Geosci. Model Dev. Discuss., 7, C1266–C1270, 2014
www.geosci-model-dev-discuss.net/7/C1266/2014/

Geoscientific Model Development Discussions

Open Access

GMDD

7, C1266–C1270, 2014

Interactive Comment

Interactive comment on "FLEXINVERT: an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid" *by* R. L. Thompson and A. Stohl

Anonymous Referee #3

Received and published: 29 July 2014

The manuscript describes an inverse emission estimation framework "FLEXINVERT" for an analytical Bayesian inversion tailored for backward Lagrangian transport models like FLEXPART. Although most of the elements are well known and have been described elsewhere, there are two main reasons why I recommend publication of the manuscript:

First of all, FLEXINVERT is presented as a comprehensive framework that combines essentially all of the elements required for such a system including a proper consideration of background concentrations (necessary because the backward transport simulation accounts only for fluxes during the recent history of an air parcel) , the computation of a variable resolution grid reflecting the true sensitivities of the observation network to upstream fluxes, the definition of observation errors and a priori uncertainties (including their spatial and temporal correlations), and the mathematical and numerical framework to solve for the optimized fluxes and their posterior uncertainties. I particularly appreciate the mathematical rigor the individual components are described with. This will greatly help any user of the framework to understand the individual steps.

The second reason why the manuscript deserves publication is that there are innovative elements that (to my knowledge) have not been presented before, at least not in the same way or to the same level of detail: These include the “Aggregation of background mixing ratios” in section 2.3 and the “Optimization of the fluxes to fine resolution” in section 2.8.

The paper is very well written and structured and well to the point. I thus have no specific recommendations regarding structure or content.

Nevertheless, I have one main concern which deserves more attention by the authors, in particular regarding future users that should also know the potential limitations of the framework: What are the computational costs and what are the corresponding limitations? 4DVAR systems have been developed because inverse problems can easily become too large to be solved analytically. FLEXINVERT is based on an analytical Bayesian inversion which involves operations with large matrices (including matrix inverse) that may become computationally expensive in particular in terms of required memory space. This will necessarily limit the applicability to small- to medium-size problems (limited number of observations/measurements sites, limited spatial and/or temporal resolution of the fluxes). Many operations involve sparse matrices but only standard linear algebra methods appear to be applied which will necessarily lead to many unnecessary computations. The observation error covariance matrix was chosen to be a diagonal matrix. Was this choice driven also by computational constraints? Although the treatment of background concentration looks appealing at first sight, it may involve considerable computational cost: The matrix H_bg may be very (excessively?) large since it is dimensioned M (number of observations) x P (number of grid

GMDD
7, C1266–C1270, 2014

cells of a global model). It would be useful if the authors could provide some information on the memory and computation time requirements e.g. for the CH_4 test case presented in Section 3. I found it difficult to judge which of the individual steps is particularly expensive and may therefore require special attention by a user. FLEXINVERT makes a number of assumptions (which are often made), and a user will have to be aware that these may not necessarily be met: - Errors are assumed to be Gaussian. As shown e.g. by Stohl et al. (2010), measurement – model residuals may be highly skewed. FLEXINVERT may therefore be sensitive to extremes. - Observation errors are assumed to be uncorrelated in time: This is likely a good assumption for weekly flask samples, but it would certainly not be a good assumption for e.g. hourly data, mainly due to correlated transport errors (not only correlated PBL errors but also errors in the wind field). This assumption can be easily verified by analyzing the autocorrelation structure of the residuals. In this way the correlation length can be determined. - The observation-based estimation of background concentrations as presented in Section 2.1.2 will not work for CO_2 which has strong negative fluxes and therefore has no clearly defined baseline. It would be good if the authors could add some word of caution on these points.

Finally, results of an inversion critically depend on the specification of a priori and model-data mis-match errors (and their correlation structure). FLEXINVERT provides a nice framework for solving the problem, but it provides little guidance with respect to the specification of these errors. The authors are obviously aware of the necessity to provide realistic error estimates as they have checked their inversion in the case study (Sect. 3) for the chi-square statistics but it should probably be stated more clearly that it is the task of the user to define these errors in a realistic way. A recent publication addressing this issue is Berchet et al. (ACP, 13, 7115–7132, 2013, doi:10.5194/acp-13-7115-2013).

Minor points:

P3753, lines 14-16: Shouldn't it be "the partial derivative of the change in mixing ratio

GMDD

7, C1266–C1270, 2014

Interactive Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

to the change in fluxes" rather than the reverse?

P3754, line 5: Other regional scale inverse modeling studies involving Lagrangian models would be Keller, C. A., M. Hill, M. K. Vollmer, S. Henne, D. Brunner, S. Reimann, S. O'Doherty, J. Arduini, M. Maione, Z. Ferenczi, L. Haszpra, A. J. Manning and T. Peter, European Emissions of Halogenated Greenhouse Gases Inferred from Atmospheric Measurements, Environ. Sci. Technol., 46, 217-225, doi:10.1021/es202453j Brunner, D., S. Henne, C. A. Keller, S. Reimann, M. K. Vollmer, S. O'Doherty, and M. Maione: An extended Kalman-filter for regional scale inverse emission estimation, Atmos. Chem. Phys., 12, 3455-3478, doi:10.5194/acp-12-3455-2012, 2012.

Page 3756, line 15: Variable resolution grids adapted to the average residence times probably have been introduced for the first time in Manning et al. (JGR, doi:10.1029/2002JD002312, 2003) and have also been applied in other studies such as Vollmer et al. (GRL, doi:10.1029/2009GL038659, 2009), Manning et al. (JGR, doi:10.1029/2010JD014763, 2011), etc. This should not be called the "method of Stohl et al.".

Page 3757, lines 15ff: It should be described more clearly that a typical setup of FLEXPART involves an outer (potentially global) domain and a finer, nested domain. The transport of particles is continued in the outer domain once they leave the nested domain. This setup is not always applicable. Consider e.g. a regional scale model such as FLEXPART-WRF where particles may terminate at the borders rather than being transported further. In this case, termination of particles may occur at any time before the end of the simulation. FLEXINVERT does not seem to be prepared for such a case.

Page 3760, line 8: Estimating background concentrations from observations is a long-standing problem that has been addressed in numerous studies prior to Stohl et al. (2010) and more sophisticated methods have been developed than presented here, e.g. Thoning et al., (JGR 94, 8549–8565, 1989), Ruckstuhl et al. (AMT, doi:10.5194/amt-5-2613-2012, 2012, 2012).

GMDD
7, C1266–C1270, 2014

Page 3763, lines 19-21: I don't understand why this is the most efficient method when the number of observations M is large compared to the dimension of the state vector. The matrix to be inverted has dimension M x M, and if M is large this is a large matrix.

Page 3764, line 21: How are the "8 surrounding grid cells" defined in a variable resolution grid?

Page 3765, line 2: Why is the dimension of B PxP? P was introduced on page 3759 as the dimension of the global model grid.

Interactive comment on Geosci. Model Dev. Discuss., 7, 3751, 2014.