

# 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<sup>1</sup>, B. Khattatov<sup>2</sup>, B. Kiryushov<sup>1</sup>, and S. Maksyutov<sup>3</sup>

<sup>1</sup>Central Aerological Observatory, Dolgoprudny, Russia

<sup>2</sup>Fusion Numerics International, Boulder, CO, USA

<sup>3</sup>Center for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan

Received: 19 November 2010 – Published in Atmos. Chem. Phys. Discuss.: 18 January 2011

Revised: 23 September 2011 – Accepted: 25 September 2011 – Published: 17 October 2011

**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 National Institute for Environmental Studies (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.

Using the EOFs we can estimate surface fluxes at higher spatial resolution, while keeping the dimensionality of the problem comparable with that in the regions approach. This also allows us to avoid potentially artificial sharp gradients in the fluxes in between pre-defined regions. EOF results generally match observations more closely given the same error structure as the traditional method.

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 extent. 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<sub>2</sub>), is the most important greenhouse gas of anthropogenic origin that affects radiative balance of the atmosphere and, eventually, the climate. Observations of CO<sub>2</sub> 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 research is to study advantages and shortcomings of traditional methods of solving the inverse problem of estimating spatial and temporal variability of surface fluxes and compare them with the EOF approach. While we do not wish to position our approach as superior, we would like to propose it as an alternative that seems to be worth exploring further. Traditional approaches to solving inverse flux estimation problem include 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. Advantage of this technique is that the problem is mathematically over-determined because the number of unknowns is much less than number of available observations. In subsequent work of Patra et al. (2005a, b, 2006), the number of the regions has been increased to 64. In both cases, monthly mean CO<sub>2</sub> surface emissions have been successfully estimated using monthly averaged ground based observations of carbon dioxide concentrations.



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

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<sub>2</sub>. In each region, distribution of CO<sub>2</sub> emission is forced to be smooth, so the resulting emission fields will be piecewise-smooth because of regions were demarcated by hard boundaries. Refining the results would necessitate increasing the number of regions, and thus increasing computational requirements, which might prove to be impractical. Also, lumping small basis regions into larger combined regions may lead to aggregation errors, because observed CO<sub>2</sub> fields are sensitive to the distribution of sources and sinks within large basis regions (Kaminski et al., 2001).

Additionally, it is reasonable to assume that at least in some cases there is a correlation in emission strengths between the regions. It is usually proposed that such correlation is negligible and grid-points within regions are perfectly correlated in space, but in reality this might not be accurate.

Another computational approach relies on the adjoint of the forward transport model (Kaminski et al., 1999). They estimated a coarse grid of fluxes at 8° latitude and 10° longitude in monthly time scales. The problem was underdetermined and a solution was found by gathering a priori information on surface fluxes. In this case, influence from each grid box could be estimated, but it could be computationally expensive due to increasing the number of unknowns, and creating adjoint versions of forward models is not straightforward. In addition, the correlation between neighboring cells is still unknown, and the results are likely to be influenced by misspecifications of such correlations. Rodenbeck (2003) performed a grid-scale inversion assuming different de-correlation length over the land and the ocean using the same spatial resolution as Kaminski (1999). For this approach reality lies somewhere between the two extremes: perfectly correlated or completely uncorrelated fluxes at a grid scale.

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 has become a widespread practice in geophysics (e.g. Wikle and Cressie, 1999). 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 grid-points. 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<sub>2</sub> 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

While solving the inverse problem we utilize the traditional Bayesian approach. We attempt to optimize the cost function consisting a sum of the two parts: a disagreement between the desired solution and the priori estimation, and a disagreement between the desired solution and the observations. Formally, the cost function is given by the following expression:

$$F(x) = \frac{1}{2}(y - \mathbf{H}x)^T \mathbf{R}^{-1}(y - \mathbf{H}x) + \frac{1}{2}(x - x_p)^T \mathbf{B}^{-1}(x - x_p) \quad (1)$$

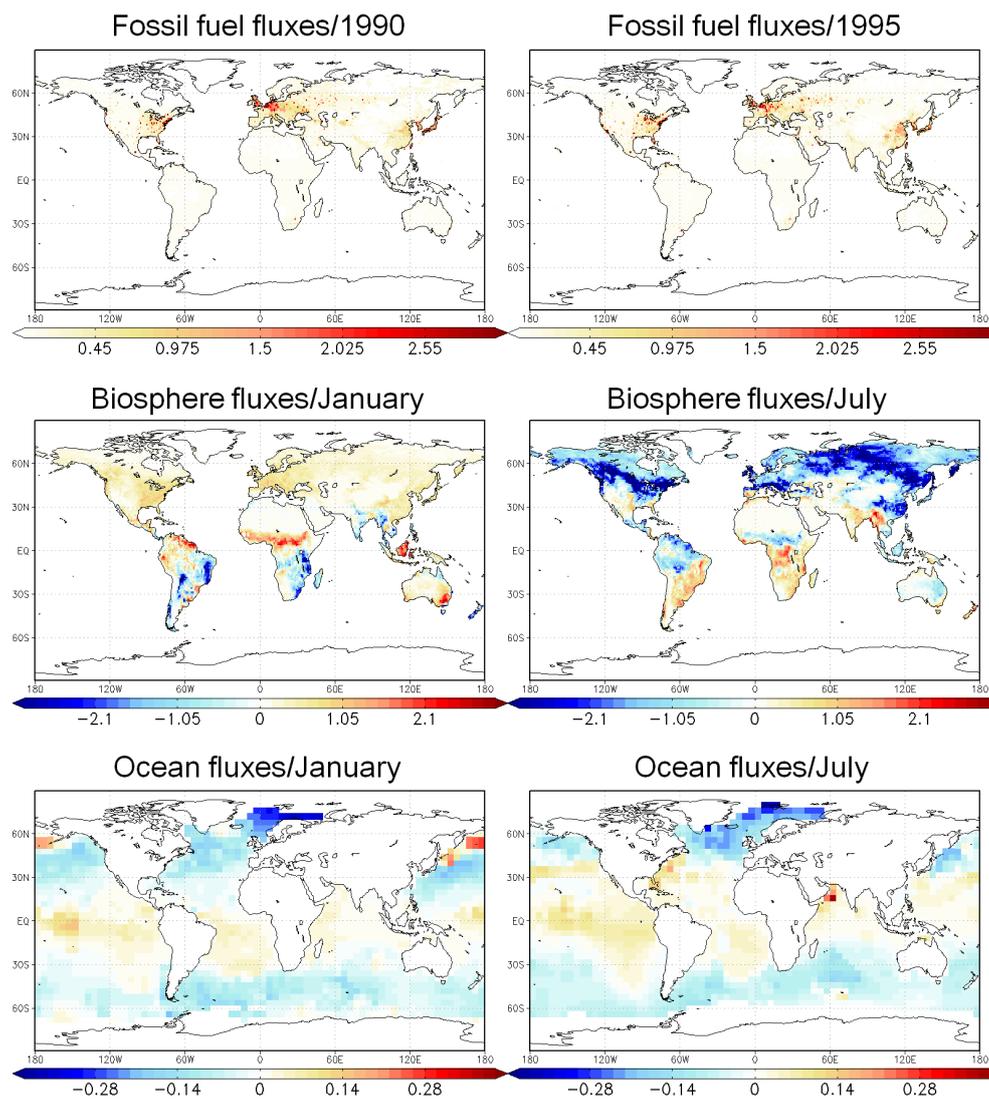
here  $y$  – is a vector of observations of size  $n$ ,  $x_p$  – is an priori estimate for the surface fluxes of size  $m$ ,  $x$  is a vector of unknown parameters (to be estimated) of size  $m$ ,  $\mathbf{H}$  – is a model operator (a matrix of dimensions  $n \times m$ ) transforming vector  $x$  to observations  $y$ ,  $\mathbf{R}$  – is a representativeness error matrix of size  $n \times n$  and  $\mathbf{B}$  – is the background error covariance matrix of size  $m \times m$ .

As mentioned above, the main idea of our approach is to use empirical orthogonal functions as the basis of decomposition for the emission sources. This would hopefully result in a more accurate representation of the spatial distribution of the emission sources while minimizing the computational requirements on the inverse problem solution. Formally for each kind of sources we have:

$$flux_{res} = flux_{pr} + \sum_{i=1}^N \alpha_i EOF_i \quad (2)$$

where  $flux_{res}$  – is a result field of flux in units  $g\ m^{-2}\ day^{-1}$ ,  $flux_{pr}$  – is a pre-set “presubtracted” fields (Fig. 1), used in the experiment T3 in units  $g\ m^{-2}\ day^{-1}$ ,  $EOF_i$  – empirical orthogonal functions,  $\alpha_i$  – are the corresponding basis expansion coefficients, and  $N$  – is the number of the EOFs. The inverse problem task essentially involves to a determination of the corresponding weights for each of the basis function (EOFs).

An advantage of our approach is that we are implicitly using information about correlation between different grid-points that could emit or absorb a trace gas in questions, and do so for each emission type. This information is embedded in the computed spatial distribution of the EOFs. No potentially subjective a priori data (for example regarding correlation lengths in land and ocean fluxes) is being used in the calculations, which may or may not be advantageous. Clearly, if external information of good quality regarding correlations of emission between different grid-ponts is available, it



**Fig. 1.** A fixed “presubtracted” average emission fields in units of  $\text{g m}^{-2} \text{day}^{-1}$  from T3 experiment for three different emission sources: fossil fuel (top figures, year data), biosphere exchange (middle figures, monthly data), ocean exchange (bottom figures, monthly data).

would be advantageous to use it in solving the inverse problem. Yet such data could be difficult to obtain, and would likely not be of similar quality for different regions of the world, thus introducing an additional bias to the solution.

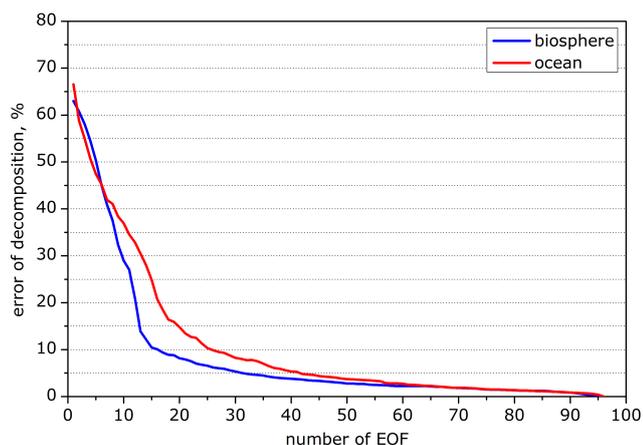
In this investigation we attempt to show that using a rather limited set of parameters (basis coefficients of EOF expansion), it is possible to quantify different surface sources and sinks of  $\text{CO}_2$ , at a fairly high spatial resolution and without the use of the regional source approach.

## 2.1 Determination of EOF

As in the T3 experiment we consider separately three kinds of surface  $\text{CO}_2$  sources: burning fossil fuel, biosphere exchanges and ocean exchanges. Therefore it is necessary to

compute EOFs for each of them, except the fossil fuel emissions (which are fixed in TransCom3 experiment).

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/>). While this data set is already a result of numerical inversion, yet in the absence of a desirable quantity of observations falling back on a variability of a set of well-established inversion results, using it seems to be a reasonable approach. CarbonTracker dataset (release 2009) provides global surface fluxes from 2000 to 2008 at 3-hour time intervals for four kinds of surface emissions: burning fossil fuels, biosphere exchanges, ocean exchanges and fires. Due to the spin-up effect we removed year 2000 from our analysis. The spatial resolution of these fields



**Fig. 2.** Error of decomposition (y-axis) between original emission fields and those restored depends on number of most significant EOFs (x-axis).

is one by one degree; hence the resolution of the EOFs will be the same.

Since the transport model (as is described in Sect. 2.2) has spatial resolution of 2.5 degrees, we should be estimating the fluxes at similar resolution. Therefore, all fields should be interpolated to a 2.5 degree grid.

Since in our investigation we are interested in monthly emissions, (as in experiment T3), we need to average the 3-hour fields to make them comparable with the monthly-mean emissions. Therefore, we get a set of 96 monthly-mean fields (8 years times 12 months) for each emission type. Note, that we are not attempting to optimize the anthropogenic emissions, as they are considered fixed in the T3 experiment setup.

Similarly to T3 experiment, prior to computing the EOFs mean T3 emission fields were subtracted from the CarbonTracker time-dependent monthly emission fields. In our present setup, due to using T3 mean fluxes we likely have bias as part of EOF set. But we cannot predict if we have a negative effect on the following optimization. The 2-D fields of deviations of surface emissions from the mean are available at monthly intervals.

After that, each 2-D field is re-shaped as a vector of size  $144 \times 72 = 10\,368$  ( $2.5 \times 2.5$  degree) and is added as column vector to the matrix representing the surface sources/sinks.

We apply SVD decomposition to the matrix (size 10 368 rows per 96 columns) and obtain the resulting 96 empirical orthogonal functions. We use a standard procedure from [9].

For the next step it is desirable to determine the minimum number of EOFs needed to reasonably accurately represent  $\text{CO}_2$  surface emissions for each source in order to reduce the dimensionality of the inverse problem. In order to do so, we compute relative error of the EOF-restored fluxes with

respect to original fluxes at 2.5 degree resolution as follows:

$$error = \frac{|flux_{orig} - flux_{EOF}|}{|flux_{orig}|} \cdot 100\% \quad (3)$$

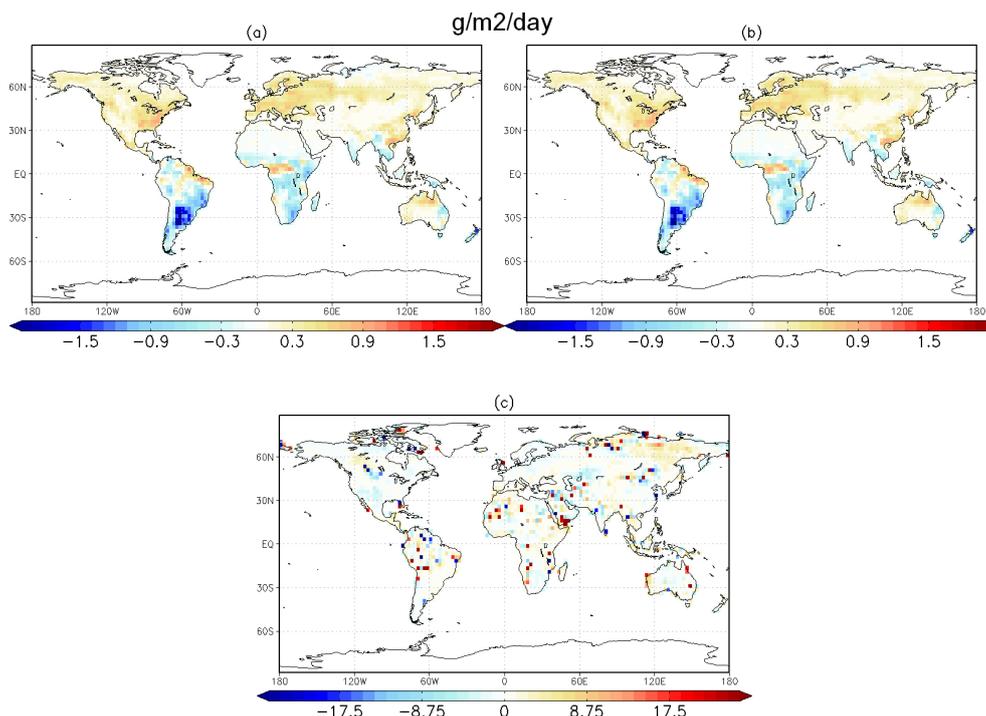
here  $flux_{orig}$  – original emission fields from CarbonTracker,  $flux_{EOF}$  – emission fields reconstructed from the EOFs. Afterwards the obtained error fields are averaged in space and time. Figure 2 shows averaged relative error between monthly averaged emission fields from CarbonTracker project and fields obtained after decomposing on EOFs. It appears that in order to be able to represent emission sources about 5% accuracy one needs to use 30 EOFs for the biospheric sources and 40 EOFs for the ocean sources (more details on the ability of the EOF decomposition to capture spatial and temporal variability of the underlying fields are given in Appendix A). Thus the highest possible dimensionality of the inversion problem for one year is  $(30 + 40) \cdot 12 + 4 = 844$ . 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. Since the EOFs themselves are obtained the re-analysis data, it is reasonable to assume that the number of the EOFs needed to adequately represent the emission fields will be similar to the number of the degrees of freedom used in the CarbonTracker framework. CarbonTracker uses 30 regions for the ocean, which comparable to the number in our simulations. For the biosphere, CarbonTracker employs surface emission regions similar to those utilized in the T3 experiment, yet introduces 19 additional ecosystems, but not in all geographical regions. As a result they have about 125 degrees of freedom in their covariance matrix. It means that CarbonTracker would be “oversolving” system, due to, as we suppose, spatial correlations between some of these regions and ecosystems.

## 2.2 Transport model

The transport model used in this work has been developed at the National Institute for Environmental Studies (Maksyutov et. al., 2008). The model effectively converts  $\text{CO}_2$  surface emissions (a derivable parameter) to the volume concentrations of  $\text{CO}_2$  (an observable parameter).

We use pressure level ECMWF operational analyses at 12-hour time step and 2.5 degrees 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 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 15<sup>th</sup> 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 ( $1 \times 1$  degree) than the model grid ( $2.5 \times 2.5$  degrees), so the



**Fig. 3.** Results of simulation experiment for biosphere. **(a)** pseudotruth, **(b)** recovered fluxes, **(c)** difference in percentage between truth and recovered fluxes.

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.

In the described work, the model was run on a grid of  $2.5 \times 2.5$  degrees, with 15 vertical levels using a semi-Lagrangian transport scheme. As such, we can reasonably constrain surface fluxes only for that model resolution, instead of  $1 \times 1$  degree. In this case the maximum dimensionality of the state vector for the grid-scale inversion would have been  $(144 \cdot 72) \cdot 12 = 124\,416$  (for one year). It is clear that using the given number of observations, the grid-scale inverse problem would have been significantly under-determined. But this problem can be solved by introducing some additional information (e.g. correlation length) and proper first guess. In case of EOF approach as described in Sect. 2.1 for the same spatial resolution we have the dimensionality of the state vector only  $(30 + 40) \cdot 12 = 840$ . That means reduction of the under-determined degree, and in case of a large number of observations per month it could be over-determined.

### 2.3 Inversion method

In our approach all non-observed parameters are estimated simultaneously at each solution step using all available observations. This technique is known as a “batch mode” inversion. In order to minimize the cost function (Eq. 1) we

utilize Kalman filter (KF) (Eqs. (4), (5) and (6)):

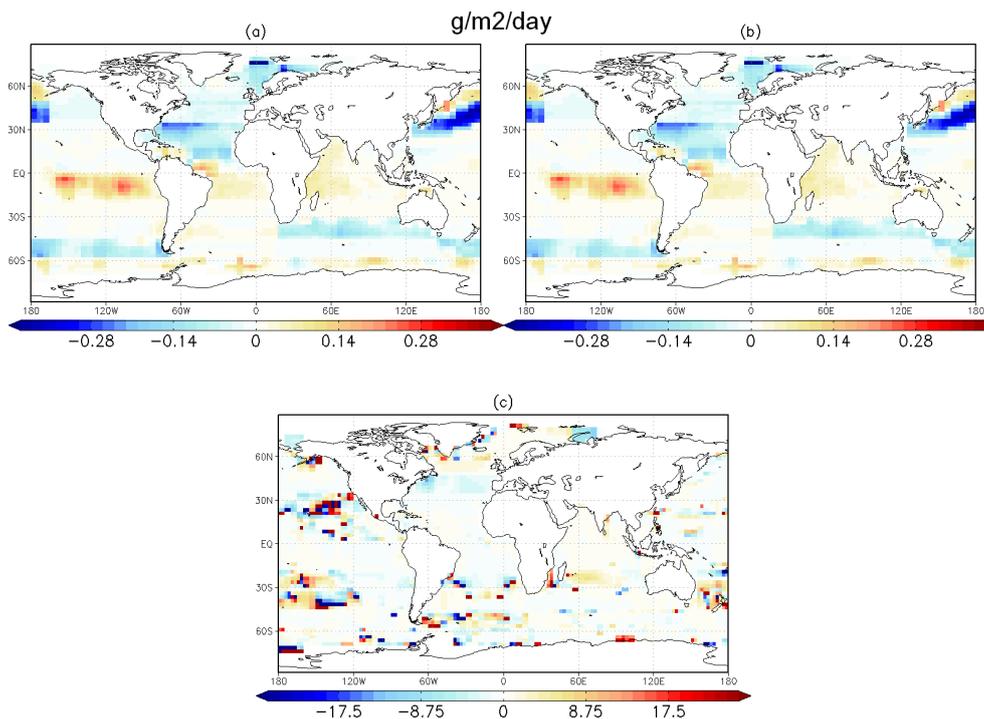
$$\mathbf{x} = \mathbf{x}_p + \mathbf{K}(\mathbf{y} - \mathbf{H}[\mathbf{x}_p]) \quad (4)$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (5)$$

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \quad (6)$$

here  $\mathbf{x}$  is the posteriori vector of EOF expansion coefficients;  $\mathbf{x}_p$  is a priori vector of EOF expansion coefficients;  $\mathbf{y}$  is observation vector;  $\mathbf{H}$  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  $\mathbf{B}$  and  $\mathbf{A}$  are the a priori and posteriori error covariance matrices for state vector.  $\mathbf{I}$  is identity matrix.

KF allow us to estimate coefficients of the EOF expansion and errors of the resulting emissions using ground-based observations of  $\text{CO}_2$  concentrations. This technique is a powerful and widespread method for inverse problem. Kalman filter applications for trace gases have been described in detail in Verlaan and Heemink (1995, 1996), Cohn and Todling (1995), Pham (1998), Zhang et al. (1999), Hanea et al. (2004). We would like to point out that in our case we describe Kalman Filter in conjunction with EOF approach.



**Fig. 4.** Results of simulation experiment for ocean. (a) pseudotruth, (b) recovered fluxes, (c) difference in percentage between truth and recovered fluxes.

## 2.4 Simulation experiment

In order to verify the proper functionality of our inverse system, we performed experiments with simulated data sets.

First, we selected a limited number of EOFs (corresponding to the Sect. 2.1) and reasonable a priori set of EOF expansion coefficients as the state vector which is to be retrieved. Monthly-mean surface sources of CO<sub>2</sub> for 2 types of emissions (biosphere and ocean) obtained from Carbon-Tracker 2001–2008 data have been de-composed using the set of EOF. As a priori values for the expansion coefficients we use monthly averages. We added 30 % Gaussian noise to the simulated data for January for the biosphere sources and 15 % to the ocean sources for the “true” values. The forward model has been run for a period of one year using these sources. Monthly mean simulated data have been computed at the locations of the 75 stations used in the T3 experiment (described in Sect. 3.5) for a total of 900 simulated observations.

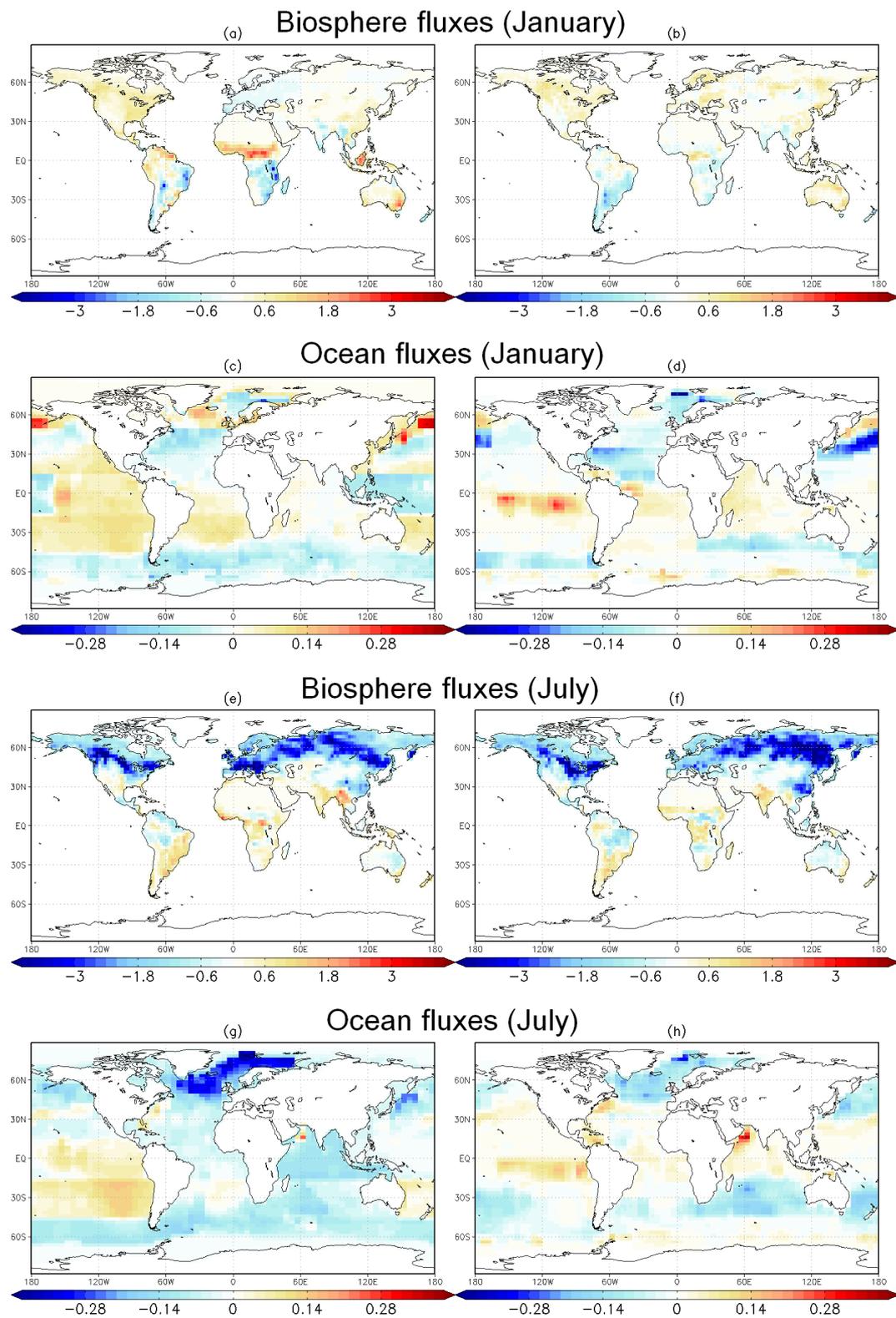
The error of observations was set to be 0.3 % for all stations. An a priori error for the coefficients of the EOF expansion was set to be 30 % for the biosphere and 15 % for the ocean. The inversion for all parameters and time points was performed in one iteration of the Kalman filter.

## 2.5 Comparison with regional inversion (TransCom3 experiment level 2)

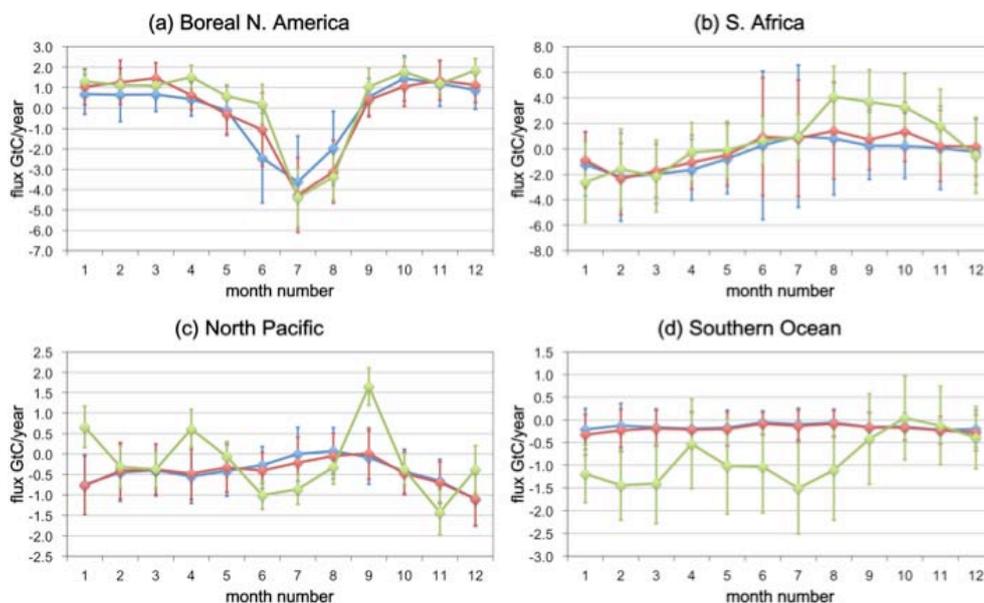
To compare our approach with regional methodology, we repeated the T3 experiment Level 2 (cyclo-stationary problem) as described by Gurney et al. (2004) and obtained monthly sources and sinks of CO<sub>2</sub>. In our project we used National Institute for Environmental Studies Transport Model (NIES TM, see Sect. 2.2). Although T3 experiment uses the averaged observation for 1992–1996 period and the set of pre-computed EOF belong to 2001–2008 period, we made the assumption that since TransCom3 Level 2 is a cyclo-stationary experiment, its results would adequately describe the main degrees of variability (EOFs) that would be approximately independent of temporal differences.

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

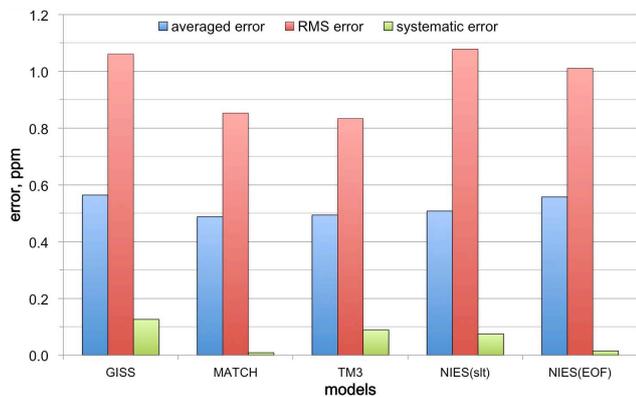
In accordance with T3 Level 2 protocol, we used NIES TM model to compute response functions for each pre-computed EOF representing CO<sub>2</sub> emissions for each month at 3-year intervals. Then all response functions were collected into one model matrix for inversion. Observations of CO<sub>2</sub> concentrations along with the covariance matrix were taken from TransCom3 data set.



**Fig. 5.** Mean annual fluxes for biosphere and ocean in units  $\text{g m}^{-2} \text{day}^{-1}$ . Column 1 represents results for region approach, column 2 represents results for EOF approach.



**Fig. 6.** Mean fluxes for T3 regions with uncertainties. (a) region 1, (b) region 6, (c) region 12, (d) region 20. Blue – prior for EOF, red – estimated with EOF, green – estimated with regions (with NIES TM).



**Fig. 7.** Averaged error, RMS error, and systematic error between observations and model simulations of CO<sub>2</sub> distributions for different transport models, and experiments with NIES model using EOF approach (last case) described in the text.

Since all unknown parameter are estimated in one iteration of our method, all observations (75 stations  $\times$  12 months = 900) represent one observational vector. Thus the size of the observation error covariance matrix is  $900 \times 900$ . Similarly, as in the T3 experiment, we estimate corrections to only biospheric and oceanic fluxes.

Therefore, the state vector for each month consists of 70 EOFs (30 for the biosphere and 40 for the ocean). The final size of the state vector is then  $70 \times 12 + 4 = 844$ , where 4 refers to the number of “presubtracted” fluxes. Thus the size of the background error covariance matrix is  $844 \times 844$ . Prior values for state vector is the same as described in

Sect. 2.4, and as the errors we used the standard deviations for each coefficient.

### 3 Results

First we will present results of the simulation inversion experiment described in Sect. 2.4.

Figures 3 and 4 display results of simulation experiment for the biosphere and ocean emissions. Figures 3c and 4c demonstrate the relative error between the “true” (simulated) emission field and the restored emission field. As one can conclude from examining these plots the error of the restored emission strengths is about 5 %.

Figure 5 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.

We will now describe the results of estimating regional emission sources. After determining spatial distribution of the global sources, we can attempt to estimate surface fluxes using the regions defined in the T3 runs for comparison.

We obtained monthly-mean fluxes in  $\text{Gt yr}^{-1}$  for each of the regions as defined in the T3 protocol.

Figure 6(a–d) shows four regions, two for the land and two for the ocean emissions. Each panel demonstrates a priori fluxes for EOF approach, inversion results using the EOF approach, and inversion results using the regional approach with the same NIES transport model for 22 regions.

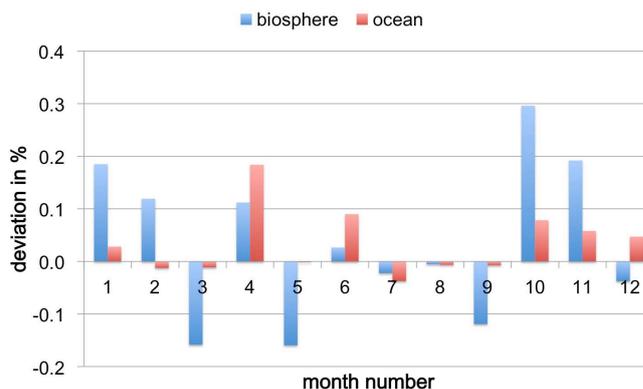
As one can see, the reduction of the uncertainties over the surface is evident from the plots. Note that the dimensionality of the problem is roughly the same or comparable in the case of the EOF and the regional approach by order of magnitude. For the ocean regions, we can observe a slight decrease in the estimation error. In all likelihood, this is related to the relatively low number of over-the-ocean observations. In summary, we conclude that our inversion results reproduce the expected annual cycle.

As another control we have used averaged error, RMS error, and systematic error between observations and model simulations of CO<sub>2</sub> distributions for different transport models, and experiments with NIES model using regional and EOF approaches. These results are presented in figure 7. Particularly interesting are results in the last two columns, as they represent inversion errors for the same model but different inversion approaches. As one can see, our approach indicates a smaller RMS error and a smaller systematic error as compared with the regions approach.

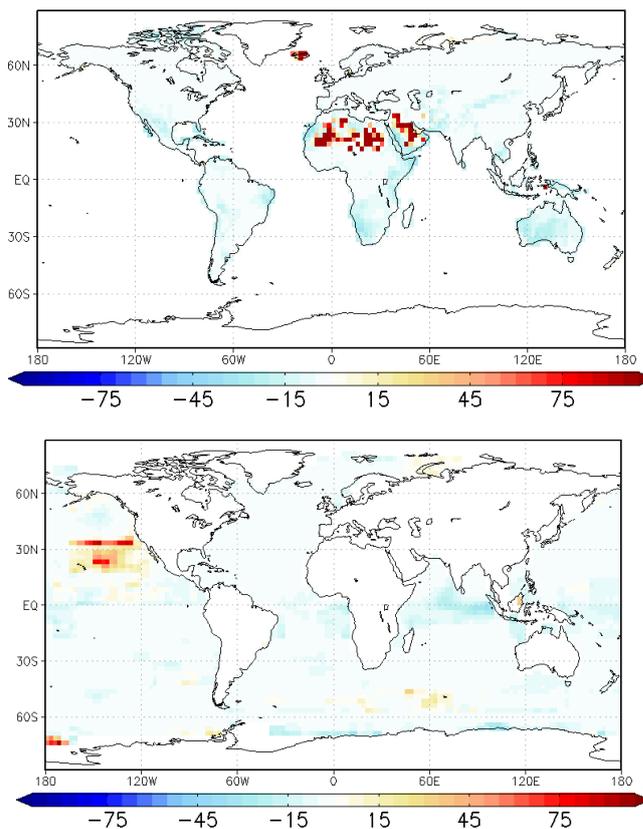
#### 4 Conclusions

We presented an alternative inversion method to the traditional approach that uses discrete geographical spatial regions. We instead relied on a pre-computed set of the empirical orthogonal functions that capture spatial variability of the CO<sub>2</sub> sources and performed the inversion with regard to the EOF expansion coefficients. The procedure of computing such EOFs has been described for the CarbonTracker reanalysis data. We also described determination of the optimal number of the EOFs for more accurate determination of the surface sources and sinks of CO<sub>2</sub>. A simulated observational system experiment has been performed to verify that the proposed inversion system captures variability of the derived parameters, as demonstrated by figures 3–4. We repeated TransCom3 Level 2 experiment. Our system demonstrated successful retrieval of the spatial distribution of the monthly mean sources similar to the regional approach of 22 regions. We also derived reasonable seasonal variability of the sources as compared to the published T3 experiment results.

As next steps in developing our approach, we would like to calculate EOFs using the global biosphere and ocean models separately. That will be the simplification of these models and its parameters. We also consider increasing the number of observations using space-borne observations in addition to the ground-based data, hopefully achieving a significantly more advanced inversion accuracy and precision.



**Fig. A1.** Relative error in percent between RMS error for spatial variability of original fields and RMS error for spatial variability of fields restored with EOFs for two kinds of sources depends on month.



**Fig. A2.** Relative error in percent between RMS error for temporal variability of original fields and RMS error for temporal variability of fields restored with EOFs for two kinds of sources. (top – biosphere, bottom – ocean).

## Appendix A

### Investigation of the ability of EOFs to capture time and space variability of surface sources of CO<sub>2</sub>

Two numerical experiments have been performed in order to assess how well EOF decomposition captures spatial and temporal variability of the emission fields:

#### Experiment 1: Spatial variability

12 monthly mean original CarbonTracker emission fields for the biosphere and oceans have been used to compute average root-mean-squared deviations for each month  $RMS_{orig}$ . Similar procedure has been performed using fields obtained from EOF expansion using 30 EOFs for biosphere and 40 EOFs for the oceans ( $RMS_{EOF}$ ). A relative error between the two has been computed as  $(RMS_{orig} - RMS_{EOF}) \cdot 100\% / RMS_{orig}$ . Results are shown in Fig. A1. As one can see, EOF decomposed fields show degree of spatial variability similar to that of the original fields.

#### Experiment 2: Temporal variability

Similar to Experiment 1 procedure has been followed in Experiment 2, except that instead of computing spatial root-mean-squared deviations for each month, we computed temporal root-mean-squared deviations at each grid point. Thus we obtained two 2-D fields of RMS deviations and the computed relative difference between them. Results are shown in Fig. A2, which demonstrates that temporal variability computed from original fields and that computed from EOF-restored fields are generally in agreement with the notable exception of Sahara region for the biosphere. This is possibly due to missing data in that region in the CarbonTracker data.

*Acknowledgements.* This study has been supported by GOSAT project at NIES, Japan. CarbonTracker 2008 results provided by NOAA ESRL, Boulder, Colorado, USA from the website at <http://carbontracker.noaa.gov>. TransCom project data was made available due to support from the National Science Foundation (OCE-9900310), the National Oceanic and Atmospheric Administration (NA67RJ0152, Amend 30) and the International Geosphere Biosphere Program/Global Analysis, Interpretation, and Modeling Project.

Edited by: M. Heimann

## References

- Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M.: The ECMWF implementation of three dimensional variational assimilation (3D-Var). I: Formulation, Q. J. Roy. Meteorol. Soc., 124, 1783–1807, 1998.
- Desbiens, R., Aoki, T., and Yokota, T.: Optimization of GOSAT Atmospheric Retrieval of CO<sub>2</sub> in Presence of Atmospheric Particles Using Empirical Orthogonal Function Representation, EOS transactions, American Geophysical Union, Fall Meeting 2007, abstract #A13D-1506, 2007.
- ECMWF, ECMWF/WCRP Level III-A Global Atmospheric Data Archive, ECMWF available online at: <http://www.ecmwf.int>, 1999.
- 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.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Pac, B. C., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fung, I., Heimann, M., John, J., Maki, T., Maksyutov, S., Peylin, P., Prather, M., and Taguchi, S.: Transcom 3 inversion intercomparison: Model mean results for the estimation of seasonal carbon sources and sinks”, Global Biogeochem. Cycles, 18, GB1010, <http://dx.doi.org/10.1029/2003GB002111>doi:10.1029/2003GB002111, 2004.
- 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. on Math. Modeling and Computation 4, 247 pp., 1998.
- Kalman, R. E.: A new approach to linear filtering and prediction problems. J. Basic Eng., 82, 35–45, 1960.
- Kaminski, T., Heimann, M., and Giering, R.: A coarse grid three-dimensional global inverse model of atmospheric transport: 2. Inversion of the transport of CO<sub>2</sub> in 1980s’, J. Geophys. Res., 104, 18555–18582, 1999.
- Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions.” J. Geophys. Res., 106, 4703–4715, 2001.
- 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 Simulator 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, <http://dx.doi.org/10.1029/2003JD004422>doi:10.1029/2003JD004422, 2004.
- Patra, P. K., Ishizawa, M., Maksyutov, S., Nakazawa, T., and Inoue, G.: Role of biomass burning and climate anomalies on land-atmosphere carbon fluxes based on inverse modelling of atmospheric CO<sub>2</sub>, Global Biogeochem. Cy., 19, GB3005, <http://dx.doi.org/10.1029/2004GB002258>doi:10.1029/2004GB002258, 2005a.
- Patra, P. K., Maksyutov, S., Ishizawa, M., Nakazawa, T., Ukita, J., and Takahashi, T.: Interannual and decadal changes in the sea-air CO<sub>2</sub> flux from atmospheric CO<sub>2</sub> inverse modelling, Global Biogeochem. Cy., 19, GB4013, <http://dx.doi.org/10.1029/2004GB002257>doi:10.1029/2004GB002257, 2005b.
- Patra, P. K., Mikaloff-Fletcher, S. E., Ishijima, K., Maksyutov, S., and Nakazawa, T.: Comparison of CO<sub>2</sub> fluxes estimated using atmospheric and oceanic inversions, and role of fluxes and their interannual variability in simulating atmospheric CO<sub>2</sub> concentrations, Atmos. Chem. Phys. Discuss., 6, 6801–6823, <http://dx.doi.org/10.5194/acpd-6-6801-2006>doi:10.5194/acpd-6-6801-2006, 2006.

- Peters et al., "An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker", PNAS, November 27, 2007, vol. 104, no. 48, 18925-18930, 2007.
- Press, W., Teukolsky, S., Vetterling, W., and Flannery, B.: Numerical Recipes in FORTRAN. The Art of Scientific Computing, Second Edition, 1995.
- Rodenbeck, C., Houweling S., Gloor, M., and Heimann, M.: CO<sub>2</sub> flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport.", Atmos. Chem. Phys., 3, 1919–1964, <http://dx.doi.org/10.5194/acp-3-1919-2003>doi:10.5194/acp-3-1919-2003, 2003.
- Wikle, C. K. and Cressie N.: A dimension reduced approach to space-time Kalman filtering, Biometrika, 86, 815–829, 1999.