Atmos. Chem. Phys. Discuss., 4, S3684–S3693, 2004
www.atmos-chem-phys.org/acpd/4/S3684/
European Geosciences Union

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Inversion of CO and NO_x emissions using the adjoint of the IMAGES model" *by* J.-F. Müller and T. Stavrakou

J.-F. Müller and T. Stavrakou

Received and published: 21 February 2005

Reply to the Referees' Comments

We would first like to thank the referees for their thorough reviews with many suggestions for improvements and helpful criticism. Their general and specific comments are adressed below.

Referee 1

General comments

1. Number of control parameters

The computational cost is indeed independent on the number of control parameters in the adjoint model approach. Two reasons, however, lie behind our choice of a limited number of control variables. Firstly, we wanted to allow for a direct comparison of our results with previous inversion studies, where the big-region approach is used. Secondly, the small number of control parameters made possible to calculate the a posteriori Hessian matrix by finite differences at a reasonable computational cost, enabling us to evaluate the BFGS and DFP methods for the estimation of posterior errors. Note furthermore, that if using smaller regions for the control parameters can help to decrease the aggregation error, it also implies a loss of information from the a priori emission distributions (which are of course not arbitrary), unless the correlations between the a priori errors on these parameters can be determined and taken into account in the inversion. These correlations might of course be very difficult to estimate. The use of larger sets of control parameters (for instance, in a grid-based approach) will be a primary objective of our future research.

2. Seasonal variations

   We agree with the referee that systematic errors may arise from the use of constant seasonalities. As explained above, however, we wanted to limit the number of control variables in order to allow the calculation of the Hessian by finite differences. We also believe that correlations between the prior errors on the monthly emissions are required if the emission seasonality is allowed to vary in the inversion. The prior emission seasonality represents an important piece of information (deduced e.g. from ATSR fire counts and natural emission models) which should not be completely neglected in this context.

3. Model meteorology and data selection

   We acknowledge that the IMAGES model has limitations related mainly to its coarse resolution, the use of climatological winds, and the absence of synoptic

ACPD

4, S3684–S3693, 2004

variations.

Although the model cannot reproduce exceptional events, it is expected to reproduce monthly averages even when they include conditions labelled as “non-background" by CMDL, since the existence of such conditions appears to be a persistent feature at many sites. Therefore, instead of introducing an additional source of error, the inclusion of “non-background" measurements should rather allow for a more realistic estimation of the temporal variability of the concentration around the monthly mean value.

We agree with the referee that, ideally, the inversion should combine meteorological data and observations for the same period, e.g. by inverting for climatological emissions using observations averaged over several years. Our choice for the year 1997 was mainly driven by our willingness to use GOME NO2 observations (available only for the year 1997 in the version used here) in order to perform a CO-NOx inversion.

4. Effectiveness of inversion in the Southern Hemisphere

As is now better explained in the revised manuscript, our motivation for omitting the “redundant" stations (i.e. those presenting the same information content) is ultimately related to the fact that the model errors at these stations are highly intercorrelated. Model errors are indeed probably as important (or more important) as representativity errors in the expression of the error associated with the data (Eq. 19 of the revised version), a fact that we forgot to mention in the discussion of this expression. Model errors include errors associated to the transport and chemical schemes used in the model, as well as the aggregation errors mentioned previously. These model errors (and a fortiori the correlations between them) are generally difficult to quantify. The combined representativity/model errors are taken equal to 10% at CMDL stations. They are assumed to be mutually independent. The validity of this last assumption is of course not verified in the case of redundant stations, which is why they were omitted. Note that taking all

Antarctic stations into account in the inversion would lead to a 5-fold increase of their weight in the cost function. The resulting posterior uncertainty reduction, however, is unlikely to be justified, given the lack of error independence.

Nevertheless, we have conducted a sensitivity test to evaluate the impact of the Antarctic stations on the inversion results. As expected, including them leads to a better agreement with the observations at these stations (modelled concentrations lower by about 5 ppb). This is achieved by a reduction of the biomass burning emissions in the Southern Hemisphere, as well as of the biogenic and oceanic source (20%, 10% and 30%, respectively). However, the performance of the model at the Tropical stations (in particular, Ascension, Pacific 15 S, South China Sea 6 N) and at the Lauder and Wollongong station gets worse.

In the revised version, model errors are briefly discussed in Section 4.1 after Eq. 19.

Specific comments

1. p.7987

   As asked by the referee, we now include information concerning the validation of IMAGES model in Sect. 2.

2. p. 7988-7989

   As suggested, a short text is added in the Introduction, and a reference to the paper of Houweling et al. (1999) has been added.

3. p. 7996, equation 2

   A short comment about non-Gaussian errors has been added in Sect. 3.1.

   The derivation of the a priori and a posteriori uncertainties from the values of Δf_j is now given in Sect. 7.

4. p. 7997

   It is now explicitly mentioned that the state $s(t_i)$ corresponds to monthly mean concentrations.

5. p. 7999/8000

   We follow the opinion of S. Houweling (Referee 2), who finds helpful to keep this part of the section “Adjoint code generation and minimizer" which provides methodological information on the adjoint technique.

6. p. 8001 (line 4)

   The 15-minute runtime refers indeed to a complete simulation with full chemistry.

7. p. 8001 (line 11)

   About 20-40 iterations are generally required to attain convergence. The optimization includes the spin-up period (9/1996-12/1996), although the model is compared to the data only in 1997.

8. p. 8004

   Mean refers to a 1997-2001 average.

9. p. 8005

   This remark is taken into account in the revised version.

10. p. 8015, line 19-21

    We agree with the referee that the posterior errors are probably underestimated due to the small number of control parameters, in particular for the parameters contributing most to the global budget of CO and NOx. The higher uncertainties which would result from a larger number of parameters would be more realistic, provided that the spatial and/or temporal correlations between the prior uncertainties are taken into account.

11. p. 8016, line 7-8

    Both referees assumed that the error increase reported for some control parameters is an artifact due to the finite difference approximation. In the case of a linear model, when the prior errors are Gaussian and mutually independent, then the a posteriori error distribution is also Gaussian, its variance is lower that the a priori one, and is independent on the observations. It can be shown, however, that non-Gaussian probability density distributions may result in posterior variances larger that the prior values, and dependent on the values of the observations (Talagrand, 2003). In order to obtain a reliable estimation for the Hessian matrix, we performed the calculation for different perturbations around the initial parameter vector ($10^{-2}, 10^{-3}, 10^{-4}$). In all cases, the estimated posterior errors on African anthropogenic and savanna burning emissions were found to be higher than their prior estimates, suggesting that this result should not be attributed to the approximation by finite differences. These higher posterior errors are related to the non-linearity of the inversion scheme, and to the fact that the African continent is practically not constrained by the data in inversion study A. In inversion study B, the NO2 observations provide an additional effective constraint on the CO sources, and reduce the posterior uncertainties. Higher a posteriori errors in non-linear models have been already reported in previous studies (Vukicevic and Bao, 1998; Fennel et al., 2001). A short comment will be included in the revised version of the paper in Sect. 7.

12. p. 8017-8018 (correlation plot)

    A figure showing the correlation matrix is now included in the paper (see Fig. 16).

    The remainder of the comments are taken into account in the revised version.

S. Houweling (Referee 2)

Major comments

ACPD

4, S3684–S3693, 2004

1. Base year 1997

   First of all, we apologize for the misleading description of the biomass burning emissions used in our study. As explained in Olivier et al., (2003), the ATSR fire counts over the period 1997-2001 are used to specify the spatial and temporal distribution of vegetation fires over that period. The average global emissions over the whole period are scaled to the Hao and Liu (1994) value. The 1997 forest fires emissions used in this study are actually about 50% higher than their 1997-2001 average, whereas savanna burning emissions are very close to their climatological average. The fact that the inversion reduces biomass burning emissions is therefore not so surprising, since the a priori values are already quite high.

   In any case, we agree that 1997 was an exceptional year, as far as biomass burning emissions are concerned. As stated previously, our choice for 1997 as a base year was motivated by the availability of GOME NO2 data. Interannual variability for the other emission categories is generally believed to be small, so that the choice of 1997 as base year should have little influence on the results for these emissions. Furthermore, as stated in the paper, the CMDL network is probably too sparse in Tropical regions to provide strong constraints on vegetation fires emissions. GOME observations turned out to be valuable for constraining savanna burning emissions, but these emissions have shown only little interannual variability over the period 1997-2001 according to the ATSR-based inventory used in this study (Olivier et al., 2003). Therefore, quite unfortunately, forest fires remain poorly unconstrained, in spite of their importance for interannual variability. A natural conclusion is that our best hope to reduce the uncertainties on these emissions is to perform multi-year inversions.

2. Major suggestion

   The main purpose of this study was to show the feasibility and interest of the adjoint model technique for optimizing ozone precursor emissions. This technique

ACPD
4, S3684–S3693, 2004

offers two main advantages: 1) coupled chemistry inversions, which are the subject of this paper, and which we think we demonstrated to be useful (chemical feedbacks, common sources), and 2) the possibility to handle an arbitrarily large number of control parameters, which is not explored in this paper but which could represent a major advance in future studies. SCIAMACHY, Aura, METOP and other important missions are going to provide a plethora of tropospheric chemical observations. The adjoint model technique might be one significant tool for the scientific exploitation of these global, high-resolution data.

Minor points

1. p. 7987

   We agree with the referee on the importance of model errors. Their quantification, however, stands beyond the scope of the present work. We have performed a comparison of a priori and optimized O3 concentrations with sonde and aircraft measurements. It did not show clear improvements, however, and we did not include these plots in the paper.

2. p. 7998

   The chi-square values for the different optimizations runs are now given in Table 10. The posterior chi-square for CO, on the order of 2.2, suggests that the model/representativity error (10%) might have been underestimated by a factor of 1.5 or so. Adopting a value of 15% for these errors (instead of 10%) would lead to results identical to those presented in the paper, provided that the value of the regularization parameter Ω is decreased by a factor 1.5 (i.e. Ω=5/1.5=3.3). The best justification for the relative weigth of the prior constraint adopted in this study resides in the results of the sensitivity tests B1 and B2, since the model/data agreement is seen to be moderately influenced when the value of Ω is doubled or

ACPD

4, S3684–S3693, 2004

Interactive Comment

halved. For instance, doubling the prior errors leads to a very minor improvement (a few percents) of the chi-square for both CO and NOx (see Table 10).

On the question regarding the selection of CMDL stations, as explained in our reply to Referee 1, our real argument for omitting “redundant" stations is to reduce the biases caused by model errors, including (as suggested by Referee 2) the use of large regions. This is now explained more clearly in the revised manuscript.

3. p. 8011

   It is very unlikely that errors in the seasonality of fossil emissions can explain such large differences between the model and the GOME data. It is indeed widely accepted that this seasonality should be relatively weak. More probable explanations include model errors (which are often season dependent), GOME retrieval errors (id.), and the lack of averaging kernels (which could account for differences between the vertical profiles used in the retrieval and those predicted by the IMAGES model).

   Computational limitations actually don't prevent the use of a larger number of parameters. Our motivations for the current discretization are explained in our response to Referee 1 (under General comments 1 and 2).

4. p. 8001

   About 20-40 iterations are generally required to attain convergence. In order to determine whether the same solution is obtained when starting from different parameter values, several tests with different initial fluxes have been conducted. Although the number of iterations needed to reach convergence was variable, in all cases, the minimization is found to produce the same optimal solution. This indicates that despite the non-linearities, local minima are not found. In the revised version of the paper, a comment is added in the end of Section 3.2.

   The methane lifetime is now added in Table 10.

ACPD

4, S3684–S3693, 2004

5. p. 8016

The point concerning a posteriori errors, also raised by Referee 1, is discussed in page 3-4 (Specific Comment 11).

Note that what the referee calls “data gaps" in Figures 6 and 7, corresponds to values of the tropospheric NO2 column lower than $5 \cdot 10^{14}$ molec.cm^{-2}.

References:

Talagrand, O: Bayesian estimation, optimal interpolation, statistical linear estimation, R. Swinbank et al. (eds.), Data Assimilation for the Earth System, 21–35, Kluwer Academic Publishers, 2003.

Vukicevic, T. and Bao, J.-W.: The effect of linearization errors on 4DVar Data Assimilation, Mon. Wea. Rev., 126., 1998.

Fennel, K., Losch, M., Schröter, J., Wenzel, M.: Testing a marine ecosystem model: sensitivity analysis and parameter optimization, J. Marine Systems, 28, 45–63, 2001.

ACPD

4, S3684–S3693, 2004

Interactive Comment