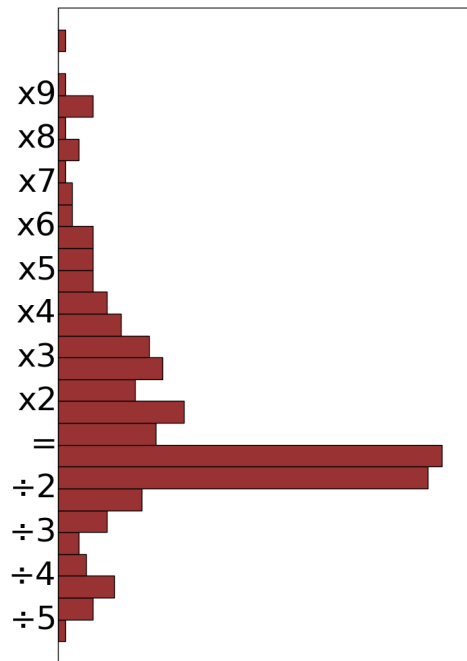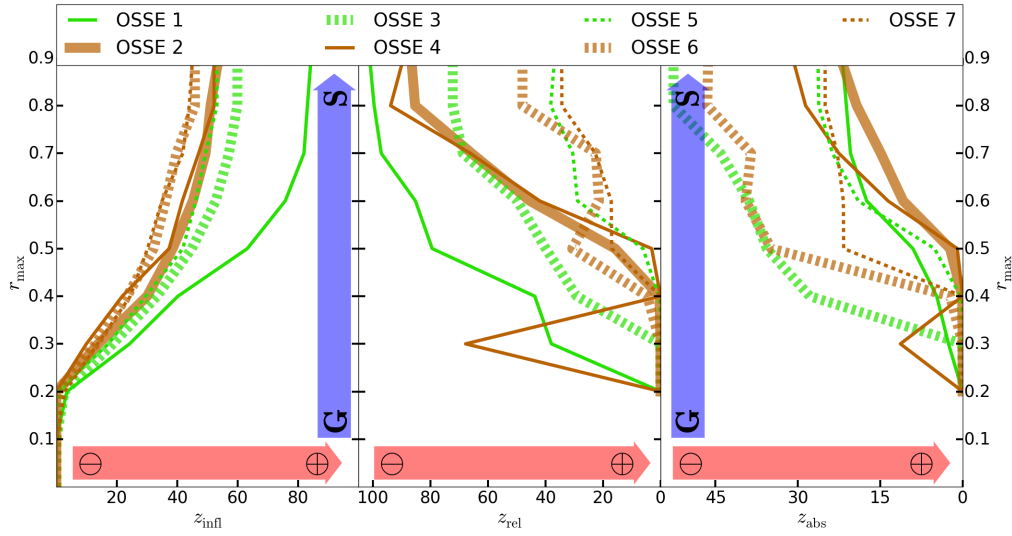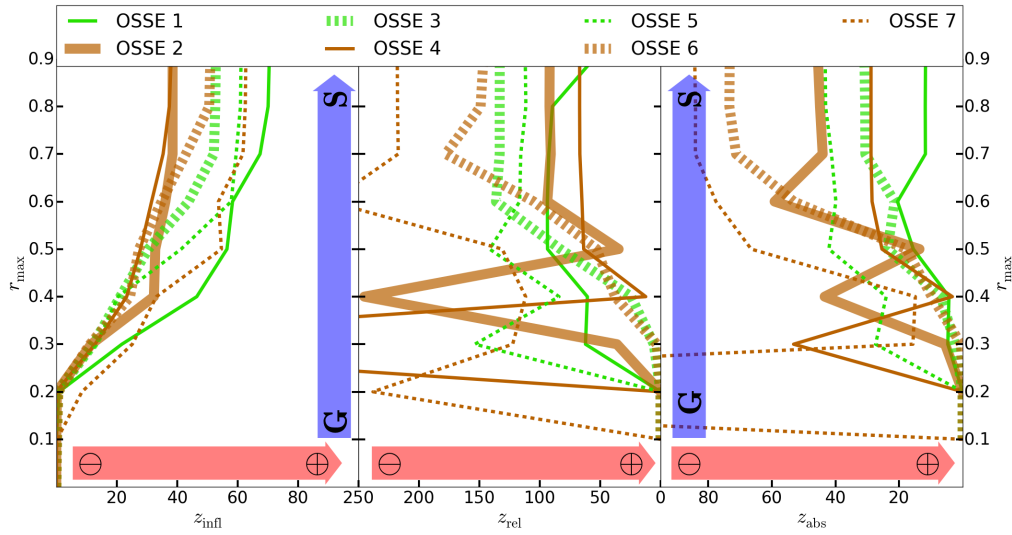
Figure 4: Distribution of the scaling factors applied to the emission databases in order to compute the 'truth'. All the emission components of the state vector have been included in the histogram. The selection of the scaling factor distribution is detailed in Sect. 5.1.

Figure 5: Score comparison on fossil fuel (up) and wetland (bottom) regions for all OSSEs along correlation thresholds r_{max} of region grouping (see details in Sect. 4.2). (left) Influence correlation $\mathbf{z}_{infl}$ profile. (center) Relative score $\mathbf{z}_{rel}$ correlation profile. (right) Absolute score $\mathbf{z}_{abs}$ correlation profile. The red arrows depict the direction from lowest scores to best ones. The blue arrows denote the direction of grouping, from all grouped ('G', $r_{max} = 0$) to all separated ('S', $r_{max} = 1$). The OSSEs are indexed along Tab. 1 numerotation. Thin (resp. thick) lines stand for correct (resp. perturbed) sub-total emissions. Green (resp. brown) lines depict correct (resp. perturbed) emission distributions. Solid (resp. dotted) lines represent correct (resp. perturbed) transport. As in Sect. 4.2, the scores are noted in %.

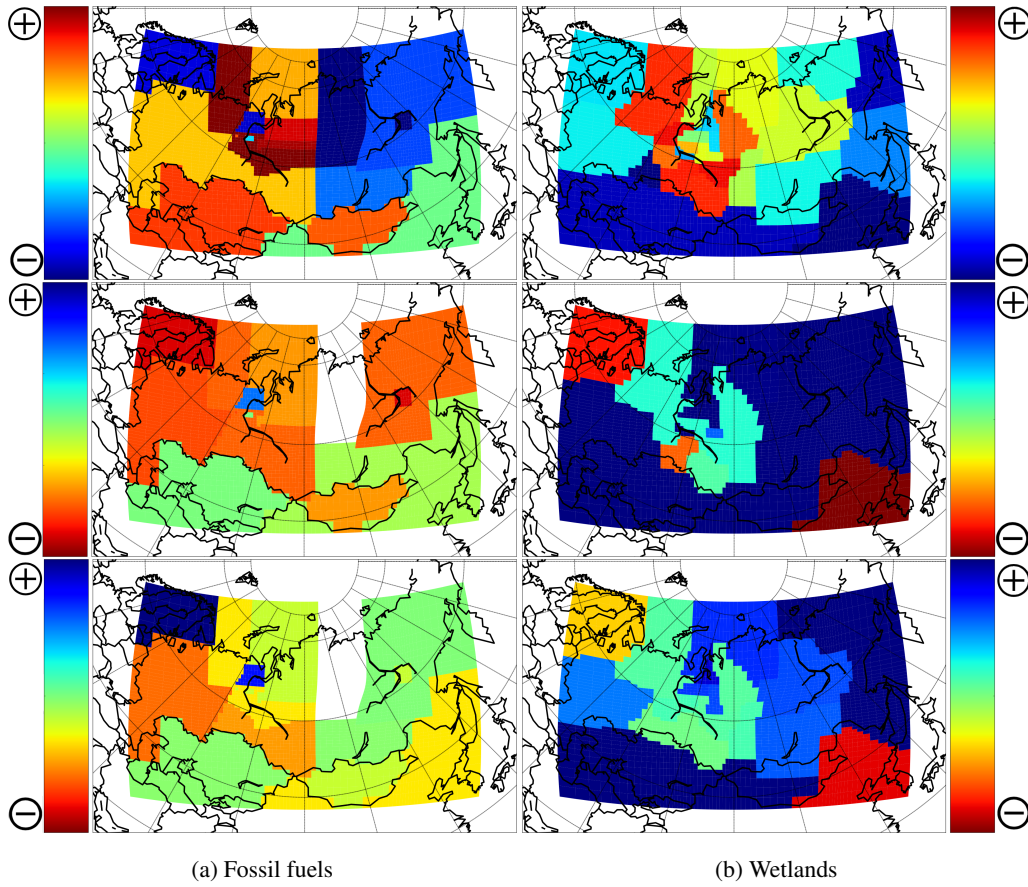(a) Fossil fuels

(b) Wetlands

Figure 6: Map of the average scores as defined in Sect. 4.2 for OSSE 1 (see Tab. 1) projected on the aggregation grid defined in Sect. 3. (up) Influence score $\mathbf{z}_{\text{infl}}$. (middle) Relative score $\mathbf{z}_{\text{rel}}$. (bottom) Absolute score $\mathbf{z}_{\text{abs}}$. The colour maps have been chosen so that redder regions correspond to better scores (denoted by $\ominus$ and $\oplus$ symbols). The zoom and map physical projection are the same as in Fig. 3.