**Atmospheric
Chemistry
and Physics
Discussions**

Technical Note: A novel approach to estimation of time-variable surface sources and sinks of carbon dioxide using empirical orthogonal functions and the Kalman filter

R. Zhuravlev[1], B. Khattatov[2], B. Kiryushov[1], and S. Maksyutov[3]

[1]Central Aerological Observatory, Dolgoprudny, Russia
[2]Fusion Numerics International, Boulder, CO, USA
[3]Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan

Correspondence to: R. Zhuravlev (ruslan.zhuravlev@gmail.com)

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

ACPD

11, 1367–1384, 2011

Estimation surface fluxes

R. Zhuravlev et al.

Title Page

Abstract	Introduction
Conclusions	References
Tables	Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

In this work we propose an approach to solving a source estimation problem based on representation of carbon dioxide surface emissions as a linear combination of a finite number of pre-computed empirical orthogonal functions (EOFs). We used NIES transport model for computing response functions and Kalman filter for estimating carbon dioxide emissions. Our approach produces results similar to these of other models participating in the TransCom3 experiment, while being more advantageous in that it is more computationally efficient, produces smooth emission fields, and yields smaller errors than the traditional region-based approach. Additionally, the proposed approach does not require additional effort of defining independent self-contained emission regions.

1 Introduction

It is well known that greenhouse gases and, in particular, greenhouse gases of anthropogenic origin, influence the Earth climate to a great extend. Accurate estimates of strengths, and spatial and temporal variability of the surface sources and sinks of greenhouse gases are thus of great interest to both the scientific community and the policy makers. Carbon dioxide (CO_2) is the most important greenhouse gas of anthropogenic origin that affects radiative balance of the atmosphere and, eventually, the climate. Observations of CO_2 concentrations in the atmosphere demonstrated short-time variability and spatial patterns reflecting influence of time-variable strengths of regional surface sources and sinks of carbon dioxide.

The objective of this study is to estimate absolute contributions of various surface regions to the total carbon dioxide budget at relatively short time scales and in a computationally efficient manner.

A traditional approach to solving this problem includes dividing the Earth's surface by a number of non-overlapping regions and estimating the strengths of sources and

sinks for each one of them. One of the most well-known and successful experiments following this approach was TransCom-3 (T3, Gurney et al., 2000), which used 22 distinct regions; 11 for land surface and 11 for ocean surface. The shapes of the ocean regions undergo seasonal variability, while the shapes of the surface regions are fixed.

5 In subsequent work of Patra et al., the number of the regions has been increased to 64. In both cases, monthly mean CO_2 surface emissions have been successfully estimated using monthly averaged ground based observations of carbon dioxide concentrations.

Recently (Feng et al., 2009), 144 distinct regions have been used and the time scale of carbon dioxide variability was reduced to 8 days using satellite observations of CO_2.

10 In each region, distribution of CO_2 emission is forced to be smooth, so the resulting emission fields will be piecewise-smooth. Refining the results would necessitate increasing the number of regions, and thus increasing computational requirements, which might prove to be impractical.

Additionally, it is reasonable to assume that at least in some cases there is a cor-

15 relation in emission strengths between the regions. It is usually proposed that such correlation is negligible, but in reality this might not be accurate.

Another computational approach relies on the adjoint of the forward transport model. In this case, influence from each grid box could be estimated, but it could be computationally expensive, and creating adjoint versions of forward models is not straightfor-

20 ward. In addition, the correlation between neighboring cells is still unknown, and the results are likely to be influenced by misspecifications of such correlations.

The main idea of our approach is to use empirical orthogonal functions (EOFs) in place of distinct geographical regions. Use of the EOFs as a tool to reduce degrees of freedom in inverse modeling became a widespread practice in geophysics (e.g.,

25 Wikle and Cressie, 1999, more citations). It is also mentioned (Desbiens et al., 2007) that use of EOF in inversion is similar to truncated SVD technique by Hansen (1987, 1998). We propose representing geographic distribution of surface emissions of carbon dioxide as a linear combination of a number of pre-computed empirical orthogonal functions. This combination contains information about climatological spatial variability

of the emissions as well as statistical correlations between different gridpoints. This approach would yield smooth surface fluxes on a global scale and it does not require additional research for defining independent self-contained emission regions. Since, as shown later in this manuscript, a relatively small number of EOFs is needed to accurately represent the CO_2 surface emissions for a number of sources, the estimation problem becomes fairly inexpensive computationally. Practical applications of the derived EOFs can also be envisioned in a framework of the geostatistical inverse modeling (Michalak et al., 2004), which requires a set of the global flux patterns to approximate optimal flux field.

2 Methodology

As in the T3 experiment we consider separately three kinds of surface CO_2 sources: burning fossil fuel, biosphere exchanges and ocean exchanges. Therefore it is necessary to compute EOFs for each of them.

For computing the EOFs we need spatial and time variable statistics of the surface emissions. This dataset has been obtained from the CarbonTracker web site (http://www.esrl.noaa.gov/gmd/ccgg/carbontracker/). CarbonTracker dataset provides global surface fluxes from 2000 to 2008 at 3 h time intervals for four kinds of surface emissions: burning fossil fuels, biosphere exchanges, ocean exchanges and fires. The spatial resolution of these fields is one by one degree; hence the resolution of the EOFs will be the same.

Similarly to T3 experiment, prior to computing the EOFs mean T3 emission fields were subtracted from the CarbonTracker time-dependent emission fields. The 2-D fields of deviations of surface emissions from the mean are available at 3-hour intervals.

One might ask what averaging time window is appropriate for constructing the error covariance matrix for EOF estimation, from both computational and physical points of view. In order to answer this question, we performed calculations of the EOFs from ensembles 3-hour emissions averaged over 1, 2, 3, and 4 days.

Using original 3-hour emission fields for the EOF analysis appears to be counter-productive, as relatively fast changes in daily solar activity and vegetation activity add statistical variability that is unlikely to be useful for the inversion on much longer transport time scales. This indeed has been confirmed by computing samples of EOFs (not shown) from 3-hour averaged fields.

Figure 1 shows a plot of the number of EOFs needed to reconstruct the original emission fields to within 10% error for the different averaging times (1, 2, 3, and 4 days).

It appears that 3-day averaging results in the least number of EOFs needed to reconstruct the original emission field to the specified accuracy of 10%. This is the averaging time interval adopted in the further calculations.

The 3-day averaged emissions were represented as a time-series of vectors for a 9-year time interval, a total of 1096 vectors. Excluding fossil fuel emissions, these vectors were separated seasonal groups. After that, covariance matrices for the emissions were computed for each group and standard singular value decomposition routine was used to compute the singular vectors (EOFs) of these matrices.

It is desirable to determine the minimum number of EOFs needed to reasonably accurately represent CO_2 surface emissions for each source in order to reduce the dimensionality of the inverse problem. Figure 2 shows standard deviations of monthly averaged emission fields from CarbonTracker project and fields obtained after decomposing on EOFs. It appears that in order to be able to represent emission sources with 10% accuracy one needs to use 5 EOFs for the anthropogenic sources, 49 EOFs for the biospheric sources and 33 EOFs for the ocean sources. Thus the highest possible dimensionality of the inversion problem is 5+(49+33)*12+4=993. In these calculation 4 accounts for the a priori fixed average emission fields: two for the anthropogenic emissions, one for the biosphere and one for the ocean.

To compare our approach with regional methodology, we repeated the T3 experiment Level 2 as described by its protocol and obtained monthly sources and sinks of CO_2. In our project we used Japanese National Institute for Environmental Studies Transport

Model (NIES TM, see Appendix A).

The seasonal inversion in T3 framework consist of a 3 year forward simulation (365 days per year) containing 4 tracers: 2 fields for fossil fuel, biosphere and ocean; and 22 CO_2 tracers (11 terrestrial, 11 oceanic). For each month observation from 75 ground-based stations are used for inversion.

In accordance with T3 protocol, we used NIES model to compute response functions for each pre-computed EOF representing CO_2 emissions at 3-year intervals. Observations of CO_2 concentrations along with the covariance matrix were taken from Transcom3 data set. Initial values for the CO_2 emissions in the inversion experiments described here were set to zero.

In order to estimate the (diagonal) covariance matrix of the state vector (CO_2 emissions) we took monthly CarbonTracker emissions for 2000 through 2008 and represented them as linear combination of chosen EOFs. After that, monthly EOF expansion coefficients were averaged over the 9 years (2000–2008).

We used Kalman filter (formulas 1 and 2) to estimate coefficients of the EOF expansion and errors of the resulting emissions using ground-based observations of CO_2 concentrations. At each inversion time step we utilize 75 observations (y) in order to improve prior estimation of the EOF expansion coefficients using standard Kalman filter procedure (Kalman, 1960):

$$x = x_a + \mathbf{K}(y - H[x_a]) \tag{1}$$
$$\mathbf{K} = \mathbf{B}H^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \tag{2}$$

Here x is the posteriori vector of EOF expansion coefficients; x_a is a priori vector of EOF expansion coefficients; H is observation operator that describes the relationship between the state vector and the observations; $\mathbf{K}$ is the Kalman gain matrix that determines the adjustment to the a priori based on the difference between model and observations and their uncertainties. $\mathbf{R}$ is the observation error covariance matrix, and $\mathbf{B}$ is the a priori error covariance matrix.

1372

3 Results

To comply with the T3 experiment setup, only biosphere and ocean EOFs are used for proper comparison with the regional inversion approach. Global distributions of emission sources after the inversion are shown in Fig. 3, while Fig. 4 shows average distributions of the a priori emissions from T3 experiment. Figure 3 presents results of the inversion for the traditional region-based approach for January and July as well as for the results of the EOF approach described here. Clearly, the overall distributions of the emission fields are similar in shape, giving some confidence in the validity of the EOF approach, yet noticeable quantitative differences are present.

The impact of these differences is quantified in Fig. 5 as RMS error, standard deviation, and systematic error between observations and model simulations of CO_2 distributions for different transport models, and experiments with NIES model using EOF approach.

The EOF-based inversion experiments were performed for 2 scenarios:

1. Using 17 EOF for the biospheric sources and 5 EOFs for the oceanic sources. Thus the total dimensionality of the problem is 22, corresponding to the 22 regions in the T3 experiment (marked 17_5 in Fig. 5);

2. Using 49 EOF for the biospheric sources and 33 EOFs for the oceanic sources, to obtain 10% accuracy in the reconstructed emission fields, as demonstrated by Fig. 2 (marked 49_33 in Fig. 5).

Two additional inversion experiments were performed. The objective of these experiments was to see how the EOF approach can be applied to correcting fossil fuel emissions, this has not been done in the T3 study. Global distributions of emission sources after the inversion are shown in Fig. 6. The experiments were set up as follows:

1. Using 3 EOF for fossil fuel sources, 15 EOF for the biospheric sources and 4 EOFs for the oceanic sources. Thus the total dimensionality of the problem is 22, corresponding to the 22 regions in the T3 experiment (marked 3_15_4 in Fig. 7);

2. Using 5 EOF for fossil fuel, 49 EOF for the biospheric sources and 33 EOFs for the oceanic sources, to obtain 10% accuracy in the reconstructed emission fields, as demonstrated by Fig. 2 (marked 5_49_33 in Fig. 7).

Intercomparison of our experiments is presented in Fig. 8.

4   Conclusions

As one can see from Fig. 5, overall the regional method yields lesser RMS error, while the EOF approach shows a clear tendency to reduce the systematic error. Increasing the dimensionality of the problem (49_33) obviously reduces the RMS error and standard deviation, but adversely affects the systematic error.

Nevertheless, an all experiments involving the EOF approach the systematic error is noticeably smaller than that of the inversion performed using regional approach, except for the inversion done with NCAR's MATCH model. Interestingly, increasing the number of EOFs beyond 49 and 33 (10% threshold error in emission field's reconstruction) does little in terms of improving results.

Another obvious difference between results obtained with the EOF approach is the natural smoothness of the emission fields, as opposed to the largely artificial boundaries present in columns 1 and 3 of Fig. 3 for the regional inversion approach.

Inversion results and computed EOFs of the carbon dioxide emission fields are available from the authors upon request.

Appendix A

The original objective of this model was to simulate the seasonal cycles of the long-lived tracer species at a relatively coarse grid resolution (2.5 to 5° longitude-latitude), and to perform sources/sinks inversion of atmospheric CO_2. The transport model has

been improved by increasing spatial resolution and driven by diurnal cycle resolving meteorology for simulating diurnal-synoptic scale variations.

The model's horizontal and vertical resolutions match those of the meteorological dataset when possible. We use pressure level ECMWF operational analyses at 12-hour time step and 2.5° horizontal resolution in model simulations (ECMWF, 1999; Courtier et al., 1998). While the same horizontal resolution is used in the model, the grid layout is different from meteorological dataset. The first model grid cell in the horizontal plane is located near the South Pole, and is confined between (0° E, 90° S) and (2.5° E, 87.5° S). The last one, at the North Pole, is confined between (357.5° E, 87.5° N) and (0° E, 90° N).

The vertical grid layout was designed to provide enough layers to match the resolution of the wind dataset (ECMWF operational analyses), and variability of the boundary layer height. The winds are interpolated from the meteorological analysis grid to the model grid using bilinear interpolation in longitude and latitude in log-pressure coordinate.

The model is designed to handle constant surface emission fields and seasonally changing emissions in the form of 12 monthly average fields per year. The monthly average emissions are interpolated linearly to daily values, so that on the 15th of each month the emission rate is equal to the monthly average for that month as provided by emission inventory files. The emission inventory fields have higher resolution (1×1°) than the model grid (2.5×2.5°), so the input dataset is mapped to a model grid by computing the overlap area of each input data cell to all model grid data cell. This assures that the global total emission flux is conserved during interpolation.

References

Desbiens, R., Aoki, T., and Yokota, T.: Optimization of GOSAT Atmospheric Retrieval of CO_2 in Presence of Atmospheric Particles Using Empirical Orthogonal Function Representation, EOS transactions, American Geophysical Union, Fall Meeting 2007, abstract #A13D-1506, 2007.

Gurney, K., Law, R., Rayner, P., and Denning, A. S.: TransCom 3 Experimental Protocol, Department of Atmospheric Science, Colorado State University, USA, Paper No. 707, 2000.

Hansen, P. C.: The truncated SVD as a method for regularization, BIT, 27, 534–553, 1987.

Hansen, P. C.: Rank-deficient and discrete Ill-posed problems: numerical aspects of linear inversion, SIAM Monogr. Math. Model. Comput., 4, 247 pp., 1998.

Kalman, R. E.: A new approach to linear filtering and prediction problems, J. Basic Engin., 82(1), 35–45, 1960.

Maksyutov, S., Patra, P. K., Onishi, R., Saeki, T., and Nakazawa, T.: NIES/FRCGC global atmospheric tracer transport model: description, validation, and surface sources and sinks inversion, J. Earth Simul., 9, 3–18, 2008.

Michalak, A. M., Bruhwiler, L., and Tans, P. P.: A geostatistical approach to surface flux estimation of atmospheric trace gases, J. Geophys. Res., 109, D14109, doi:10.1029/2003JD004422, 2004.

Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, P. Natl. Acad. Sci. USA, 104, 48, 18925–18930, 2007.

Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical Recipes in FORTRAN (The Art of Scientific Computing, 2nd edn.), Cambridge University Press, New York, USA, 1992

Wikle, C. K. and Cressie, N.: A dimension reduced approach to space-time Kalman filtering, Biometrika, 86, 815–829, 1999.

Estimation surface fluxes

R. Zhuravlev et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

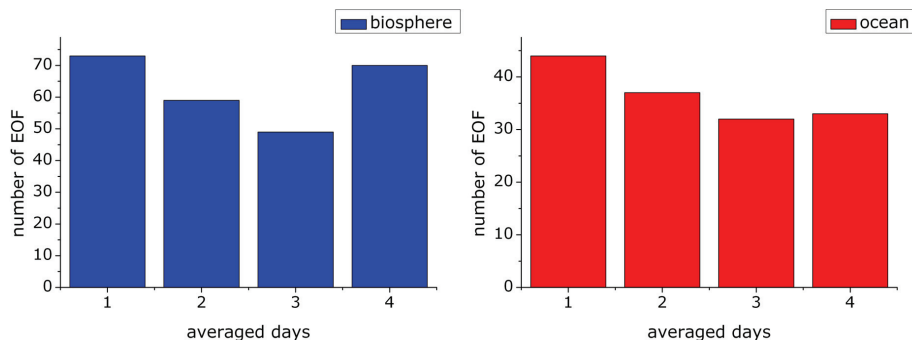Printer-friendly Version

Interactive Discussion

Fig. 1. Number of EOFs needed to restore the original emissions with 10% accuracy for different averaging times. Biosphere emissions are shown on the left, ocean emissions on the right.

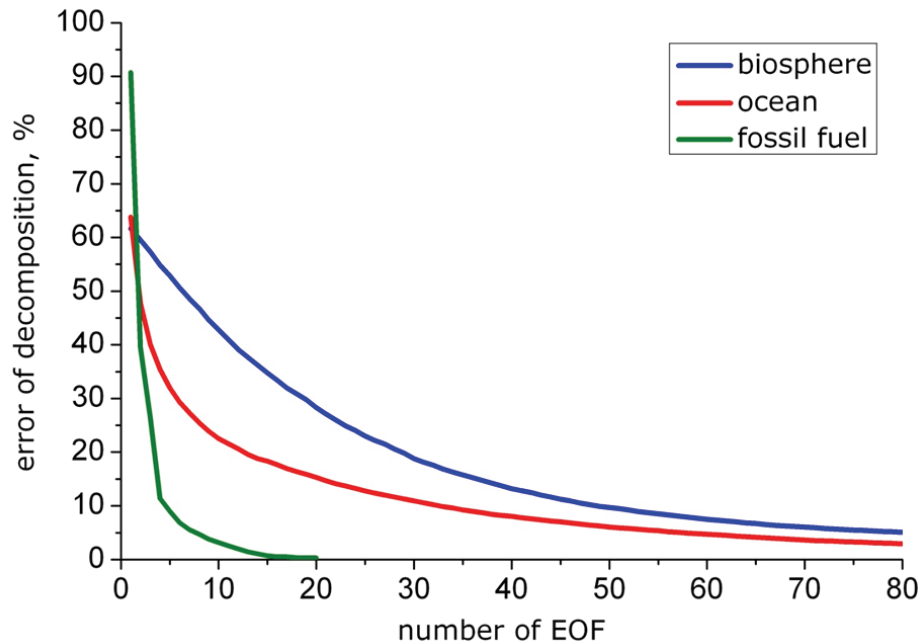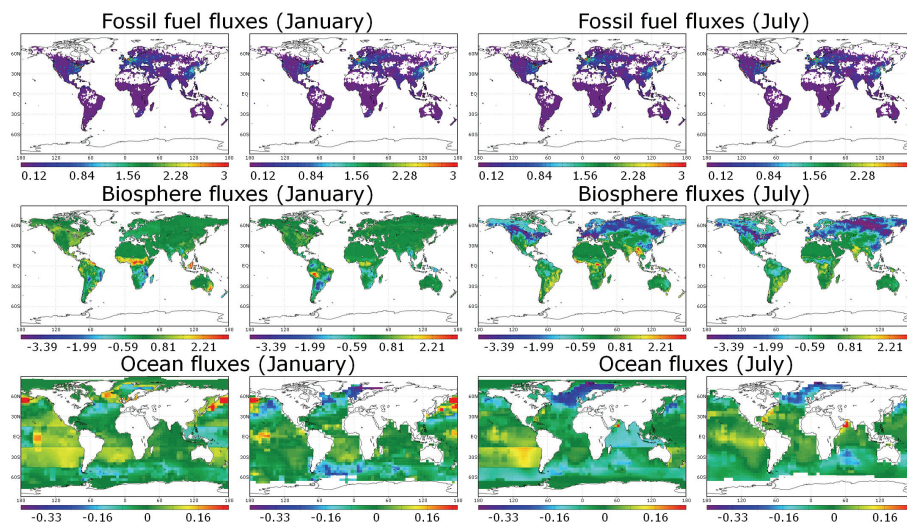**Fig. 2.** RMS deviation (*y*-axis) between original emission fields and those restored depends on number of most significant EOFs (*x*-axis).
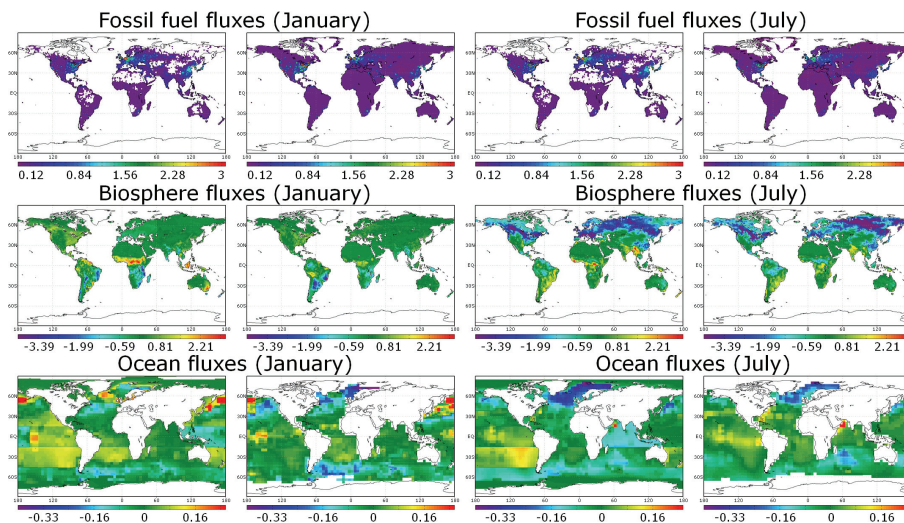
Fig. 3. Estimated fluxes in units of g/m^2/day for January and July using the traditional region approach (columns 1 and 3), and the EOF approach (columns 2 and 4).

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion
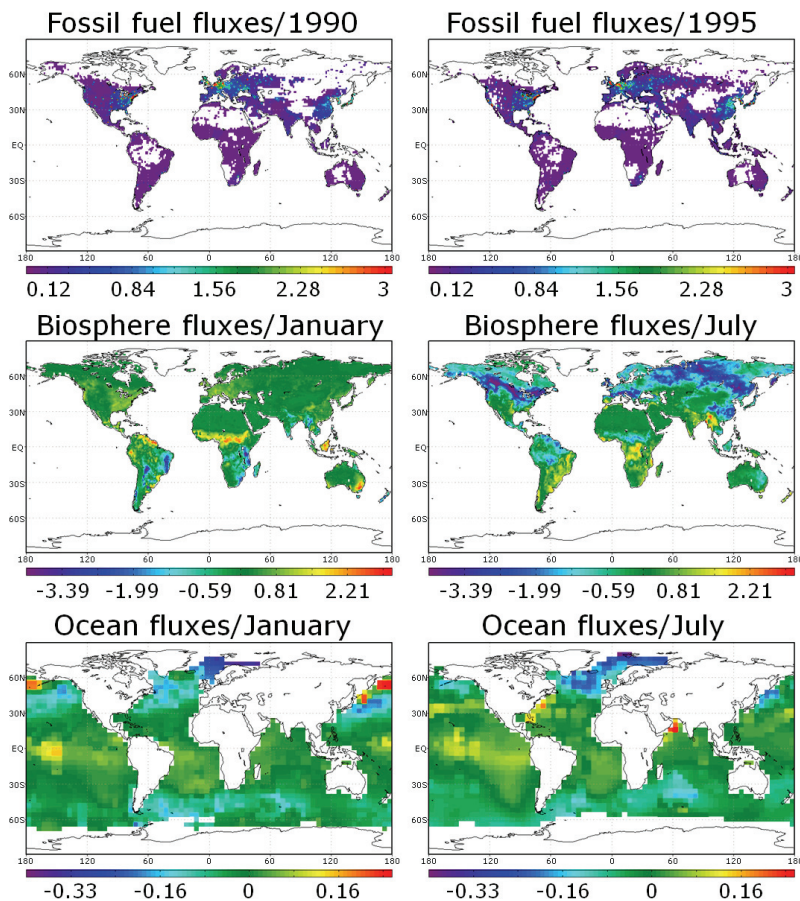
Fig. 4. A priori average emission fields in units of g/m^2/day from T3 experiment for 3 different emission sources: fossil fuel (top figures, year data), biosphere exchange (middle figures, monthly data), ocean exchange (bottom figures, monthly data).
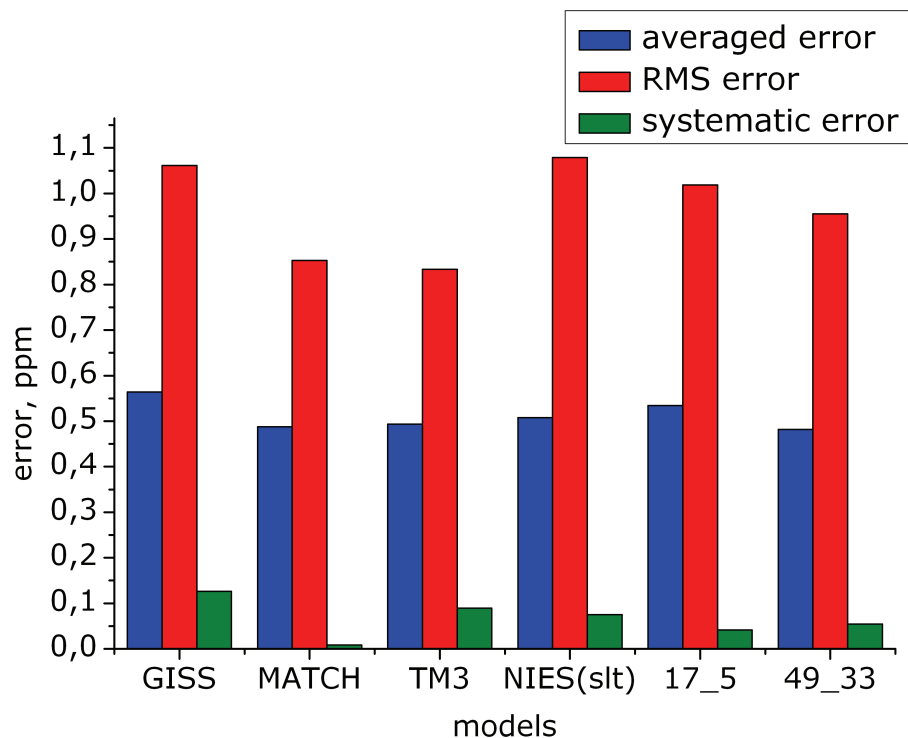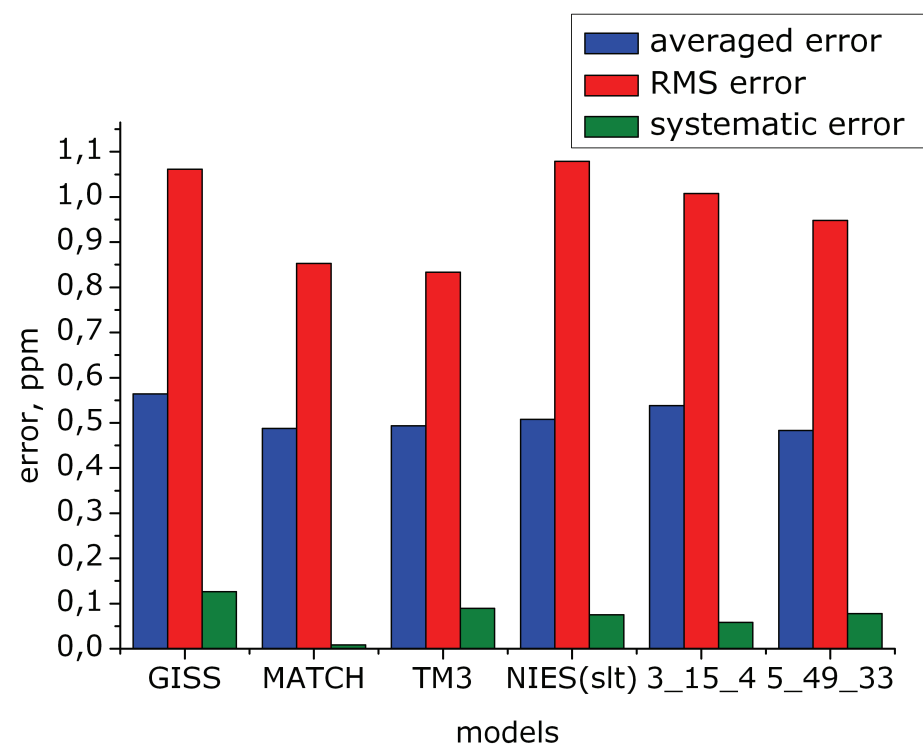
Fig. 5. Averaged error, RMS error, and systematic error between observations and model simulations of CO_2 distributions for different transport models, and experiments with NIES model using EOF approach (last two cases) described in the text.

Fig. 6. Estimated fluxes in units of g/m^2/day for January and July using the traditional region approach (columns 1 and 3), and the EOF approach with addition anthropogenic EOFs (columns 2 and 4).

Fig. 7. Averaged error, RMS error, and systematic error between observations and model simulations of CO_2 distributions for different transport models, and experiments with NIES model using EOF approach with addition anthropogenic EOFs (last two cases) described in the text.
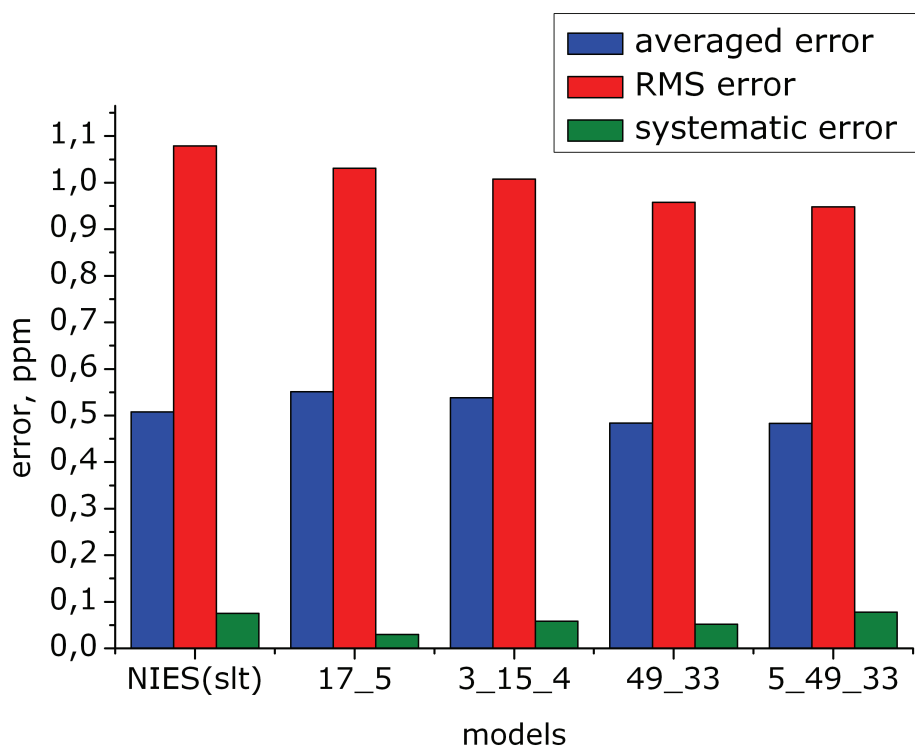
Fig. 8. Averaged error, RMS error, and systematic error between observations and model simulations of CO_2 distributions for NIES TM with regional approach (1 column) and EOFs experiments described in the text.