Atmos. Chem. Phys., 8, 209–250, 2008
www.atmos-chem-phys.net/8/209/2008/

Atmospheric Chemistry and Physics

Retrieving global aerosol sources from satellites using inverse modeling

O. Dubovik[1,2], T. Lapyonok[3,4], Y. J. Kaufman[5], M. Chin[5], P. Ginoux[6], R. A. Kahn[5,7], and A. Sinyuk[3,4]

[1]Laboratoire de Optique Atmosphérique, Université de Lille 1/CNRS, Villeneuve d 'Ascq, France
[2]Substantial part of this study was done while worked at: Laboratory for Terrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[3]Laboratory for Terrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[4]Science Systems and Applications, Inc., Lanham, MD, USA
[5]Laboratory for Atmospheres, NASA Goddard Space Flight Center, Greenbelt, MD, USA
[6]Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, NJ, USA
[7]Geophysical Jet Propulsion Laboratory, Pasadena, CA, USA

Received: 22 November 2006 – Published in Atmos. Chem. Phys. Discuss.: 12 March 2007
Revised: 4 October 2007 – Accepted: 4 December 2007 – Published: 18 January 2008

Abstract. Understanding aerosol effects on global climate requires knowing the global distribution of tropospheric aerosols. By accounting for aerosol sources, transports, and removal processes, chemical transport models simulate the global aerosol distribution using archived meteorological fields. We develop an algorithm for retrieving global aerosol sources from satellite observations of aerosol distribution by inverting the GOCART aerosol transport model.

The inversion is based on a generalized, multi-term least-squares-type fitting, allowing flexible selection and refinement of a priori algorithm constraints. For example, limitations can be placed on retrieved quantity partial derivatives, to constrain global aerosol emission space and time variability in the results. Similarities and differences between commonly used inverse modeling and remote sensing techniques are analyzed. To retain the high space and time resolution of long-period, global observational records, the algorithm is expressed using adjoint operators.

Successful global aerosol emission retrievals at $2^\circ \times 2.5^\circ$ resolution were obtained by inverting GOCART aerosol transport model output, assuming constant emissions over the diurnal cycle, and neglecting aerosol compositional differences. In addition, fine and coarse mode aerosol emission sources were inverted separately from MODIS fine and coarse mode aerosol optical thickness data, respectively. These assumptions are justified, based on observational coverage and accuracy limitations, producing valuable aerosol source locations and emission strengths. From two weeks of daily MODIS observations during August 2000, the global placement of fine mode aerosol sources agreed with available independent knowledge, even though the inverse method did not use any a priori information about aerosol sources, and was initialized with a "zero aerosol emission" assumption. Retrieving coarse mode aerosol emissions was less successful, mainly because MODIS aerosol data over highly reflecting desert dust sources is lacking.

The broader implications of applying our approach are also discussed.

Correspondence to: O. Dubovik
(dubovik@loa.univ-lille1.fr)

1 Introduction

Knowledge of the global distribution of tropospheric aerosols is important for studying the effects of aerosols on global climate. Satellite remote sensing is the most promising way to collect information about global aerosol distributions (King et al., 1999; Kaufman et al., 2002). However, in spite of recent advances in space technology, the satellite data do not yet provide the required accuracy nor the level of detail needed to assess aerosol property time and space variability. Tropospheric aerosols may display strong local variations, and any single satellite instrument needs at least several days of observations to obtain sufficient cloud-free images for global coverage. Also, most satellite aerosol data records are limited to daytime, clear-sky conditions. Comprehensive, global atmospheric aerosol simulations having adequate time and space resolution can be obtained using global models that rely on estimated emissions and account for aerosol transport and removal processes.

At present, there are a number of well-established Global Circulation Models (GCMs) that generate their own meteorology (e.g. models by Roechner et al., 1996; Tegen et al.,

1997, 2000; Koch et al., 1999; Koch, 2001; Ghan et al., 2001a, b; Reddy and Boucher, 2004) and Chemical Transport Models (CTMs) that incorporate meteorological data from external sources into the model physics (e.g. models by Balkanski et al., 1993; Chin et al., 2000, 2002; Ginoux et al., 2001; Takamura et al., 2000, 2002). However, the accuracy of global aerosol models is limited by uncertainties in aerosol emission source characteristics, knowledge of atmospheric processes, and the meteorological field data used. As a result, even the most recent models are mainly expected to capture only the principal global features of aerosol transport; among different models, quantitative estimates of average regional aerosol properties often disagree by amounts exceeding the uncertainty of remote sensing aerosol observations (e.g. Kinne et al., 2003, 2006; Sato et al., 2003). Therefore, there are diverse, continuing efforts to harmonize and improve global aerosol modeling by refining the meteorology, atmospheric process representations, emissions, and other modeling components used.

The availability of aerosol remote sensing products, especially global aerosol fields provided by satellite observations, is of critical importance for verifying and constraining aerosol models. For example, the direct comparisons of model outputs with observed aerosol properties are used for evaluating model accuracy and for identifying possible modeling problems (e.g. Takamura et al., 2000; Chin et al., 2002, 2003, 2004; Kinne et al., 2003, 2006). The observations can also be used to optimize the agreement between tracer transport model predictions and observation. For example, model predictions can be adjusted and enhanced by assimilating observations into the model. Collins et al. (2000, 2001) improved regional aerosol model predictions by assimilating the available satellite retrievals of aerosol optical thickness. Weaver et al. (2006) suggested a procedure for assimilating satellite-level radiances into a radiative transfer model driven by GOCART global transport model aerosol field predictions. Another way of improving global aerosol modeling is retrieving (or adjusting) aerosol emissions from available observations by inverting a global model. This approach is particularly promising because aerosol emission uncertainty is widely recognized as a major factor limiting global aerosol model accuracy. It has been shown that inversion techniques are rather effective at improving the accuracy of trace gas chemical models (e.g. Kaminski et al., 1999b; Khattatov et al., 2000; Kasibhatla et al., 2000; Elbern et al., 1997; Para et al., 2003).

However, implementing the same techniques for inverting aerosol models appears to be more challenging. Indeed, a description of the aerosol field generally requires a larger number of parameters compared to a description of atmospheric gases, partly because of relatively high aerosol temporal and spatial variability (see discussion in Sect. 2.5). In addition, direct implementation of basic inversion methods (that use the Jacobi matrices of first derivatives) is computationally demanding and, therefore, hardly applicable in aerosol global modeling. In these regards, designing an inversion in a variational formalism framework, using adjoint operators, is rather promising. The adjoint operators (Marchuk, 1977, 1986; Cacuci, 1981; Tarantolla, 1987) allow direct calculation of the gradients of the quadratic form in respect to model input parameters, without explicit use of Jacobi matrices. Such calculations have computational requirements similar to those of forward modeling. Correspondingly, using adjoint operators allows efficient implementation of the model inversion by minimizing quadratic form (quantifying mismatch between observations and modeling results) with the gradient methods, provided the methods converge rapidly enough.

Adjoint techniques are widely used in meteorology and oceanography for variational data assimilation (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Courtier and Talagrand, 1987; Navon, 1997, etc.), and have been successfully applied to inverse modeling analyses involving atmospheric gases (Kaminski et al., 1999a; Elbern et al., 2000; Menut et al., 2000; Vukicevic and Hess, 2000; Vautard et al., 2000; Elbern and Schmidt, 2001; Schmidt and Martin, 2003; Menut, 2003; Elbern et al., 2007). Hakami et al. (2005) used an adjoint approach to retrieve regional sources of black carbon from aircraft, shipboard, and surface black carbon measurements collected during the ACE-Asia field campaign.

Our paper explores the possibility of deriving the global distribution and strength of aerosol emission sources from satellite observations, using the adjoint operator formulation to invert an aerosol transport model. Figure 1 illustrates the general retrieval concept. In addition, we analyze possible parallels and analogies between inverse modeling and retrieval approaches widely used in atmospheric remote sensing. Such analyses may be useful sources of efficient methods developed in remote sensing, that could be adapted to inverse modeling. For example, numerous remote sensing applications use the Phillips-Tikhonov-Twomey inversion technique developed in the early sixties by Phillips (1962), Tikhonov (1963) and Twomey (1963). The technique suggests constraining ill-posed problems using a priori limitations on the derivatives of the retrieved function. Here, we discuss the possibility of constraining temporal and/or spatial aerosol variability by applying a priori limitations on aerosol mass derivatives with respect to time and space coordinates. Also, we formulate the inversion problem using a multi-term least squares approach, convenient for including multiple a priori constraints in the retrieval (Dubovik, 2004).

We applied our approach to retrieving global aerosol sources by inverting the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model. Algorithm performance is illustrated by numerical tests, as well as by deriving global aerosol emissions, applying the algorithm to actual MODIS aerosol observations. The algorithms potential and limitations are also discussed more generally.

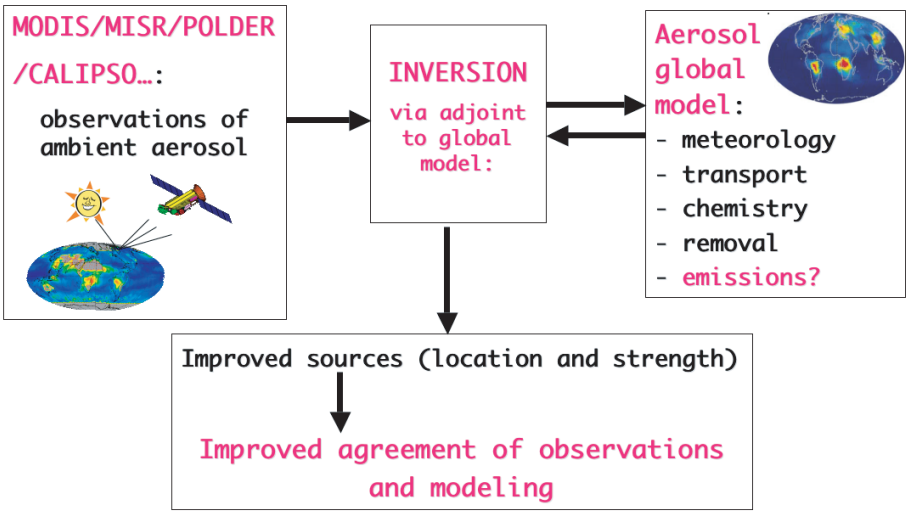

Fig. 1. Flowchart of the retrieval scheme concept.

2 The methodology of inverse modeling

The spatial and temporal behavior of atmospheric constituents is simulated in chemistry models by solving the continuity equation (Brasseur et al., 1999; Jacob, 1999):

$$\frac{\partial m}{\partial t} = -\boldsymbol{v}\,\nabla m + \left(\frac{\partial m}{\partial t}\right)_{\text{diff}} + \left(\frac{\partial m}{\partial t}\right)_{\text{conv}} + S - R, \qquad (1)$$

where $\boldsymbol{v}$ is the transport velocity vector, m is mass and suffixes "diff" and "conv" denote turbulent diffusivity and convection, respectively. S and R denote source and loss terms, respectively. The characteristics m, $\boldsymbol{v}$, S and R in Eq. (1) are explicit functions of time t and spatial coordinates $\boldsymbol{x}$=(x, y, z). The continuity equation does not yield a general analytical solution and is usually solved numerically using discrete analogues. Each component process in the numerical equivalent of Eq. (1) is isolated and treated sequentially at each time step Δt (e.g. Jacob, 1999):

$$m\,(t+\Delta t, \boldsymbol{x}) = T\,(t, \boldsymbol{x})\;\;(m\,(t, \boldsymbol{x}) + s\,(t, \boldsymbol{x}))\,\Delta t, \qquad (2)$$

where $s(t, \boldsymbol{x})$ – mass emission, $T(t, \boldsymbol{x})$ is transport operator, that can be approximated as:

$$T\,(t, \boldsymbol{x}) = T_q T_{q-1} \ldots T_3 T_2 T_1 \quad , \qquad (3)$$

and T_i *(i=1,…,q)* are operators for isolated transport processes such as advection, diffusion, convection and wet scavenging. Thus, the calculation of mass at any given time can be reduced to the numerical integration of known transport and source functions:

$$m\,(t, \boldsymbol{x}) = \int_{t_0}^{t} T\,\left(t', \boldsymbol{x}\right)\;\left(m\,\left(t', \boldsymbol{x}\right) + s\,\left(t', \boldsymbol{x}\right)\right)\,dt'. \qquad (4)$$

If the transport operator $T(t, \boldsymbol{x})$ is linear, Eq. (4) can be equivalently written in terms of the matrix equation. For example, Fig. 2 illustrates one of many possible approaches to representing the global mass distribution by a vector $\boldsymbol{M}$. Using the same approach for representing the global emission distribution by a vector $\boldsymbol{S}$, the matrix equivalent of Eq. (4) can be written as (see explicit derivation in Appendix B):

$$\boldsymbol{M} = \mathbf{T}\boldsymbol{S} + \mathbf{T}_0\boldsymbol{M}_0, \qquad (5)$$

where $\boldsymbol{M}_0$ is a vector of mass values at all locations at time t_0; $\boldsymbol{M}$ and $\boldsymbol{S}$ are the corresponding vectors of mass and emission values at all locations and considered times t_0, t_1,…, t_{n-1},t_n (i.e. these vectors represent the 4-dimensional (4D) aerosol mass and emission variability); $\mathbf{T}$ is the coefficient matrix defining the mass transport to each location $\boldsymbol{x}$ and time step t_k from all locations $\boldsymbol{x}$ and previous time steps $t_{i<n}$. $\mathbf{T}_0$ is the coefficient matrix defining the transport of mass to each location $\boldsymbol{x}$ and time step t_k from mass present at all locations $\boldsymbol{x}$ and at time step t_0. Figure 3 illustrates the relation between integral Eq. (4) and the representation of aerosol transport modeling in vector-matrix form by Eq. (5). Thus, the source vector can be retrieved by solving the matrix equation if the mass measurements $\boldsymbol{M}^{\text{meas}}$=$\boldsymbol{M}$+$\boldsymbol{\Delta}_M$ are available.

Using Eq. (5), the inversion of aerosol transport can be implemented numerically as the solution of a system of algebraic equations. However, vectors $\boldsymbol{M}$, $\boldsymbol{S}$ and matrix $\mathbf{T}$ can have extremely large dimensions (see discussion in Sect. 2.5), and direct implementation of some matrix operations can be difficult. Therefore, inverting the transport equation is commonly formulated in a calculus of variations framework, a field of mathematics that deals with functions of functions. In this formalism, emission estimation is achived using 4D-variational (4D-var) data assimilation techniques (e.g. Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Courtier and Talagrand, 1987; Elbern et al.,

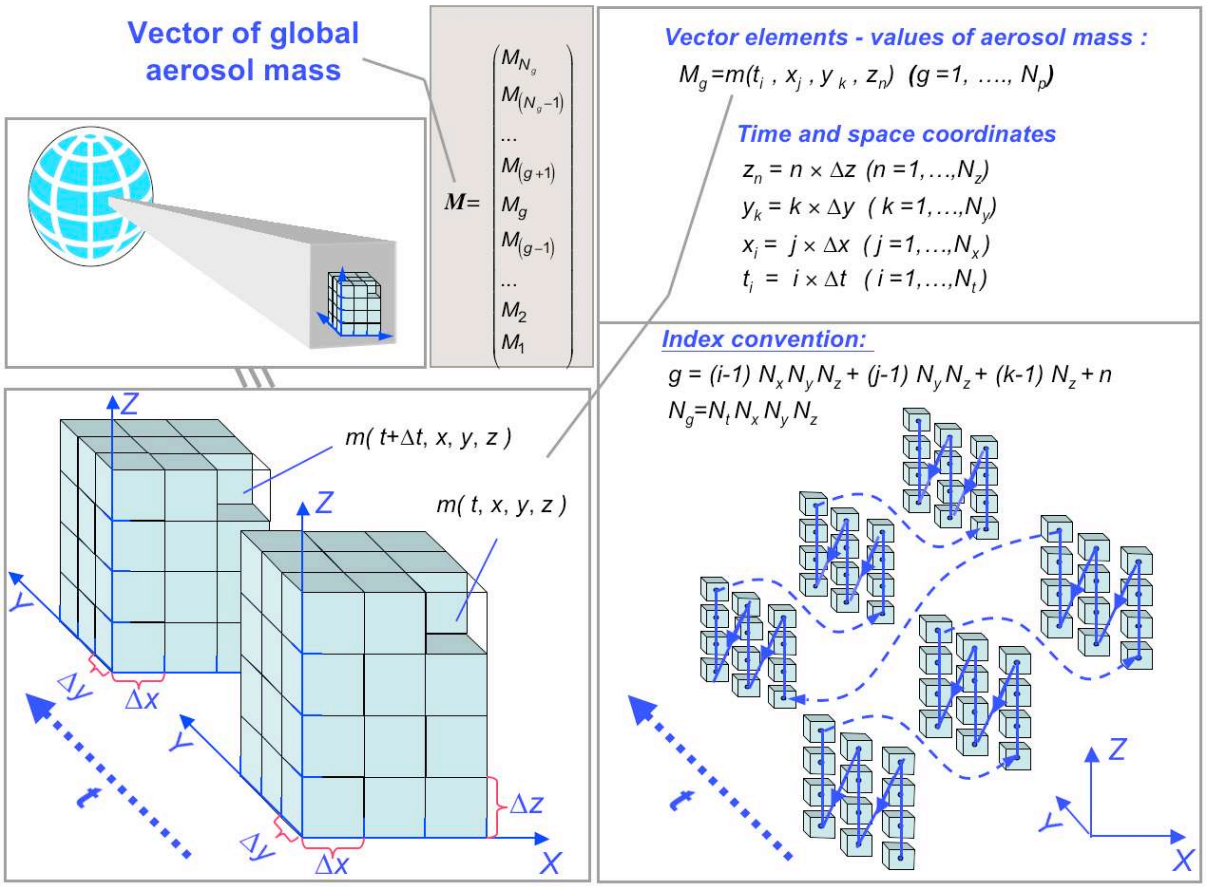

Fig. 2. Illustration representing the aerosol global mass distribution in vector form.

2000; Vukicevic and Hess, 2000; Elbern and Schmidt, 2001; Schmidt and Martin, 2003; Menut, 2003; etc.). Nevertheless, here we invert the transport equation using the matrix formulation given by Eq. (5), to retain full analogy with remote sensing inversion approaches. We hope this analogy will highlight the parallels between these two research areas, and will make it easier to identify positive developments in remote sensing that can be applied to inverse modeling algorithms. Thus, below we discuss the inversion of the transport equation as a formal linear system inversion problem, shown in Eq. (5).

2.1 Statistical optimization of the linear inversion

If the statistical behavior of the errors $\boldsymbol{\Delta}_M$ is known, one can use this knowledge to optimize the solution of Eq. (5). In that way, the solution $\hat{\boldsymbol{S}}$ should not only closely reproduce observations $\boldsymbol{M}^{\text{meas}}$, but in addition, the remaining deviations $\hat{\boldsymbol{\Delta}}_M = \boldsymbol{M}^{\text{meas}} - \boldsymbol{M}(\hat{\boldsymbol{S}})$ should have a distribution close to the expected error properties described by the Probability Density Distribution (PDF) of errors $P(\boldsymbol{\Delta}_M)$. According to the well-known Method of Maximum Likelihood (MML), the optimum solution $\hat{\boldsymbol{S}}$ corresponds to a maximum of the PDF as follows (e.g. Edie et al., 1971):

$$P(\boldsymbol{\Delta}_M) = P(\boldsymbol{M}^{\text{meas}} - \boldsymbol{M}(\boldsymbol{S})) = P(\boldsymbol{M}(\boldsymbol{S}) | \boldsymbol{M}^{\text{meas}}) = \max . \quad (6)$$

Where the PDF $P(\boldsymbol{M}(\boldsymbol{S}) | \boldsymbol{M}^{\text{meas}})$, written as a function of retrieval parameters $\boldsymbol{S}$ for a given set of available observations $\boldsymbol{M}^{\text{meas}}$, is known as a Likelihood Function. The MML is a fundamental principle of statistical estimation that provides a statistically optimum solution in many senses. For example, the asymptotic error distribution (infinite number of Δ_M realizations) of MML estimates has the smallest possible variance. Most statistical properties of the MML solution remain optimal for a limited number of observations (e.g. see Edie et al., 1971). The normal (or Gaussian) distribution is widely considered as the best model for describing actual error distribution (Tarantola, 1987; Edie et al., 1971; etc.):

$$P\left(\boldsymbol{M}(\boldsymbol{S}) | \boldsymbol{M}^{\text{meas}}\right) = \left((2\pi)^{\text{m}} \det(\mathbf{C}_M)\right)^{-1/2}$$
$$\exp\left(-\frac{1}{2}\left(\boldsymbol{M}(\boldsymbol{S}) - \boldsymbol{M}^{\text{meas}}\right)^{\text{T}} \mathbf{C}_M^{-1} \left(\boldsymbol{M}(\boldsymbol{S}) - \boldsymbol{M}^{\text{meas}}\right)\right), \quad (7)$$

where $(\ldots)^{\text{T}}$ denotes matrix transposition, $\mathbf{C}_M$ is the covariance matrix of $\boldsymbol{\Delta}_M$, det($\mathbf{C}$) denotes the determinant of $\mathbf{C}_M$,

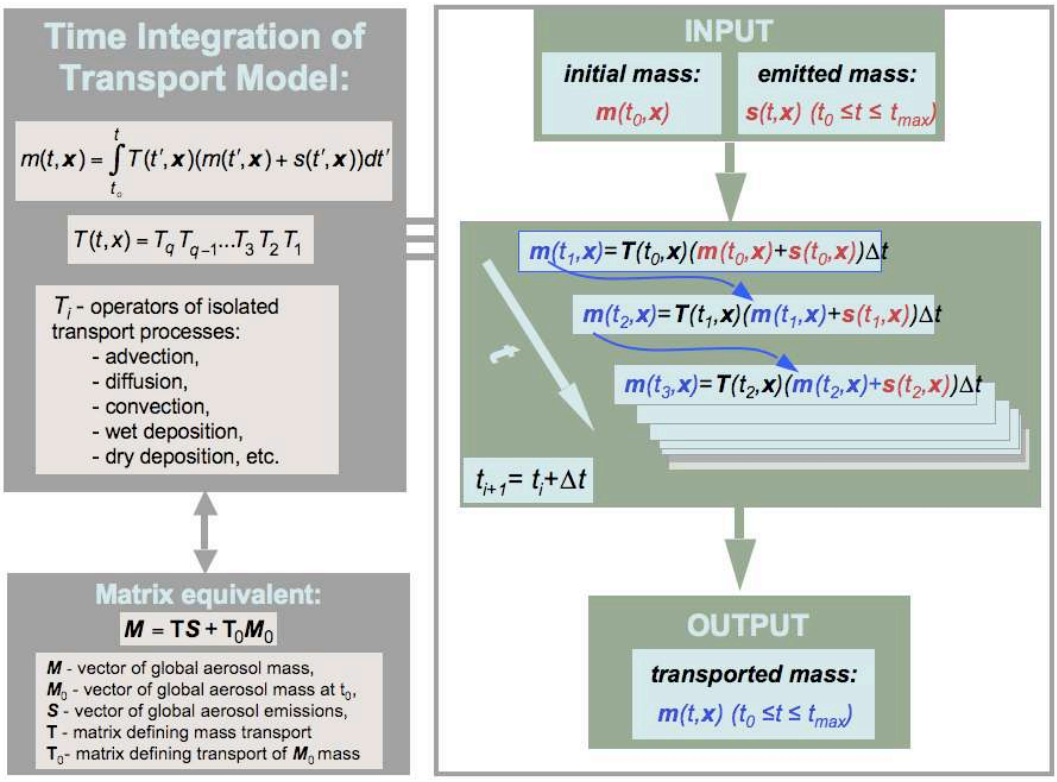

Fig. 3. Illustration of the relation between sequential time integration of aerosol transport and the representation of aerosol transport modeling in vector-matrix form.

and m is the dimension of the vectors $\boldsymbol{M}(\boldsymbol{S})$ and $\boldsymbol{M}^{\text{meas}}$. The maximum of the PDF exponential term in Eq. (7) corresponds to the minimum of the quadratic form in the exponent. Therefore, the MML solution is a vector $\hat{\boldsymbol{S}}$ corresponding to the minimum of the following quadratic form:

$$\Psi(\boldsymbol{S}) = \frac{1}{2}\left(\boldsymbol{M}(\boldsymbol{S}) - \boldsymbol{M}^{\text{meas}}\right)^{\mathrm{T}} \mathbf{C}_M^{-1} \left(\boldsymbol{M}(\boldsymbol{S}) - \boldsymbol{M}^{\text{meas}}\right) = \min. \quad (8)$$

Thus, with the assumption of normal noise, the MML principle requires searching for a minimum in the product of the squared terms of $(\boldsymbol{M}^{\text{meas}} - \boldsymbol{M}(\boldsymbol{S}))$ in Eq. (7). This is the basis for the widely known Least Square Method (LSM).

For linear $\boldsymbol{M}(\boldsymbol{S})$ (as in Eq. 5), the LSM solution can be written as (e.g. Rao 1965):

$$\hat{\boldsymbol{S}} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\mathbf{T}\right)^{-1} \mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\ \boldsymbol{M}^{*}. \quad (9)$$

Here, $\boldsymbol{M}^*$ is the vector of mass measurements corrected for the backdround aerosol $\boldsymbol{M}_0$ present in the atmosphere at t_0 (i.e. prior observations): $\boldsymbol{M}^*=\boldsymbol{M}^{\text{meas}}-\mathbf{T}\boldsymbol{M}_0$.

2.2 Inversion constrained by a priori estimates of unknowns

If the problem is ill-posed and Eq. (5) does not have a unique solution, then some a priori constraints need to be applied. The expected distribution of sources is commonly used as an a priori constraint in inverse modeling. In that case the inversion can be considered as a joint solution of Eq. (5) and constraining a priori the system:

$$\begin{cases} \boldsymbol{M}^{\text{meas}} = \boldsymbol{M}(\boldsymbol{S}) + \boldsymbol{\Delta}_M \\ \boldsymbol{S}^{*} = \boldsymbol{S} + \boldsymbol{\Delta}_S \end{cases}, \quad (10)$$

where $\boldsymbol{S}^*=\boldsymbol{S}+\boldsymbol{\Delta}_S$ is the vector of a priori estimates of the sources and $\boldsymbol{\Delta}_S$ is vector of the errors that usually considered statistically independent of $\boldsymbol{\Delta}_M$ and normally distributed with zero mean and covariance matrix $\mathbf{C}_S$. To solve Eq. (10), MML should be applied to the joint PDF of the measurements and a priori estimates:

$$P(\boldsymbol{M}(\boldsymbol{S})|\boldsymbol{M}^{\text{meas}}, \boldsymbol{S}^{*}) = P(\boldsymbol{M}(\boldsymbol{S})|\boldsymbol{M}^{\text{meas}})P(\boldsymbol{S}|\boldsymbol{S}^{*}) = \max, \quad (11a)$$

i.e.

$$P(\boldsymbol{M}(\boldsymbol{S})|\boldsymbol{M}^{\text{meas}}, \boldsymbol{S}^{*}) = \sim \exp\left(-\frac{1}{2}\left(\Delta\boldsymbol{M}^{\mathrm{T}}\mathbf{C}_m^{-1}\Delta\boldsymbol{M}\right)\right)$$
$$\exp\left(-\frac{1}{2}\left(\Delta\boldsymbol{S}^{\mathrm{T}}\mathbf{C}_S^{-1}\Delta\boldsymbol{S}\right)\right) = \max, \quad (11b)$$

where $\Delta\boldsymbol{M}=\boldsymbol{M}(\boldsymbol{S})-\boldsymbol{M}^*$ and $\Delta\boldsymbol{S}=\boldsymbol{S}-\boldsymbol{S}^*$.

Accordingly, the MML solution of joint Eq. (11) corresponds to a minimum of the following quadratic form:

$$2\,\Psi(\boldsymbol{S}) = 2\,(\Psi_m + \Psi_S) = \Delta\boldsymbol{M}^{\mathrm{T}}\mathbf{C}_m^{-1}\Delta\boldsymbol{M} + \Delta\boldsymbol{S}^{\mathrm{T}}\mathbf{C}_s^{-1}\Delta\boldsymbol{S}. \quad (12)$$

Thus, unlike with Eq. (8), including a priori constraints requires simultaneously minimizing both the measurement term $2\Psi_m$ and the a priori $2\Psi_S$ term. Defining the solution as a minimization of the above two-terms quadratic form is probably the most popular approach for implementing constrained inversions, particularly in geophysical inverse modeling applications. Indeed, the formulations of Eqs. (11) and (12) are practically equivalent to the basic formulations used in the Bayesian approach (e.g. Tarantolla 1997) widely used in inverse modeling (e.g. Rodenbeck et al., 2003; Michalak et al., 2004). In the Bayesian approach, the PDF of the measurements and a priori estimates $P(\boldsymbol{S}|\boldsymbol{S}^*)$ is defined as the prior PDF of the state $\boldsymbol{S}$, and $P(\boldsymbol{M}(\boldsymbol{S})|\boldsymbol{M}^{\text{meas}}, \boldsymbol{S}^*)$ is defined as the posterior PDF of the state $\boldsymbol{S}$. Therefore, the Bayesian definition directly assumes a priori properties of the unknown vector $\boldsymbol{S}$. In a contrast, Eq.(10) treats the a priori estimates $\boldsymbol{S}^*$ as simply a kind of "measurements" of unknowns $\boldsymbol{S}$. Technically, this is equivalent to the Bayesian approach, but it allows more flexibility in formulating a constrained inversion: for example, it can easily be extended to use simultaneously multiple constraints in the inversion (see discussion in Sect. 2.4).

The solution minimizing Eq. (12) can be found using the following equations:

$$\hat{\boldsymbol{S}} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\mathbf{T} + \mathbf{C}_s^{-1}\right)^{-1}\left(\mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\boldsymbol{M}^* + \mathbf{C}_s^{-1}\,\boldsymbol{S}^*\right), \quad (13a)$$

or

$$\hat{\boldsymbol{S}} = \boldsymbol{S}^* - \mathbf{C}_s\mathbf{T}^{\mathrm{T}}\left(\mathbf{C}_m + \mathbf{T}\mathbf{C}_s^{-1}\mathbf{T}^{\mathrm{T}}\right)^{-1}\left(\mathbf{T}\boldsymbol{S}^* - \boldsymbol{M}^*\right). \quad (13b)$$

The covariance matrix of estimates $\hat{\boldsymbol{S}}$ can also be obtained using two formally equivalent formulations:

$$\mathbf{C}_{\hat{\boldsymbol{S}}} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\mathbf{T} + \mathbf{C}_s^{-1}\right)^{-1}, \quad (14a)$$

or

$$\mathbf{C}_{\hat{\boldsymbol{S}}} = \mathbf{C}_s - \mathbf{C}_s\mathbf{T}^{\mathrm{T}}\left(\mathbf{C}_m + \mathbf{T}\mathbf{C}_s^{-1}\mathbf{T}^{\mathrm{T}}\right)^{-1}\mathbf{T}\mathbf{C}_s \quad (14b)$$

Most efforts in deriving emission sources, and generally in assimilating geophysical parameters, rely on these basic equations (e.g. Hartley and Prinn, 1993; Elbern et al., 1997; Dee and Da Silva, 1998; Khattatov et al., 2000; Kasibhatla et al., 2000; Para et al., 2003).

Equations (13a) and (13b), as well as (14a) and (14b), are considered to be generally equivalent (e.g. see Tarantola, 1987). One of the important differences is that the matrix ($\mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\mathbf{T}$+$\mathbf{C}_S^{-1}$) inverted in Eqs. (13a) and (14a) has dimension N_S (the number of retrieved parameters) whereas ($\mathbf{C}_m$+$\mathbf{T}\mathbf{C}_S\mathbf{T}^{\mathrm{T}}$) inverted in Eqs. (13b) and (14b) has the dimension N_m (the number of measurements). In these regards, the pairs of Eqs. (13) and (14) are fully equivalent when N_m=N_S. Equations (13a) and (14a) are preferable for inverting redundant measurements (N_m>N_S), whereas Eqs. (13b) and (14b) are preferable for inverting underdetermined measurement sets (N_m<N_S). Indeed, Eq. (13a) directly relates to LSM Eq. (9), where the estimate $\hat{\boldsymbol{S}}$ is mostly determined by the measurement term $\mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\boldsymbol{M}^*$ and the generally minor a priori term is mainly expected to provide uniqueness and stability to the solution. In contrast, in Eq. (13b) the solution $\hat{\boldsymbol{S}}$ is expressed in the form of an a priori estimate $\boldsymbol{S}^*$ corrected or "filtered" by measurements, which is the situation when the number measurements N_m is small (N_m<N_S), and cannot fully determine the set of unknowns $\boldsymbol{a}$, but can improve the assumed a priori values $\boldsymbol{S}^*$. Also, it should be noted that the problem of source retrieval as formulated by Eq. (5) assumes the simultaneous retrieval of the entire vector $\boldsymbol{S}$, which includes global emission sources for the entire time period considered. However, the problem of emissions retrieval (e.g. Hartley and Prinn, 1993) and data assimilation in general (Dee and Da Silva, 1998; Khattatov et al., 2000) is often formulated as a time-sequential correction to a known parameter field based on observations, whereas the optimal estimation Eq. (13b) is used to optimize the forecast of $\boldsymbol{S}(t_i)$, i.e. emission at time t_i, based on known values of emission at previous time t_{i-1}:

$$\boldsymbol{S}_t = \boldsymbol{S}_{t-1} - \mathbf{C}_{S_{t-1}}\mathbf{T}_t^{\mathrm{T}}\left(\mathbf{C}_{m_{t+1}} + \mathbf{T}_t\mathbf{C}_{S_{t-1}}^{-1}\mathbf{T}_t^{\mathrm{T}}\right)^{-1}\left(\mathbf{T}_t\boldsymbol{S}_t - \boldsymbol{M}_{t+1}^*\right), \quad (15a)$$

and the covariance matrix $\mathbf{C}_{S_t}$ is the following:

$$\mathbf{C}_{S_t} = \mathbf{C}_{S_{t-1}} - \mathbf{C}_{S_{t-1}}\mathbf{T}_t^{\mathrm{T}}\left(\mathbf{C}_{m_{t+1}} + \mathbf{T}_t\mathbf{C}_{S_{t-1}}^{-1}\mathbf{T}_t^{\mathrm{T}}\right)^{-1}\mathbf{T}_t\mathbf{C}_{S_{t-1}}. \quad (15b)$$

where the index "t" indicates that the vectors are associated with time step t. Correspondingly, Eqs. (15) does not solve Eq. (5) directly, rather, it searches for a solution by solving the following sequence of the equations, formulated for a single time step:

$$\boldsymbol{M}_{t+1}^* = \boldsymbol{M}_{t+1}^{\text{meas}} - \mathbf{T}_t\boldsymbol{M}_t = \mathbf{T}_t\boldsymbol{S}_t, \quad (16)$$

where $\boldsymbol{M}_{t+1}^*$=$\boldsymbol{M}_{t+1}^{\text{meas}}$–$\mathbf{T}_t$ $\boldsymbol{M}_t$ is the vector of mass measured at time step t+1, corrected for the effect of aerosol mass $\boldsymbol{M}_t$ present in the atmospheres at the previous time step t; $\mathbf{S}_t$ is the vector of emission sources at time step t, $\mathbf{T}_t$ is the matrix describing the aerosol mass transport from time step t to time step t+1. The vectors $\boldsymbol{S}_t$ and $\boldsymbol{M}_t$ relate to vectors $\boldsymbol{S}$ and $\boldsymbol{M}$used in Eq. (5) as follows:

$$\boldsymbol{S}^{\mathrm{T}}=(\boldsymbol{S}_{t+n}, \ldots, \boldsymbol{S}_{t+1}, \boldsymbol{S}_t)^{\mathrm{T}} \text{ and } \boldsymbol{M}^{\mathrm{T}}= (\boldsymbol{M}_{t+n}, \ldots, \boldsymbol{M}_{t+1}, \boldsymbol{M}_t)^{\mathrm{T}}. \quad (17)$$

The relationship between matrixces $\mathbf{T}_t$ and $\mathbf{T}$ can be seen from Eq. (B5) in the Appendix.

Correspondingly, instead of the joint system given by Eq. (10), Eqs. (15) solves the following joint system:

$$\begin{cases} \boldsymbol{M}_{t+1}^{\text{meas}} = \boldsymbol{M}_t\left(\boldsymbol{S}_t\right) + \boldsymbol{\Delta}_M \\ \boldsymbol{S}_t^* = \boldsymbol{S}_{t-1} + \boldsymbol{\Delta}_S \end{cases}, \quad (18)$$

where the second line describes an a priori assumption of continuity between emissions at time steps t and $t-1$. (Note

that in Eq. (10) the a priori estimates $\boldsymbol{S}^*$ do not directly assume such continuity). The solution given by Eqs. (15) corresponds to a minimum of the following quadratic form:

$$2\,\Psi_{t+1}(\boldsymbol{S}_t) = \Delta\boldsymbol{M}_{t+1}^{\mathrm{T}}\mathbf{C}_{m_{t+1}}^{-1}\Delta\boldsymbol{M}_{t+1} + \Delta\boldsymbol{S}_t^{\mathrm{T}}\mathbf{C}_{S_{t-1}}^{-1}\Delta\boldsymbol{S}_t, \quad (19)$$

where $\Delta\boldsymbol{M}_{t+1}=\mathbf{T}_t\boldsymbol{S}_t-\boldsymbol{M}_{t+1}^*$ and $\Delta\boldsymbol{S}_t=\boldsymbol{S}_t-\boldsymbol{S}_{t-1}$.

This sequential correction (filtering) given by Eqs. (15) are widely known as a "Kalman filter", named after the author (Kalman, 1960) who originated the technique for engineering purposes.

Constrained inversion techniques are also widely used in remote sensing for retrieving vertical profiles of atmospheric properties (pressure, temperature, gaseous concentrations, etc.), where Eqs. (13–14) are associated with studies by Strand and Westwater (1968) and Rodgers (1976). It should be noted that in remote sensing, Eq. (13b) is not related to a sequential time retrieval (as considered by Kalman (1960)), but instead it is formulated for retrieving the entire vector $\boldsymbol{S}$ of unknowns as given by Eq. (17) (Rodgers, 1976). The important difference between Eqs. (13b), (14b) and the Kalman filter Eq. (15) is that the solution $\boldsymbol{S}_t$ of Eq. (15) is influenced only by the observations performed at one time step $t+1$, whereas in Eqs. (13b), (14b) (as well as in Eqs. 13a, 14a), the component $\hat{\boldsymbol{S}}_t$ of the entire solution $\hat{\boldsymbol{S}}$ can be influenced by observations of aerosol ontained at later time steps.

2.3 Inversion constrained by a priori smoothness constraints (limiting derivatives of the solution)

Equations (13–14) illustrate only the group of methods for performing constrained inversions, where the constraints explicitly contain the a priori estimates $\boldsymbol{S}^*$ of unknowns. Another group of popular constrained inversion methods does not restrict the magnitudes of the solution $\hat{\boldsymbol{S}}$; instead these methods use smoothness constraints that limit only the differences between elements $\hat{S}_j$ of the solution vector $\hat{\boldsymbol{S}}$. If the vector $\boldsymbol{S}$ is discrete analog of a continuous function, then the smoothness constraints can be considered as a priori limitations on the function $S(t, x, y, z)$, so the smoothness constraints can be considered a priori limitations on the derivatives of the function $S(t, x, y, z)$ with respect to time or spatial coordinates. The potential advantage of smoothness constraints is the fact that, in principle, using smoothness constraints imposes weaker limitations on the solution than using a priori constraints (since knowledge of function derivatives is less constraining than knowledge of function itself).

Numerous atmospheric remote sensing retrievals using smoothness constraints are based on the constrained inversion approach originated by Phillips (1962), Tikhonov (1963) and Twomey (1963). If one formally applies the Phillips-Tikhonov-Twomey approach for solving Eq. (5), the solution would be the following:

$$\hat{\boldsymbol{S}} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{T} + \gamma\Omega\right)^{-1}\mathbf{T}^{\mathrm{T}}\boldsymbol{M}^*, \quad (20)$$

where γ is the Lagrange parameter and Ω is the so-called smoothness matrix of n-th differences. For example, for the second differences, the matrix Ω is the following:

$$\Omega = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \ldots & & & \\ -2 & 5 & -4 & 1 & 0 & 0 & \ldots & & \\ 1 & -4 & 6 & -4 & 1 & 0 & 0 & \ldots & \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & 0 & \ldots \\ & & & & \ldots & & & & \\ & & & \ldots & 0 & 1 & -4 & 5 & -2 \\ & & & & \ldots & 0 & 1 & -2 & 1 \end{pmatrix}. \quad (21)$$

The solution of Eq. (20) corresponds to a minimum of the following quadratic form:

$$2\,\Psi(\boldsymbol{S}) = 2\,(\Psi_m + \Psi_{\mathrm{smooth}}) = \Delta\boldsymbol{M}^T\Delta\boldsymbol{M} + \gamma\boldsymbol{S}^{\mathrm{T}}\Omega\boldsymbol{S}. \quad (22)$$

In contrast with Eqs. (13–14), the original Phillips-Tikhonov-Twomey technique was not based on direct assumptions about the error statistics. Nevertheless, this formula can be generalized within the statistical formalism by using normal noise assumptions (e.g. see Dubovik, 2004). The principal difference of Eq. (20) from Eqs. (13–14) is the fact that Eq. 20) does not use a priori values of unknowns S_i. Instead, Eq. (20) limits the differences between the components S_i of the vector $\boldsymbol{S}$. For example, if the vector $\boldsymbol{S}$ is a discrete analog of a continuous function of one parameters x, e.g.

$$S_i = S(x_i), \quad (23)$$

where x_i are equidistant points ($x_{i+1}=x_i+\Delta x$), then the a priori term in the minimized quadratic form (Eq. 22) would represent the norm of n-th derivatives (see Twomey, 1977; Dubovik, 2004):

$$\int_{x_{\min}}^{x_{\max}}\left(\frac{d^nS(x)}{dx^n}\right)^2 dx \approx \sum_{x_i=x_{\min}}^{x_i=x_{\max}}\left(\frac{\Delta^n(x_i)}{(\Delta x)^n}\right)^2 =$$

$$(\Delta x)^{-n}\,(\mathbf{D}_n\boldsymbol{S})^{\mathrm{T}}\,(\mathbf{D}_n\boldsymbol{S}) \sim \boldsymbol{S}^{\mathrm{T}}\left(\mathbf{D}_n^{\mathrm{T}}\mathbf{D}_n\right)\boldsymbol{S} = \boldsymbol{S}^{\mathrm{T}}\Omega_n\boldsymbol{S}, \quad (24)$$

where $\mathbf{D}_n$ is the matrix of n-th differences:

$$\begin{aligned} \Delta^1 &= S_{i+1} - S_{i,} && (n=1),\\ \Delta^2 &= S_{i+2} - 2S_{i+1} + S_{i,} && (n=2),\\ \Delta^3 &= S_{i+3} - 3S_{i+2} + 3S_{i+1} - S_{i,} && (n=3). \end{aligned} \quad (25)$$

For example, matrix $\mathbf{D}_2$ of second differences is the following:

$$\mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & \ldots & & & & \\ 0 & 1 & -2 & 1 & 0 & \ldots & & & \\ 0 & 0 & 1 & -2 & 1 & 0 & \ldots & & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \\ \ldots & \ldots & \ldots & \ldots & \ldots & 0 & 1 & -2 & 1 \end{pmatrix}. \quad (26)$$

The corresponding smoothness matrix $\Omega_2=\mathbf{D}_2^{\mathrm{T}}\mathbf{D}_2$ is given by Eq. (21).

Thus, in many remote sensing applications where parameter functions $S(x_i)$ are retrieved, using smoothness constraints as shown in Eq. (20) is fruitful and popular. For example, such constraints are widely used in aerosol size distribution retrievals, for eliminating unrealistically strong oscilations in the dependence of aerosol particle concentration on particle size (e.g. Twomey, 1977; King et al., 1978; Nakajima et al., 1996; Dubovik and King, 2000, etc.). Using a priori estimates as solution constraints in those applications tends to over-constrain the retrievals.

Using a priori limitations on the derivatives (shown above) does not seem to be as popular for geophysical parameter data assimilation and inversion of tracer modeling. Such constraints are certainly included in general formalations of assimilation techniques (e.g. Navon, 1997), and they have been utilized for oceanographic data assimilations (e.g. Thacker, 1988; Thacker and Long, 1988; Yaremchul et al., 2001, 2002). Nonetheless, inverse modeling techniques commonly favor Bayesian formulations that constrain the solution with a priori estimates of terms, as shown in Eq. (13) (e.g. see the review by Lahoz et al., 2007). Constraints on time and space variability are often included in Bayesian formulations, by using in Eqs.(13) the covariance matrix $\mathbf{C}_S$ of a priori estimates having non-zero non-diagonal elements (e.g. Rodenbeck et al., 2003; Michalak et al., 2004; Houweling et al., 2004).

One of the many possible reasons for the unpopularity of a priori limitations on the derivatives in inverse modeling is probably the fact that tracer modeling deals with 4D characteristics. For example, the unknown vector $\boldsymbol{S}$ in Eq. (5) represents global aerosol sources. Correspondingly, instead of one parametric function shown by Eq. (23), we should consider vector $\boldsymbol{S}$ as the discrete equivalent of the 4D function:

$$S_i = S(t_i, x_j, y_k, z_m), \tag{27}$$

i.e. vector $\boldsymbol{S}$ has a total of $N_t \times N_x \times N_y \times N_z$ elements, where N_t,N_x ,N_y and N_z are the total numbers of discrete points for coordinates t, x, y and z, respectively. Obviously, the variability of emissions $S(t, x, y, z)$ with time t, vertically with z and horizontally with y and x does not have to be the same. This is why, using a single smoothness term in Eq. (22) with a single smoothness matrix $\boldsymbol{\Omega}$ (as the one given in Eq. 21) is not appropriate for constraining the retrieval of four-dimensional characteristic $S(t, x, y, z)$. At the same time, some temporal and spatial horizontal and vertical continuity of aerosol emission can naturally be expected (the same is applicable for most of geophysical parameters). Therefore, applying smoothness constraints on the variability of $S(t, x, y, z)$ with each coordinate instead of using a single variability constraint can be useful. However, that would require using several constraints simultaneously. A possible approach for using multiple constraints is discussed by Dubovik and King (2000) and Dubovik (2004).

2.4 Constrained inversion within multi-term LSM

Dubovik and King (2000) and Dubovik (2004) demonstrated that Eqs. (13a) and (20) can be naturally derived and generalized by considering inversions with a priori constraints as a version of multi-term LSM. Formally, both measured and a priori data can be written as

$$\boldsymbol{f}_k^* = \boldsymbol{f}_k(\boldsymbol{a}) + \Delta\boldsymbol{f}_k, \qquad (k = 1, 2, \ldots, K), \tag{28}$$

where $\boldsymbol{f}_k^*$ are vectors of measurements, $\Delta\boldsymbol{f}_k$ are vectors of measurement errors, and $\boldsymbol{f}_k(\boldsymbol{a})$ are forward models that allow adequate simulations of $\boldsymbol{f}_k$ from predetermined parameters $\boldsymbol{a}$. Index k denotes different data sets. The separations of data sets assume that the data from the same data set have similar error structure, independent of errors in the data from other sets. Assuming that $\Delta\boldsymbol{f}_k$ is normally distributed with covariance matrices $\mathbf{C}_k$, the MML optimum solution of Eq. (28) corresponds to the minimum of the following quadratic form:

$$2\Psi(\boldsymbol{a}) = \sum_{k=1}^{K} (\Delta\boldsymbol{f}_k)^{\mathrm{T}} (\mathbf{C}_k)^{-1} (\Delta\boldsymbol{f}_k) = \min, \tag{29a}$$

where $\Delta\boldsymbol{f}_k = \boldsymbol{f}_k(\boldsymbol{a}) - \boldsymbol{f}_k^*$. This condition does not prescribe the value of the minimum and, therefore, it can be formulated via weighting matrices:

$$2\Psi'(\boldsymbol{a}) = \sum_{k=1}^{K} \gamma_k (\Delta\boldsymbol{f}_k)^{\mathrm{T}} (\mathbf{W}_k)^{-1} (\Delta\boldsymbol{f}_k) = 2\sum_{k=1}^{K} \gamma_k \Psi_k'(\boldsymbol{a}) = \min, \tag{29b}$$

where weighting matrices $\mathbf{W}_k$ defined as:

$$\mathbf{W}_k = \frac{1}{\varepsilon_k^2}\mathbf{C}_k \text{ and } \gamma_k = \frac{\varepsilon_1^2}{\varepsilon_k^2}. \tag{30}$$

Here ε_k^2 is the first diagonal element of $\mathbf{C}_k$, i.e. $\varepsilon_k^2 = \{\mathbf{C}_k\}_{11}$. Using the weighting matrices $\mathbf{W}_k$ is, in principle, equivalent to using covariance matrices $\mathbf{C}_k$, although sometimes it is more convenient because it explicitly shows that the minimization depends only on the relative contribution of each term Ψ_k' to the total Ψ'. The Lagrange parameters γ_k weight the contribution of each source relative to the contribution of first data source (obviously, γ_1=1). The minimum of the multi-term quadratic form given by Eq. (29) can be found by the multi-term equivalent of Eq. (9):

$$\hat{\boldsymbol{a}} = \left(\sum_{k=1}^{K} \gamma_k (\mathbf{K}_k)^{\mathrm{T}} (\mathbf{W}_k)^{-1} (\mathbf{K}_k)\right)^{-1} \left(\sum_{k=1}^{K} \gamma_k (\mathbf{K}_k)^{\mathrm{T}} (\mathbf{W}_k)^{-1} \boldsymbol{f}_k^*\right). \tag{31}$$

The corresponding covariance matrix can be estimated from the following:

$$\mathbf{C}_{\hat{\boldsymbol{a}}} \approx \left(\sum_{k=1}^{K} \gamma_k (\mathbf{K}_k)^{\mathrm{T}} (\mathbf{W}_k)^{-1} (\mathbf{K}_k)\right)^{-1} \hat{\varepsilon}^2, \tag{32}$$

where $\hat{\varepsilon}^2$ is estimated from the minimum of Ψ' as: $\hat{\varepsilon}^2 = \Psi'/(N_f - N_a)$, N_f is the total number of elements $\{\boldsymbol{f}_k\}_j$

(in all sets $\boldsymbol{f}_k$), N_a is the total number of unknown parameters a_i.

Using the above multi-term equations, one can formulate an inversion with smoothness constraints on the variability of $S(t, x, y, z)$, separately for each coordinate. Specifically, such multiple smoothness constraints represent a solution of the following joint system:

$$\begin{cases} \boldsymbol{M}^{\text{meas}} = \boldsymbol{M}(\boldsymbol{S}) + \Delta_M \\ \mathbf{0}_t^* = \Delta_t^n(\boldsymbol{t}, \boldsymbol{x}) + \Delta_t \\ \mathbf{0}_x^* = \Delta_x^n(\boldsymbol{t}, \boldsymbol{x}) + \Delta_x \\ \mathbf{0}_y^* = \Delta_y^n(\boldsymbol{t}, \boldsymbol{x}) + \Delta_y \\ \mathbf{0}_z^* = \Delta_z^n(\boldsymbol{t}, \boldsymbol{x}) + \Delta_z \end{cases}, \tag{33}$$

where $\Delta^n(...)$ denotes the n-th difference (see Eq. 25) of aerosol sources with respect to time, or to coordinates x, y or z. For example, for the time coordinate t, second differences (Eq. 25) can be written as:

$$\{\Delta_t^2(t, \boldsymbol{x})\}_g = \Delta_t^2(t_i, x_j, y_k, z_m) = S(t_{i+1}, x_j, y_k, z_m) - 2S(t_i, x_j, y_k, z_m) + S(t_{i-1}, x_j, y_k, z_m). \tag{34a}$$

where the index g can be calculated (see Fig. 2), for example, as follows

$$g = (i-1)N_xN_yN_z + (j-1)N_yN_z + (k-1)N_z + (m-1). \tag{34b}$$

The second line in Eq. (33) states that differences $\Delta^n(t_i)$ are equal to zero with errors Δ_{t_i}. Accordingly, for n=2 the vectors $\mathbf{0}_t^*$, $\boldsymbol{\Delta}^2(t)$ and Δ_t consist of $(N_t-2)\times N_x \times N_y \times N_z$ zeros, $\Delta_t^2(t_i, x_j, y_k, z_m)$ and Δ_{t_i}, respectively. The 3rd, 4th and 5th lines in Eq. (33) are defined in the same way for coordinates x, y and z respectively. Assuming that Δ_t, Δ_x, Δ_y and Δ_z are normally distributed with zero means and diagonal covariance matrices $\mathbf{C}_t = \varepsilon_t^2 \mathbf{I}_t$, $\mathbf{C}_x = \varepsilon_x^2 \mathbf{I}_x$, $\mathbf{C}_y = \varepsilon_y^2 \mathbf{I}_y$ and $\mathbf{C}_z = \varepsilon_z^2 \mathbf{I}_z$, the multi-term LSM solution of Eq. (33) can be written as follows:

$$\hat{\boldsymbol{S}} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{W}_m^{-1}\mathbf{T} + \gamma_t\Omega_t + \gamma_x\Omega_x + \gamma_y\Omega_y + \gamma_z\Omega_z\right)^{-1} \mathbf{T}^{\mathrm{T}}\mathbf{W}_m^{-1}\boldsymbol{M}^*, \tag{35}$$

where

$$\mathbf{W}_m = \frac{1}{\varepsilon_m^2}\mathbf{C}_m, \quad \gamma_t = \frac{\varepsilon_t^2}{\varepsilon_m^2}, \quad \gamma_x = \frac{\varepsilon_x^2}{\varepsilon_m^2}, \quad \gamma_y = \frac{\varepsilon_y^2}{\varepsilon_m^2}, \quad \gamma_z = \frac{\varepsilon_z^2}{\varepsilon_m^2},$$

and $\varepsilon_m^2 = \{\mathbf{C}_m\}_{11}$, $\varepsilon_t^2 = \{\mathbf{C}_t\}_{11}$, $\varepsilon_x^2 = \{\mathbf{C}_x\}_{11}$, $\varepsilon_y^2 = \{\mathbf{C}_y\}_{11}$ and $\varepsilon_z^2 = \{\mathbf{C}_z\}_{11}$. The matrices Ω are determined via corresponding matrices of n-th differences $\Omega = \mathbf{D}_n^{\mathrm{T}}\mathbf{D}_n$. Equation (35) yields the minimum of the following quadratic form:

$$2\,\Psi(\boldsymbol{S}) = 2\Psi_m(\boldsymbol{S}) + 2\sum_{(q=t,x,y,z)} \gamma_q\,\Psi_q(\boldsymbol{S}) = 2\,(\Delta\boldsymbol{M})^{\mathrm{T}}\,\mathbf{W}_m^{-1}\,\Delta\boldsymbol{M} + 2\sum_{(q=t,x,y,z)} \gamma_q \boldsymbol{S}^{\mathrm{T}}\Omega_q\boldsymbol{S}. \tag{36}$$

Each of the smoothness terms in this equation can be considered as a discrete equivalent of the norm of the n-th partial derivative. For example, for the second term, which corresponds to the time coordinate t, one can write:

$$\Psi_t(\boldsymbol{S}) = \sum_{(i,j,k,m)} \left(\frac{\Delta_t^n(t_i, x_j, y_k, z_m)}{(\Delta t)^n}\right)^2 \approx \Psi_t'(s) = \int_{t_{\min}}^{t_{\max}} \left(\frac{\partial^n s(t, \boldsymbol{x})}{\partial t^n}\right)^2 dx \tag{37a}$$

and

$$\Psi_t(\boldsymbol{S}) = (\Delta t)^{1-2n}\left(\mathbf{D}_{(n,t)}\boldsymbol{S}\right)^{\mathrm{T}}\mathbf{D}_{(n,t)}\boldsymbol{S} = (\Delta t)^{1-2n}\,\boldsymbol{S}^{\mathrm{T}}\left(\mathbf{D}_{(n,t)}^{\mathrm{T}}\mathbf{D}_{(n,t)}\right)\boldsymbol{S} = (\Delta t)^{1-2n}\,\boldsymbol{S}^{\mathrm{T}}\Omega_t\boldsymbol{S}, \tag{37b}$$

where the matrix $\mathbf{D}_{(n,t)}$ is the matrix of differences corresponding to n-th partial derivative with respect to time. For example, $\mathbf{D}_{(2,t)}\boldsymbol{S}$ would produce a vector with elements equal to the second differences, as shown in Eq. (34).

Thus, it was shown above that using the multi-term LSM approach, one can apply multiple smoothness constraints in the retrieval of emission sources. Therefore, it is possible to utilize knowledge about typical time, horizontal and vertical variability of the emissions as a priori constraints on the retrieval. As shown in Eqs. (33–37), such smoothness constraints are included as restrictions on the n-th partial derivatives of $S(t, x, y, z)$ assuming zero values for the corresponding differences in Eq. (33), and that values of the Lagrange parameters determine the variations from zero. The order of the differences assumed relates to the character of expected variability; for example for a one-parameter function $S(t)$, there are following relationships:

$$\begin{aligned} &\Delta^1(t) = 0 \rightarrow S(t) = \text{const} && - \text{constant}, \\ &\Delta^2(t) = 0 \rightarrow S(t) = A + Bt && - \text{straight line}, \\ &\Delta^3(t) = 0 \rightarrow S(t) = A + Bt + Ct^2 && - \text{parabola, etc.} \end{aligned} \tag{38}$$

Note that the constraints employed in Kalman filter Eq. (15) are equivalent to restricting the first differences (see Eq. 18), i.e. assuming a priori linear continuity of the source variability. In a contrast, Eqs. (35–36) with multiple constraints allow using higher order constraints on time variability, and can constrain not only time, but also the space and vertical variability of the emissions.

2.5 Inversion using adjoint equations

Methods analogous to Eq. (13) are used for retrieving CO_2 sources from surface-based and satellite observations (e.g. see Enting et al., 1995; Patra et al., 2003). However, direct implementation of Eqs. (9, 13) for retrieving aerosol emission sources is not feasible, due to the very large dimensions of matrix $\mathbf{T}$ and vectors $\boldsymbol{S}$ and $\boldsymbol{M}$. For example, CO_2 emission sources can be assumed constant for monthly or yearly periods over large geographic areas (e.g., Patra et al., 2003, used 22 and 53 global regions). The temporal and spatial variability of tropospheric aerosols and their sources are much higher. For aerosols, the GOCART model (see below)

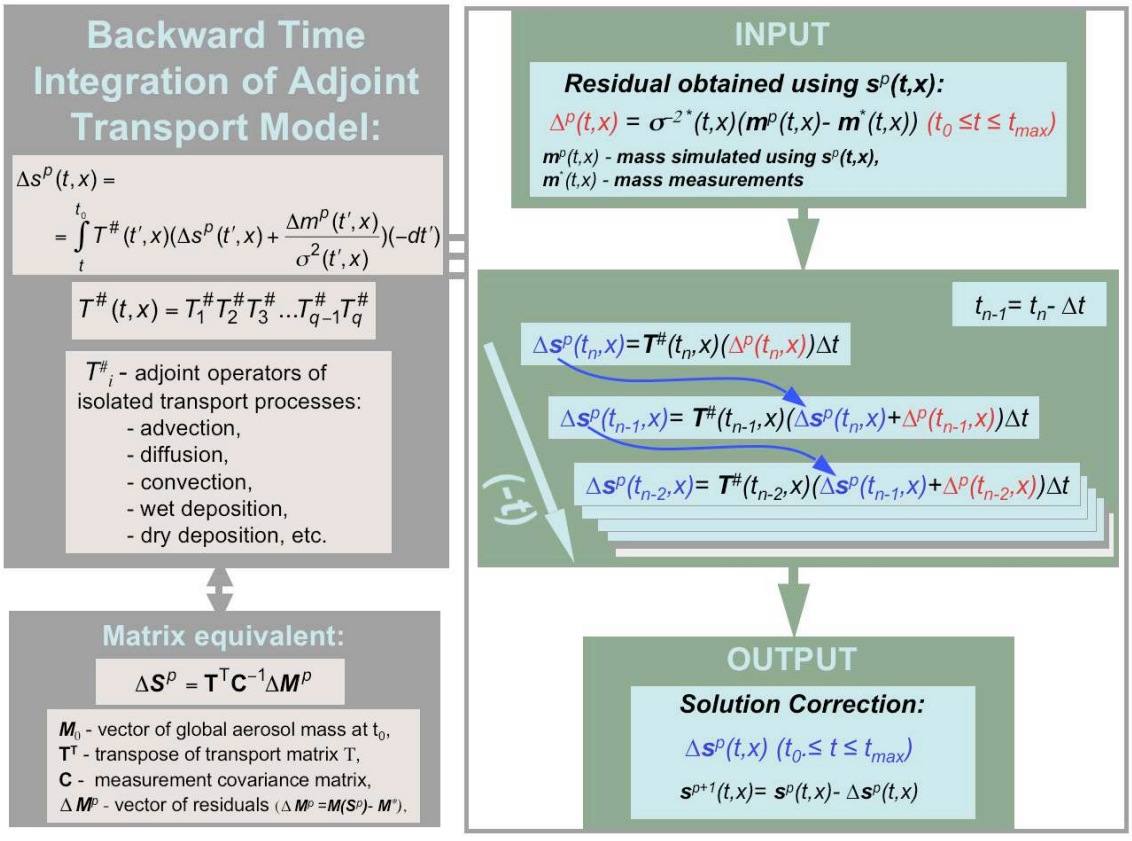

Fig. 4. Illustration of the calculation of gradient $\nabla\Psi_m(\boldsymbol{S})$ (Eq. 40) by means of implementing the sequential backward time integration of the adjoint aerosol transport model.

has $2^\circ \times 2.5^\circ$ horizontal resolution (144 longitudes, 91 latitudes) and 30 vertical layers, with the possibility of having variable sources in each layer. As a result, inverting a few weeks of observations using Eqs. (9) or (13) requires dealing with the vector $\boldsymbol{S}$ having dimension N_S far exceeding 200 000, even under the conservative assumption that sources are near surface and constant during 24 hours. Performing the vector and matrix operations of Eqs. (9, 13) directly on terms of such high dimensionality is problematic. One way to avoid dealing with such large vectors and matrices is to perform the inversion using time sequential retrievals, as is given by Kalman filter formulation of Eq. (15), where the retrieval uses generally smaller matrices and vectors containing parameter values at only a single time step t_i. However, in Kalman filter procedure given in Eq. (15), the retrieval relies only on observations at a single time step, and on assuming linear continuity of the emission strength. However, the emitted aerosol is transported over a period of time, and therefore, observations during that entire period (a week) can be useful for the retrieval. In these regards, using Eqs. (10–13) seem preferable to Eqs. (15–19), and can be implemented with computational requirements close to those of forward modeling. To achieve this, the inversion routine must adopt the strategy used for global model forward simulations. As shown by Eq. (5), transport modeling can be formulated as a matrix operator; however, in practice, transport models are implemented with numerical time integration (Eq. 4), by sequentially computing chemical transports during each time step Δt (Eq. 2), and with separate treatment of isolated processes (Eq. 3). Figure 3 illustrates the relationship between the matrix formulation of aerosol transport Eq. (2), and direct time integration. A similar approach can be employed in inverse modeling, by developing so-called "adjoint" transport operators as formulated in a variational assimilation framework (e.g. Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Elbern et al., 1997; Menut et al., 2000; Elbern et al., 2007). The analogies between the variational and matrix formulations are rather apparent. (In order to assist the reader in understanding the considerations discussed below, Figs. 4–5 provide diagrams outlining matrix formulations and their continuous analogs). Indeed, any inversion can be implemented by iterations without the explicit use of matrix inversion. For example, a solution equivalent to the one of Eq. (9) can be obtained by the steepest descent iterative method:

$$\hat{\boldsymbol{S}}^{p+1} = \hat{\boldsymbol{S}}^p - t_p \Delta \hat{\boldsymbol{S}}^p, \tag{39}$$

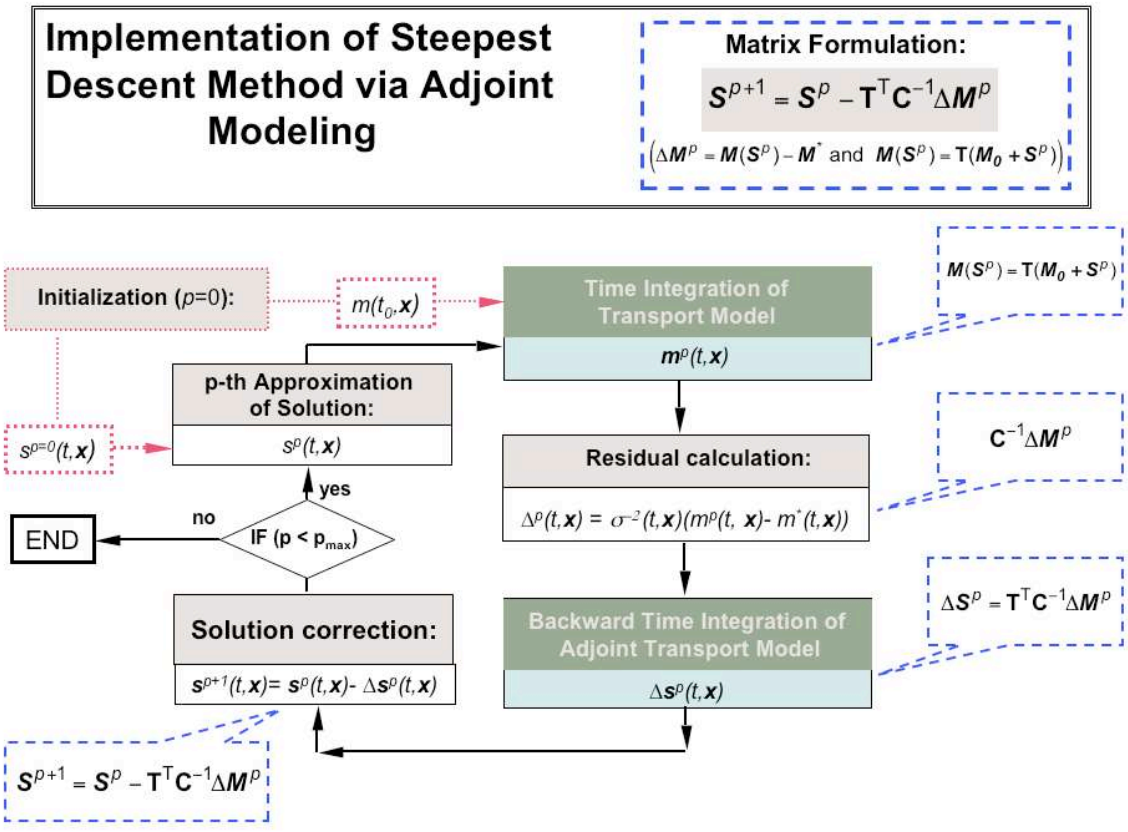

Fig. 5. Diagram illustrating implementation of the steepest descent iteration in terms of the adjoint modeling approach.

$$\Delta\hat{\boldsymbol{S}}^p = \nabla\Psi_m(\boldsymbol{S}^p) = \mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\Delta\boldsymbol{M}^p\,, \tag{40}$$

where $\Delta\boldsymbol{M}^p = \boldsymbol{M}(\boldsymbol{S}^p) - \boldsymbol{M}^*$, $\nabla\Psi_m(\boldsymbol{S})$ denotes the gradient of $\Psi_m(\boldsymbol{S})$ and t_p is a non-negative coefficient. This method uses the fact that the gradient $\nabla\Psi_m(\boldsymbol{S})$ points in the direction of maximal local changes of $\Psi_m(\boldsymbol{S})$, and this direction ($t_p\nabla\Psi_m(\boldsymbol{S}^p)$, generally $t_p < 1$) can always be used to correct $\boldsymbol{S}^p$, so it moves toward the solution $\boldsymbol{S}^{'}$ that minimizes $\Psi_m(\boldsymbol{S})$, i.e $\Psi_m(\boldsymbol{S}^{p+1}) < \Psi_m(\boldsymbol{S}^p)$. Equations (39–40) do not require inversions of high dimension matrices (inverting a diagonal covariance matrix is trivial). The gradient $\nabla\Psi_m(\boldsymbol{S})$ can be simulated using a time integration scheme similar to the one employed for forward modeling so the matrix solution of the steepest descent method Eqs. (39–40) can be replaced by an analogous continuous operation. Namely, the elements of the gradient vector $\nabla\Psi_m(\boldsymbol{S})$ can be simulated in a manner similar to Eqs. (2–4) and the inversion can be implemented using a continuous analog of the gradient vector $\nabla\Psi_m(\boldsymbol{S})$ (see also Figs. 3–4). A continues equivalent of Eq. (40) can be written as follows (the detailed derivations is given in Appendix B):

$$\Delta\hat{s}^p(t, \boldsymbol{x}) = \int_t^{t_0} T^{\#}(t', \boldsymbol{x})\left(\Delta\hat{s}^p(t', \boldsymbol{x}) + \sigma^{-2}(t', \boldsymbol{x})\,\Delta m^p(t', \boldsymbol{x})\right)(-dt'), \tag{41a}$$

where

$$\Delta m^p(t, \boldsymbol{x}) = m^*(t, \boldsymbol{x}) - \int_{t_0}^{t} T(t', \boldsymbol{x})\left(m(t', \boldsymbol{x}) + s^p(t', \boldsymbol{x})\right)dt', \tag{41b}$$

and $T^{\#}(t, \boldsymbol{x})$ is the adjoint of the transport operator $T(t, \boldsymbol{x})$, that is composed of adjoints $T_i^{\#}(t, \boldsymbol{x})$ of the component processes $T_i(t, \boldsymbol{x})$:

$$T^{\#}(t, \boldsymbol{x}) = T_1^{\#}T_2^{\#}T_3^{\#}\ldots T_{q-1}^{\#}T_q^{\#}. \tag{41c}$$

The vectors $\Delta\hat{s}^p(t, \boldsymbol{x})$ and $\sigma^{-2}(t, \boldsymbol{x})\,\Delta m^p(t, \boldsymbol{x})$ denote functional equivalents of vectors $\Delta\hat{\boldsymbol{S}}^p$ and $\mathbf{C}_m^{-1}\Delta\boldsymbol{M}^p$ respectively. For example, if one intends to use the continuous function $\Delta\hat{s}^p(t, \boldsymbol{x})$ in numerical calculations, it can be represented by a vector $\Delta\hat{\boldsymbol{S}}^p$ with the following elements:

$$\left\{\Delta\hat{\boldsymbol{S}}^p\right\}_l = \Delta\hat{s}^p(t_i, x_j, y_k, z_m)\,, \tag{42a}$$

where the index l is determined in the same way as in Eq. (34b).

Similarly, if the observational errors are uncorrelated, i.e. the covariance matrix of the measurements $\mathbf{C}_m$ is diagonal, with the diagonal elements equal to $\sigma^2(t_i, x_j, y_k, z_m)$, the elements of vector $\mathbf{C}_m^{-1}\Delta\boldsymbol{M}^p$ relate to the continuous function

$\sigma^{-2}(t,\boldsymbol{x})\,\Delta m^p(t,\boldsymbol{x})$ in straightforward way:

$$\left\{\mathbf{C}_m^{-1}\Delta\boldsymbol{M}^p\right\}_l \rightarrow \sigma^{-2}\left(t_i,x_j,y_k,z_m\right)\Delta m^p\left(t_i,x_j,y_k,z_m\right). \quad (42b)$$

If observational errors do not have time correlations but do have spatial correlations, $\mathbf{C}_m$ has an array structure that can be included in the algorithm (see Eqs. (B12–B15) in Appendix B), provided one can formulate a weighting function $C^{-1}(t,\boldsymbol{x},\boldsymbol{x}')$ from the covariance function $C(t,\boldsymbol{x},\boldsymbol{x}')$, to perform a role analogous to the one of matrix $\mathbf{C}_m^{-1}$ in the discrete representation. Including the spatial correlations during each time moment is feasible because the model is integrated by time steps, and each step can be treated rather independently. However, accounting for observational errors that are correlated in time is not feasible without changing the structure of Eq. (41).

It is important to note that Eq. (41) is convenient for practical implementation of the inversion. Indeed, as outlined in Fig. 4 (compare with Fig. 3), Eq. (41) is related to Eq. (4), with the difference that it uses $\Delta m^p(t,\boldsymbol{x})$ in place of $s(t,\boldsymbol{x})$, and $\Delta\hat{s}^p(t,\boldsymbol{x})$ in place of $m(t,\boldsymbol{x})$, and it performs the backward time integration of the adjoint operator $T^{\#}(t,\boldsymbol{x})$. If $T(t,\boldsymbol{x})$ is functionally equivalent to the matrix operator $\mathbf{T}$, then the adjoint operator $T^{\#}(t,\boldsymbol{x})$ is an equivalent to the transposed matrix $\mathbf{T}^{\mathrm{T}}$ (Appendix A). Therefore, the main reason for developing the adjoint operator $T^{\#}$ from T can be illustrated by considering matrix transposition. For example, since the transport operator integration can be approximated using the split operator approach (e.g. see Jacob, 1999), where matrices corresponding to different atmospheric processes are multiplied at each time step (e.g. see Eq. 3). The following matrix identity is helpful:

$$(\mathbf{T}_3\,\mathbf{T}_2\,\mathbf{T}_1)^{\mathrm{T}} = (\mathbf{T}_1)^{\mathrm{T}}(\mathbf{T}_2)^{\mathrm{T}}(\mathbf{T}_3)^{\mathrm{T}}. \quad (43)$$

This reversing of the order of operations by transposition produces an overturned sequence of component process applications within each time step Eq. (41c), and reverses the order of integration in Eq. (41a), i.e. in backward time integration Eq. (41a). Also, the transposition of matrix $\mathbf{T}_i$ changes rows and columns, so if $\mathbf{T}$ is non-square, the input of $(\mathbf{T}_i)^{\mathrm{T}}$ should have the dimensions of $\mathbf{T}_i$ output, and vice versa. Thus, the adjoint model (Eq. 41) can be developed on the basis of the original model (Eq. 4) by reversing the order of operations and switching the inputs and outputs of the routines (e.g. Elbern et al., 1997; Menut et al., 2000).

Thus, using the adjoint of the transport model allows us to implement the LSM inversion (Eq. 9) without using explicit matrix inversions, and therefore demands only moderate computational efforts. As is shown in Fig. 5 each iteration in Eq. (41) requires one forward integration of the transport model (Eq. 41b) followed by one backward integration of the adjoint transport model (Eq. 41a).

The need to perform a number of iterations in Eq. (41) is a potential drawback of implementing inversions via adjoint modeling. Indeed, the steepest descent method of Eqs. (39–40), in general, converge to the exact solution after a very large number of iterations. The even faster method of conjugated gradients may require up to N_S iterations (Press et al., 1992). Nevertheless, a rather limited number of simple iterations appears to be sufficient for global inverse modeling of high dimensionality. For instance, the iterations of Eqs. (39–40) converge from an arbitrary initial guess to the solution rapidly if the following sequence tends toward the zero matrix (Dubovik, 2004):

$$\prod_{p=1}^{\infty}\left(\mathbf{I}-t_p\mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\mathbf{T}\right)\Rightarrow 0, \quad (44)$$

where $\mathbf{I}$ is unity matrix. It is clear that rapid convergence of Eqs. (39–40) can be achieved only if $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ is predominantly diagonal ($\mathbf{C}_m$ is often diagonal and does not cause problems). Fortunately, in transport modeling, the diagonal elements of $\mathbf{T}^{\mathrm{T}}\mathbf{T}$ dominate, because local aerosol emission typically influence only nearby locations (i.e. matrix $\mathbf{T}$ is rather sparse and has a large number of zeros, see Eq. (B5) in Appendix B).

It should be noted that Eqs. (41) expressing the inversion via adjoint operators, are generally analogous to techniques used in variational assimilation (e.g. Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Menut et al., 2000; Vukicevic et al., 2001). Nevertheless, the statistical estimation approach employed in our study makes it possible to establish direct relationships between Eqs. (41) and conventional LSM minimization which therefore improves flexibility in implementing the inversion. For example, using error covariances directly makes it possible to account for different levels of accuracy in the inverted observations. Moreover, formulating the inversion using a statistical approach is convenient for including several a priori constraints in the same retrieval, for example, by following multi-term LSM strategy discussed in Sect. 2.4.

2.6 Including a priori constraints in inversion, using adjoint equations

Equations (41) can be easily adopted for constrained inversion. Figures 6–7 illustrate the considerations discussed in this Section. For example, the inversion constraining the solution $\hat{\boldsymbol{S}}$ with its a priori estimates $\boldsymbol{S}^*$, shown as a matrix inversion in Eq. (13), can be implemented iteratively, e.g. using steepest descent iterations:

$$\hat{\boldsymbol{S}}^{p+1} = \hat{\boldsymbol{S}}^p - t_p\Delta\hat{\boldsymbol{S}}^p, \quad (45a)$$

$$\Delta\hat{\boldsymbol{S}}^p = \nabla\Psi_m(\boldsymbol{S}^p)+\nabla\Psi_S(\boldsymbol{S}^p) = \mathbf{T}^{\mathrm{T}}\mathbf{W}_m^{-1}\Delta\boldsymbol{F}^p+\gamma_s\mathbf{W}_s^{-1}\Delta\boldsymbol{S}^p. \quad (45b)$$

Here we used weighting matrices $\mathbf{W}_{\ldots}$ instead of covariance matrices $\mathbf{C}_{\ldots}$ in order to align these equations with the LSM multi-term formulations given by Eqs. (28–32). If we assume

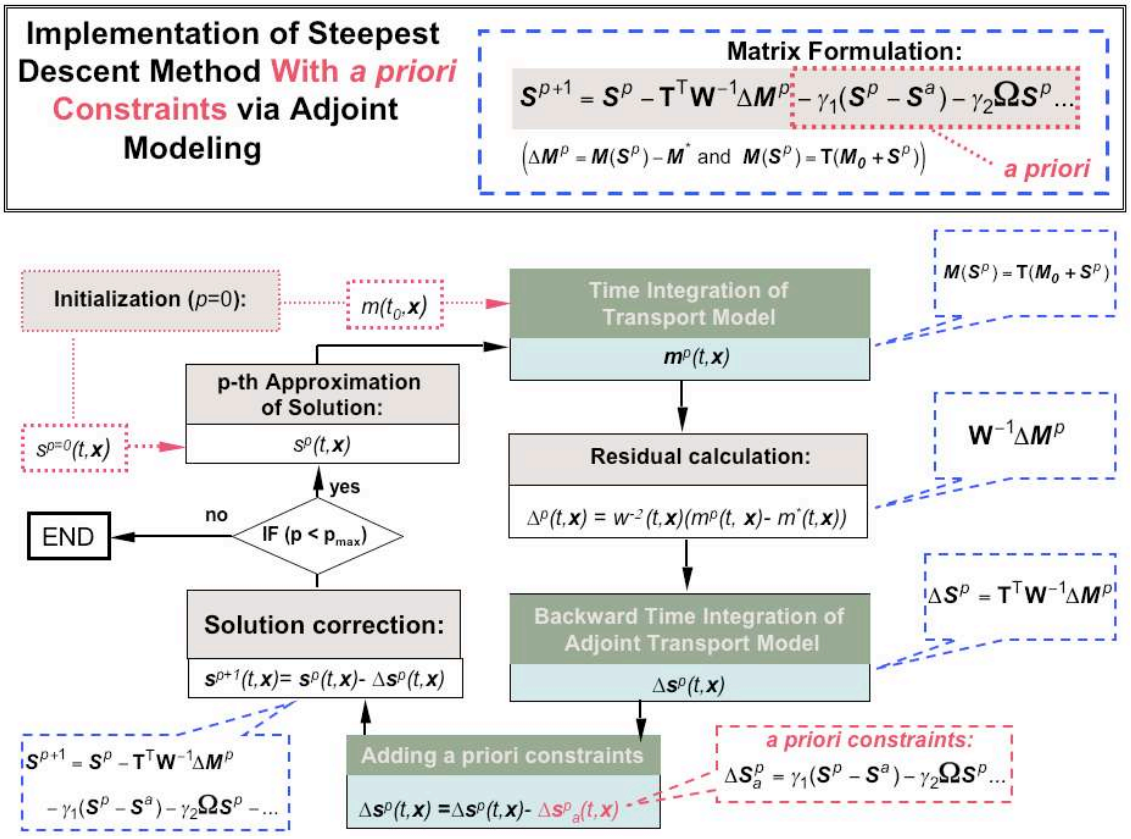

Fig. 6. illustrating implementation of the steepest descent method with a priori constraints, by means of adjoint aerosol transport modeling.

for simplicity that all measurements are statistically independent and have the same accuracy ε_m (i.e. $\mathbf{C}_m=\mathbf{I}\,\varepsilon_m^2 \rightarrow \mathbf{W}_m=\mathbf{I}$), and that all a priori estimates are statistically independent and have the same accuracy ε_s ($\mathbf{C}_S=\mathbf{I}\,\varepsilon_S^2 \rightarrow \mathbf{W}_S=\mathbf{I}$), then we can write the continuous analog to Eq. (45b) as follows:

$$\Delta\hat{s}^p(t,\boldsymbol{x}) = \int_t^{t_0} T^{\#}(t',\boldsymbol{x})\ \left(\Delta\hat{s}^p(t',\boldsymbol{x}) + \Delta m^p(t',\boldsymbol{x})\right)(-dt') + \gamma_S\left(\hat{s}^p(t',\boldsymbol{x}) - \hat{s}^*(t',\boldsymbol{x})\right). \quad (45c)$$

where $\gamma_S=\varepsilon_m^2/\varepsilon_S^2$.

The iterative analog to Eq. (35), constraining the solution by limiting the time and spatial derivatives of $\hat{S}(t,\boldsymbol{x})$, can be written as follows:

$$\hat{\boldsymbol{S}}^{p+1} = \hat{\boldsymbol{S}}^p - t_p\Delta\hat{\boldsymbol{S}}^p, \quad (46a)$$

$$\Delta\hat{\boldsymbol{S}}^p = \nabla\Psi_m(\boldsymbol{S}^p) + \sum_{(q=t,x,y,z)} \gamma_q\,\nabla\Psi_q(\boldsymbol{S}^p) = \mathbf{T}^T\mathbf{W}_m^{-1}\,\Delta\boldsymbol{F}^p + \sum_{(q=t,x,y,z)} \gamma_q\mathbf{D}_n^T\mathbf{D}_n\boldsymbol{S}^p. \quad (46b)$$

The function $\Delta\hat{s}^p(t,\boldsymbol{x})$, corresponding to vector $\Delta\hat{\boldsymbol{S}}^p$ can be formulated as follows:

$$\Delta\hat{s}^p(t,\boldsymbol{x}) = \int_t^{t_0} T^*(t,\boldsymbol{x})\ \left(\Delta\hat{s}^p(t',\boldsymbol{x}) + \Delta m^p(t',\boldsymbol{x})\right)(-dt') + \sum_{(q=t,x,y,z)} \gamma_q\,D_n^{\#}D_n s^p(t,\boldsymbol{x}) = \int_t^{t_0} T^*(t,\boldsymbol{x})\left(\Delta\hat{s}^p(t',\boldsymbol{x}) + \Delta m^p(t',\boldsymbol{x})\right)(-dt') + \sum_{(q=t,x,y,z)} \gamma_q\frac{\partial^{(2n)}s^p(t,\boldsymbol{x})}{\partial q^{(2n)}}, \quad (47)$$

where $\boldsymbol{x}'=(x_j, y_k, z_m)$, D_n denotes n-th derivative operator and $D_n^{\#}$ denotes the adjoint to the n-th derivative operator. For the adjoint operator $D_n^{\#}$, one can write the following:

$$D_n^{\#} = (-1)^n D_n. \quad (48a)$$

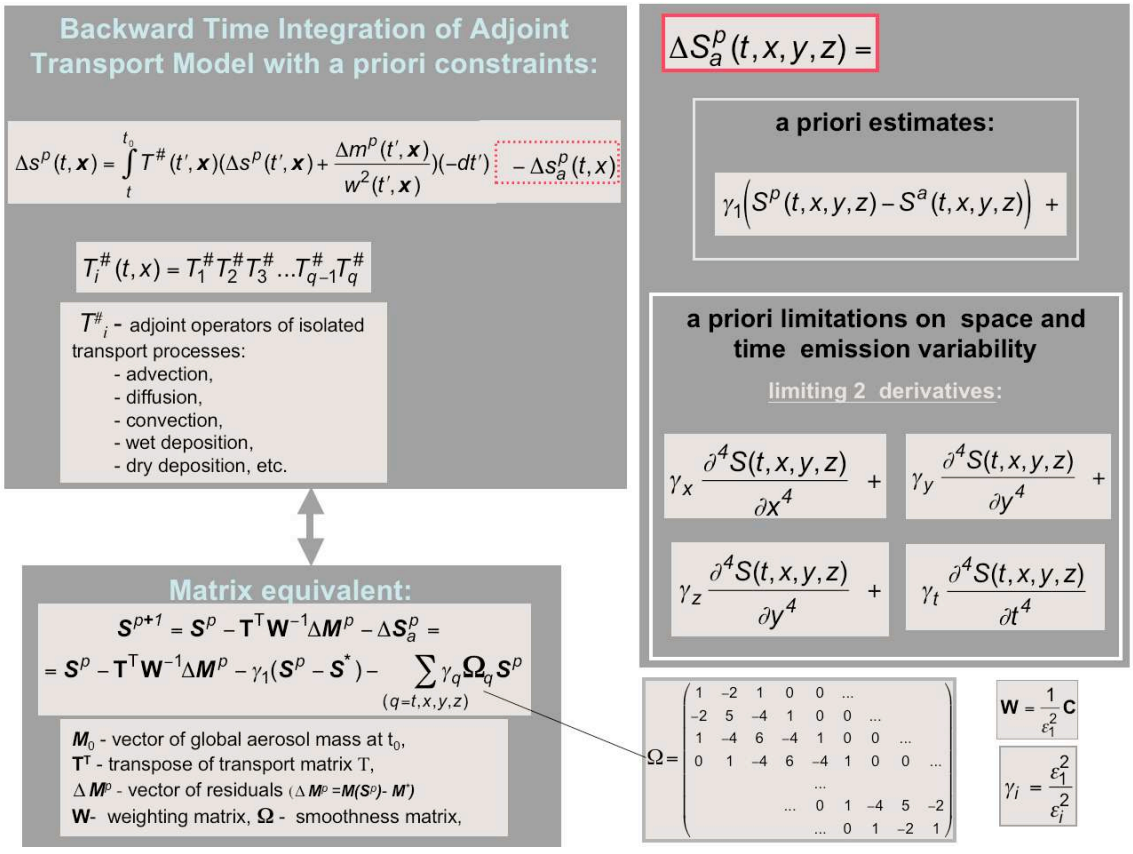

Fig. 7. Illustration of the combination of sequential backward time integration of the adjoint aerosol transport model with a priori constraints.

This identity can be obtained from the transposition of the matrices $\mathbf{D}_n$. For example, for $\mathbf{D}_1{}^{\mathrm{T}}$ we have the following:

$$D_1^{\mathrm{T}} = \begin{pmatrix} 1 & -1 & 0 & \dots & & & & & \\ 0 & 1 & -1 & 0 & \dots & & & & \\ 0 & 0 & 1 & -1 & 0 & \dots & & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & 1 & -1 & 0 & \\ \dots & \dots & \dots & \dots & \dots & 0 & 1 & -1 & \end{pmatrix}^{\mathrm{T}}$$

$$= \begin{pmatrix} 1 & 0 & \dots & & & & \\ -1 & 1 & 0 & \dots & & & \\ 0 & -1 & 1 & 0 & \dots & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & -1 & 1 & 0 \\ \dots & \dots & \dots & \dots & 0 & -1 & \end{pmatrix}. \quad (48b)$$

Here one can see that with exception of the first and last lines of $\mathbf{D}_1{}^{\mathrm{T}}$, each row corresponds to first differences. Similarly, it is easy to demonstrate that the lines of the matrices $\mathbf{D}_n{}^{\mathrm{T}}(n > 1)$ corespond to n-th differences, with exception of first and last n lines. If both the number of lines and columns in $\mathbf{D}_n{}^{\mathrm{T}}$ are much larger than n, this difference between $\mathbf{D}_n$ and $\mathbf{D}_n{}^{\mathrm{T}}$ can be neglected, e.g. in the continuous case when $\Delta q -> 0$. For example, when the norm of the second derivatives of $s(t,\boldsymbol{x})$ over time is constrained, the following relationship can written for the p-th element of the gradient $\nabla\Psi_t(\boldsymbol{S}^p)$:

$$D_2^{\#}D_2 S^p(t,\boldsymbol{x})|_{t_i} = \left.\frac{\partial^4 S^p(t,\boldsymbol{x})}{\partial t^4}\right|_{t_i} \approx \{\Psi_t(\boldsymbol{S}^p)\}_g =$$
$$= (\Delta t)^{-4}\left(S(t_{i+2},\boldsymbol{x}) - 4S(t_{i+1},\boldsymbol{x}) + 6S(t_i,\boldsymbol{x})\right.$$
$$\left.-4S(t_{i-1},\boldsymbol{x}) + S(t_{i-2},\boldsymbol{x})\right) \quad (49)$$

where the index g can be calculated according Eq. (34) and $2<i<N_t-2$ (see Eq. 21). Equations analogous to Eq. (49) can be written for terms corresponding to the spatial coordinates x, y and z. It should be noted that in practice, the calculation of the first "transport term" and the second "a priori term" in Eq. (47) can be performed rather independently as shown in Fig. 6. For example, in Section 3 (where Eq. (47) is implemented) the "transport term" is integrated with the time step of the GOCART model (20 min), whereas the "a priori term" is calculated as shown in Eq. (49) with time step $\Delta t = t_{i+1} - t_i$ equal to 24 h.

These formulations can also adopt the same a priori constraints as those used in the Kalman filter, i.e. when only the continuity of sources is constrained a priori by assuming $\boldsymbol{S}_t^* = \boldsymbol{S}_{t-1} + \Delta_S$ (see Eqs. 15 and 18). In this situation, one can assume that the first derivative of $s(t,\boldsymbol{x})$ over time is

close to zero, i.e. $\partial s(t,\boldsymbol{x})/\partial t \approx 0$. Similarly, Eq. (47) should be used with only one a priori term, corresponding to the norm of the first derivatives $\partial s(t,\boldsymbol{x})/\partial t$ i.e. $q=t$ and $n=1$. This way Eq. (47) relies on the same a priori constraints as those used in the Kalman filter Eq. (47), and observations $m(t,\boldsymbol{x})$ during the entire time period $t>t_i$ can contribute to the solution $S(t_i,\boldsymbol{x})$, whereas Eq. (15) relies only on observations $m(t_i,\boldsymbol{x})$ at time t_i.

Also, it should be noted that for clarity Eqs. (45c) and (47) were written for the case of the simplest measurements covariance matrices and a priori data errors $\mathbf{C}_{...}=\varepsilon^2_{...}\,\mathbf{I}_{...}$. However, the generalization of these equations to cases when the accuracies within each data set are different ($\{\mathbf{C}\}_{ii}\neq\{\mathbf{C}\}_{jj}$, $i\neq j$), or the covariance matrices are non-diagonal, is rather straightforward (similar to that shown by Eqs. (41–43).

2.7 Inverting models having non-linearities

Previous sections described an approach to inverting a linear transport model (Eq. (5)), provided global aerosol mass $\boldsymbol{M}^*$ measurements are available. In practice, the transport model may be non-linear, and the global aerosol data fields may be available only in the form of satellite optical measurements:

$$f = F(m(t,\boldsymbol{x}),\lambda,\theta;...), \tag{50}$$

where $f(\ldots)$ is generally a non-linear function depending on aerosol mass $m(t,\boldsymbol{x})$, instrument spectral characteristics λ, observation geometry θ, etc. Therefore, the following non-linear equation should be solved instead of Eq. (5):

$$\boldsymbol{F}^* = \boldsymbol{F}\left(\boldsymbol{M}(\boldsymbol{S})\right) + \Delta\boldsymbol{F}, \tag{51}$$

where $\boldsymbol{F}$ and $\Delta\boldsymbol{F}$ are vectors of global optical data and their uncertainties. Since, the steepest descent method can be applied to both linear and non-linear problems, Eqs. (41, 45, 47), that use adjoint operators, can be expanded to solve Eq. (51). For example, for a basic case when only optical measurements $\boldsymbol{F}^*$ are inverted with no a priori constraints, the steepest descent solution can be written as:

$$\hat{\boldsymbol{S}}^{p+1} = \hat{\boldsymbol{S}}^{p} - t_p \Delta\hat{\boldsymbol{S}}^{p}, \tag{52a}$$

$$\Delta\hat{\boldsymbol{S}}^{p} = \nabla\Psi_f(\boldsymbol{S}^p) = \mathbf{K}_p^{\mathrm{T}}\mathbf{C}_f^{-1}\Delta\boldsymbol{F}^p = \mathbf{T}_p^{\mathrm{T}}\mathbf{F}_p^{\mathrm{T}}\mathbf{C}_f^{-1}\Delta\boldsymbol{F}^p, \tag{52b}$$

where $\Delta\mathbf{F}^p=\boldsymbol{F}(\boldsymbol{S}^p)-\boldsymbol{F}^*$. Matrices $\mathbf{K}_p$ $\mathbf{T}_p$ and $\mathbf{F}_p$ denote Jacobi matrices of the first derivatives *df/ds, dm/ds* and *df/dm* calculated in the vicinity of the vector $\boldsymbol{S}^p$:

$$\left\{\mathbf{K}_p\right\}_{ji} = \left.\frac{df_j\left(m(\boldsymbol{S}),\lambda,\theta,...\right)}{dS_i}\right|_{\boldsymbol{S}=\boldsymbol{S}^p}, \tag{53a}$$

and

$$\left\{\mathbf{T}_p\right\}_{j'i} = \left.\frac{dm_{j'}(\boldsymbol{S},t,\boldsymbol{x})}{dS_i}\right|_{\boldsymbol{S}=\boldsymbol{S}^p},$$

$$\left\{\mathbf{F}_p\right\}_{ji'} = \left.\frac{df_j\left(\boldsymbol{M},\lambda,\theta,...\right)}{\partial m_{i'}}\right|_{\boldsymbol{M}^p=\boldsymbol{M}(\boldsymbol{S}^p)}, \tag{53b}$$

where indices j, j', i and i' are indices for the elements $f_j, m_{j'}, m_{i'}, S_i$ of the corresponding vectors $\boldsymbol{F}^{\mathrm{T}}=(f_1,f_2,\ldots)$, $\boldsymbol{M}^{\mathrm{T}}=(m_1,m_2,\ldots)$, and $\boldsymbol{S}^{\mathrm{T}}=(S_1,S_2,\ldots)$. The following relationship between the Jacobi matrices of *df/ds, dm/ds* and the *df/dm* derivatives was used in Eq. (53):

$$\mathbf{K}_p = \mathbf{F}_p\mathbf{T}_p \quad . \tag{53c}$$

The function $\Delta\hat{s}^p(t,\boldsymbol{x})$, corresponding to vector $\Delta\hat{\boldsymbol{S}}^p$, can formulated as follows:

$$\Delta\hat{s}^p(t,\boldsymbol{x}) = \int_{t}^{t_0} T_p^{\#}\left(t',\boldsymbol{x}\right) F_p^{\#}\left(t',\boldsymbol{x}\right) \left(\Delta\hat{s}^p\left(t',\boldsymbol{x}\right) + \Delta f^p\left(t',\boldsymbol{x}\right)\right)(-dt'), \tag{54}$$

where $T_p^{\#}(t,\boldsymbol{x})$ and $F_p^{\#}(t,\boldsymbol{x})$ are adjoint operators for the mass transport $T(s(t,\boldsymbol{x}))$ and the optical model $F(m(t,\boldsymbol{x}))$, and index p indicates that these adjoint operators are equivalents of transposed Jacobi matrices $\mathbf{T}_p^{\mathrm{T}}$ and $\mathbf{F}_p^{\mathrm{T}}$. The derivation of $F_p^{\#}(t,\boldsymbol{x})$ is quite transparent because optical properties $f(m(t,\boldsymbol{x}),\ldots)$ usually are related only to local aerosols, so in practical implementations of Eq. (54) (that are usually performed in discrete representations), $F_p^{\#}(t,\boldsymbol{x})$ can be explicitly replaced by the transposed Jacobi matrix $\mathbf{F}_p^{\mathrm{T}}$.

Equation (54) can be expanded easily to implement a constrainted inversion of $f=F(m(t,\boldsymbol{x}),\lambda,\theta,...)$. For example, when the solution is constrained by a priori limits on the temporal and spatial derivatives of $\hat{S}(t,\boldsymbol{x})$ (utilized in Eqs. 35 and 48), Eq. (54) can be written as follows:

$$\Delta\hat{s}^p(t,\boldsymbol{x}) = \int_{t}^{t_0} T_p^{\#}\left(t',\boldsymbol{x}\right) F_p^{\#}\left(t',\boldsymbol{x}\right)\left(\Delta\hat{s}^p\left(t',\boldsymbol{x}\right) + \Delta f^p\left(t',\boldsymbol{x}\right)\right)(-dt') + \sum_{(q=t,x,y,z)} \gamma_q \frac{\partial q^{(}2n)}{\left(s^p(t,\boldsymbol{x})\right)\partial q^{(}2n)}. \tag{55}$$

Using this $\Delta\hat{s}^p(t,\boldsymbol{x})$, (if the same discrete representation is used), the iterative retrieval would amount to minimizing the following quadratic form:

$$2\,\Psi(\boldsymbol{S}) = 2\Psi_f(\boldsymbol{S}) + 2\sum_{(q=t,x,y,z)} \gamma_q \Psi_q(\boldsymbol{S}) = 2\left(\Delta\boldsymbol{F}\right)^{\mathrm{T}}\Delta\boldsymbol{F} + 2\sum_{(q=t,x,y,z)} \gamma_q \boldsymbol{S}^{\mathrm{T}}\mathbf{D}_{n,q}^{\mathrm{T}}\mathbf{D}_{n,q}\boldsymbol{S}. \tag{56a}$$

This quadratic form can be generalized by the following functional:

$$2\,\Psi' = 2\int_t \iiint_{x,y,z} \Delta f^{\#}(t,x,y,z)\,\Delta f(t,x,y,z)dxdydzdt + 2\sum_{(q=t,x,y,z)} \gamma_q \int_{q_{\min}}^{q_{\max}} \left(\frac{\partial^n s(q,...)}{\partial q^n}\right)^2 dq, \tag{56b}$$

where $\Delta f(t,x,y,z)=f^*-f(t,x,y,z)$ and $\Delta f^{\#}(t,x,y,z)$ denotes the adjoint of $\Delta f(t,x,y,z)$.

Thus, the above derivations show the high potential of using a statistical estimation approach for implementing aerosol transport inverse modeling. For example, it was demonstrated that by following the multi-term LSM strategy, there is flexibility to apply various types of a priori constraints in inverse modeling. For example, Eqs. (46–49) show how constraints on aerosol (or other tracer) emission derivatives with respect to spatial coordinates or time can be included in the adjoint integration of tracer models, that are widely used in variational assimilation for atmospheric tracer source identification (e.g. Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987). Using such constraints in inverse modeling may have high potential because, in principle, a priori limitations on the derivatives of emission variability is a weaker and more flexible way of constraining the solution than assuming a priori values of the emission. Equations (55–56) give the formulation of the approach for inverting satellite observations (e.g. radiances measured by passive satellite observations). This generalization may have the high potential, because using satellite observations directly in inverse modeling and satellite data assimilation has advantages compared to relaying on satellite retrieval products (Weiver et al., 2006). However, using the steepest descent iteration in Eqs. (55–56) makes the application of Eqs. (55–56) less attractive in practice, because generally, inverting radiative transfer equations by the method of steepest descent requires a large number of iterations (Dubovik and King, 2000). Therefore, it might be useful to explore the possibility of adopting iterative strategy of the conjugated gradients method (Appendix C), as this method is known to have superior convergence properties than steepest descent.

3 Application of the inverse methodology for aerosol source retrieval from satellite observations

- First, we consider the inversion of an aerosol transport model, to derive the "unknown input" (aerosol emissions) to the model from the "known output" (aerosol mass distribution). Our inverse algorithm developments are based on the GOCART aerosol transport model.

- Second, we discuss differences between inverting the model output and satellite data and outline the modifications required for applying the model based inverse algorithm to the satellite observations. We consider inversion of the aerosol observations from MODIS.

- Finally, we illustrate the performance of developed algorithm by numerical tests and then we apply the algorithm to the actual MODIS aerosol data.

3.1 Algorithm for inverting the GOCART model

The inversion algorithm (Eqs. 39–41) treats the strength of the aerosol emission at each global location as an unknown. Therefore, in an ideal situation, when the observations provide enough information to retrieve all the emission parameters, the emissions derived from the observations theoretically could replace the original module prescribing the aerosol emissions in the chemical transport model. Such an ideal situation is likely if reliable observations about all the aerosol characteristics are provided by the transport model output, i.e. if these observations are sensitive to all the time and space (4D) aerosol variations provided by model. As we will discuss in the next Section, the real observations are not sensitive to all the aerosol distribution details that can be modeled, therefore successful inversion of model output does not guarantee the successful inversion of real observations. Nonetheless, inverting the detailed model output can be helpful for verifying the performance of different blocks in the inversion algorithm. As a first step in implementing the inverse algorithm, we developed an algorithm to invert GOCART output, and carried out series of numerical tests to verify algorithm performance under highly constrained conditions (the entire output is prescribed) and in a "no error" environment.

The GOCART – Goddard Chemistry Aerosol Radiation and Transport model is described in papers by Chin et al., (2000, 2002) and Ginoux et al. (2001). The model uses the assimilated meteorological data from the Goddard Earth Observing System Data Assimilation System (GEOS DAS) and provides four-dimensional aerosol mass distributions in 20 to 30 atmospheric layers, at a horizontal resolution of 2° latitude by 2.5° longitude. The model calculates aerosol composition and size distribution, optical thickness and radiative forcing. There are seven modules representing atmospheric processes: emission, chemistry, advection, cloud convection, diffusion (boundary layer turbulent mixing), dry deposition, and wet deposition. The model solves the continuity Eq. (1) using an operator-splitting technique (Eqs. 2–3), with a time step of 15 min for advection, convection and diffusion, and 60 min for the other processes.

GOCART provides 4D distributions of about 16 aerosol particle types/size bins: sulfate, hydrophilic and hydrophobic Organic Carbon (OC), hydrophilic and hydrophobic Black Carbon (BC), four size-differentiated sea salt bins, and up to seven dust size-differentiated bins, depending on the GOCART model version (Chin et al., 2002, 2004; Ginoux et al., 2001). The model does not include interactions between different aerosol particles, with the exception of transformations between hydrophilic and hydrophobic components of BC and OC. Therefore, GOCART simulates distributions of sulfates, BC, OC, desert dust and sea salt independently. Calculations for different size bins of dust and sea salt are also independent, and the atmospheric processes sensitive to particle size (e.g. sedimentation) are incorporated accordingly. The same concept can be adopted for inverting the model output, i.e. each aerosol component of the output can be inverted independently with an inverse algorithm that uses an adjoint model tuned to the properties of each aerosol component.

The inverse algorithm described by Eqs. (41, 45, 47, 54 and 55) and illustrated by diagrams in Figs. 5 and 6 was implemented for the GOCART model. The adjoint transport operator $T*_p(t, \boldsymbol{x})$ was developed by redesigning GOCART modules for each atmospheric process. Namely, the adjoint operation for advection was performed with the original GOCART advection algorithm (Lin and Rood, 1996), using sign-reversed wind fields (Vukicevic et al., 2001). The equivalence between such physically derived retro-transport and adjoint equations has been proven rigorously by Hourdin and Talagrand (2006). The adjoints of the local processes were developed by analogy with the corresponding transpose matrix operators. Specifically, cloud convection, diffusion, dry deposition, and wet deposition affect only vertical aerosol motion. All these processes have local character in the sense that for a single time step in the model, they work independently in each horizontally resolved vertical column. Therefore, such processes can be easily modeled via explicit use of matrices of rather small dimension, and the corresponding adjoint operators can be obtained by direct transposition of those matrices. First, we arranged the cloud convection, diffusion, dry deposition, and wet deposition in matrix form, and derived the transpose of those matrices. Then, for achieving faster calculation time, we redesigned the original programs so that their application to a vector provides a product equivalent to the application of the corresponding transpose matrices. Chemical aging transformations of black and organic carbon aerosols require changing only the proportions of different components, and do not induce any vertical or horizontal aerosol motions. Therefore, the adjoints of these chemical processes can be constructed simply by changing the direction of chemical transformation. Once the adjoint model was developed, we implemented several inversions of 4D mass fields simulated by GOCART for different aerosol components. These numerical tests showed that the algorithm retrieves the emissions of all aerosols accurately and no inversion issues that could substantially limit the accuracy of the aerosol source retrievals in such well-constrained situations were revealed. Particular emphasis was placed on testing inversions of BC, OC and dust, since we expect algorithm applications to be focused on observations of these types of aerosol.

3.2 Application of the algorithm for inverting satellite data

Naturally, satellite observations (as well as observations of any other type) do not provide the same aerosol quantities, coverage, and sampling as model simulations (see Fig. 8). Accordingly, the inverse algorithm settings, as well as some aspects of inversion concept, needs to be adjusted when inverting observations. Below we discuss in detail the application of inverse modeling for retriving global aerosol emissions from MODIS data, though many of the aspects considered are relevant to inverse modeling with any other satellite data.

MODIS – the MODerate resolution Imaging Spectroradiometer aboard both NASA's Terra and Aqua satellites, provides near-global daily observations of Earth over a wide spectral range (0.41 to 15.0 μm). These measurements are used to derive spectral aerosol optical thickness and aerosol size parameters over land and ocean (Kaufman et al., 1997; Tanré et al., 1997; Remer et al., 2005). The primary aerosol products avaiable include aerosol optical thickness at three visible wavelengths over land and seven wavelengths over ocean, aerosol effective radius, and fraction of optical thickness attributed to the fine mode. The present study uses the MODIS aerosol optical thickness product aggregated to 1° by 1° spatial resolution. The expected accuracy of MODIS optical thickness is $\Delta\tau = \pm 0.03 \pm 0.05\tau$ over ocean (Tanré et al., 1997; Remer et al., 2005) and $\Delta\tau = \pm 0.05 \pm 0.15\tau$ over land (Kaufman et al., 1997; Remer et al., 2005).

3.2.1 Main issues for applying inverse modeling to the MODIS data

First, to use MODIS observations as input to the inversion, the adjoint formulation must include the conversion from modeled aerosol mass into measured aerosol optical parameters. Accordingly, the operator F_p is rearranged into the adjoint $F_p^{\#}$ and is used in the inversion according to Eqs. (54–55). Since the aerosol optical thickness operator F_p sums the contributions from different layers and aerosol types, its adjoint $F_p^{\#}$ redistributes the total sum to the individual layers and aerosol types.

Second, the MODIS global $\tau(0.55)$ observations reported at 1° by 1° need to be rescaled to the 2° by 2.5° GOCART horizontal resolution.

Third, as mentioned above, satellite data provide less information than global model output. Specifically, a passive, multi-spectral, polar-orbiting, cross-track scanning, single view remote sensor such as MODIS has the following main limitations (Kaufman et al., 1997; Tanré et al., 1996, 1997; Remer et al., 2005).:

- no sensitivity to aerosol vertical distribution;
- global coverage only once in two days, only for cloud-free conditions, and not over bright surfaces such as deserts or in glint regions over water;
- limited capability to identify aerosol type based on coarse/fine mode size discrimination only; no information about particle shape or composition.

Indeed, the top-of-atmosphere radiances are sensitive mainly to the total effective aerosol content in the atmospheric column. But it is worth noting that coarse/fine mode discrimination contains some particle type information, as desert dust and maritime aerosols are dominated by coarse mode particles, whereas biomass burning and urban pollution are dominated by fine mode particles (Dubovik et al., 2002). In summary, the problem of retrieving all the aerosol

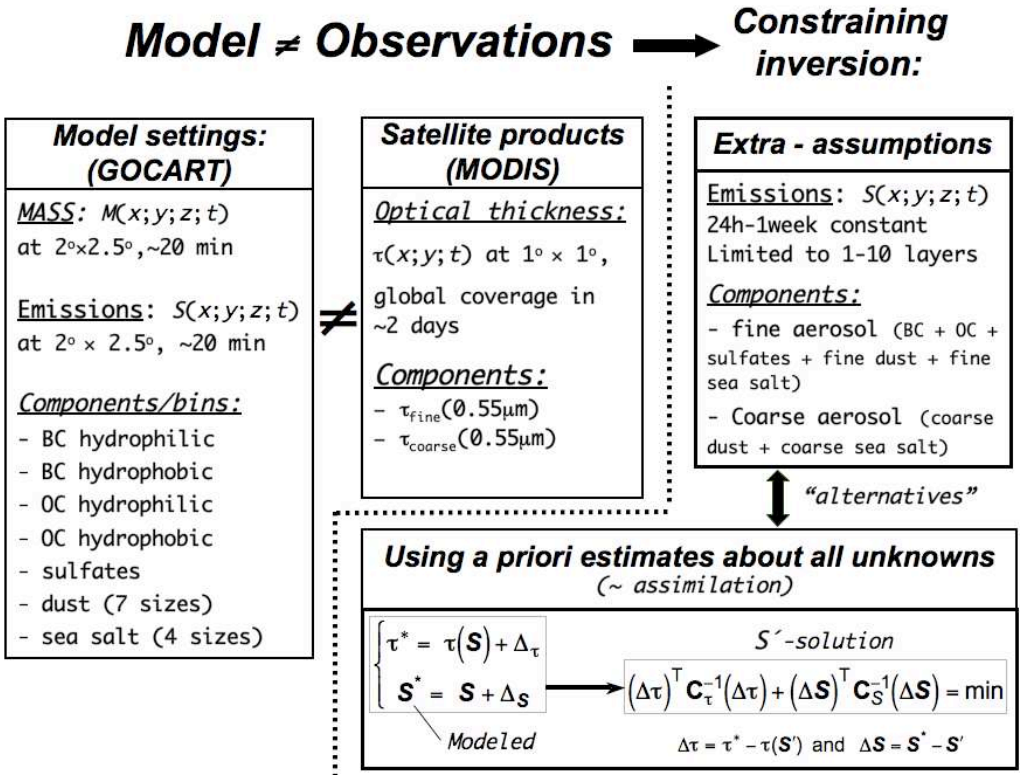

Fig. 8. The diagram summarizing concepts of constraining global emission retrievals from satellite observations.

sources used by GOCART from MODIS observations is ill-posed (see Fig. 8). From a formal viewpoint, the number of observations is significantly smaller than a number of retrieved parameters.

3.2.2 Approaches to constraining aerosol emission retrieval from MODIS data

Constraining inversion by a priori estimates

One way to assure retrieval uniqueness is to use a priori constraints, and search for a solution of equation system Eq. (10) that satisfies both the measurements and the a priori estimates (see illustration on Fig. 8). The solution of Eq. (10) is always unique, since formally, the number of measurements N_m+N_S always exceeds the number of unknowns N_S. The solution can be obtained by the matrix inversion of Eq. (13), or by Eqs. (45), that uses the adjoint transport model. (To invert optical measurements, the adjoint transport integration should be implemented, as in Eq. (54)). For example, following the strategy commonly employed in assimilations techniques, one can use the original GOCART emissions as a priori estimates of unknown emissions $\boldsymbol{S}^*$. With this approach, the a priori emission estimate errors are equivalent to uncertainties in the emission modeling. The potential drawback of this approach is that in some cases, a priori emission estimates can over-restrict the freedom of the algorithm to search for the actual solution. For example, if the values of modeled emissions $\boldsymbol{S}^*$ in certain locations depart from real emissions $\boldsymbol{S}'$ much more than expected (i.e. if the errors of $\boldsymbol{S}^*$ are underestimated), then the solution given by Eq. (13) or Eqs. (45) will not provide a solution having the required departure at that location from $\boldsymbol{S}^*$ toward $\boldsymbol{S}'$, which will cause a bias in the results. The appropriate choice of corresponding Lagrange multipliers in Eq. (45) (or covariance matrices of a priori estimates in Eq. (13)) is supposed to prevent the appearance of solution biases caused by incorrect a priori information. However, the optimum choice of Lagrange multipliers (or covariance matrices of a priori estimates) is nontrivial, and having some criteria that indicate problems in the assumed $\boldsymbol{S}^*$ uncertainties (as well as in all other assumptions) would be highly desirable. The value of the quadratic form 2Ψ obtained by the minimization could be used as such an indicator. Specifically if all assumptions are correct, the values of the form 2Ψ given by Eq. (29) follow a χ^2 distribution with "$m-n$" degrees of freedom, the mean minimum is (see detailed discussion by Dubovik, 2004):

$$\left\langle \left(2\Psi\left(\hat{\boldsymbol{a}}\right)\right)_{\min}\right\rangle = \sum_{k=1}^{K}\left(\Delta \boldsymbol{f}_k\left(\hat{\boldsymbol{a}}\right)\right)^{\mathrm{T}}\left(\mathbf{C}_k\right)^{-1}\left(\Delta \boldsymbol{f}_k\left(\hat{\boldsymbol{a}}\right)\right)$$
$$-\left(\Delta\hat{\boldsymbol{a}}\right)^{\mathrm{T}}\left(\mathbf{C}_{\hat{\boldsymbol{a}}}\right)^{-1}\left(\Delta\hat{\boldsymbol{a}}\right) = \sum_{k=1}^{K} N_k - N_a, \qquad (57a)$$

where N_k = rank($\mathbf{C}_k$) and N_a = rank($\mathbf{C}_{\hat{a}}$). Similarly, for $2\Psi'$ formulated for the joint system Eq. (10) formalized in the same manner as Eq. (29b), one can write:

$$\left\langle \left(2\Psi'\left(\hat{\boldsymbol{S}}\right)\right)_{\min}\right\rangle = \sum_{k=1}^{2}\gamma_k^2(\Delta \boldsymbol{f}_k)^{\mathrm{T}}(\mathbf{W}_k)^{-1}(\Delta \boldsymbol{f}_k) - \left(\Delta\hat{\boldsymbol{S}}\right)^{\mathrm{T}}\left(\mathbf{W}_{\hat{\boldsymbol{S}}}\right)^{-1}\left(\Delta\hat{\boldsymbol{S}}\right) = \left(N_1+N_2-N_{\hat{\boldsymbol{S}}}\right)\varepsilon_1^2. \quad (57b)$$

Here we in addition assumed statistical independence of measurements $\boldsymbol{f}_1^*$ and a priori estimates $\boldsymbol{S}^*$, i.e. N_1 = rank($\mathbf{W}_1$), N_2 = rank($\mathbf{W}_2$) and $N_{\hat{\boldsymbol{S}}}$ = rank($\mathbf{W}_{\hat{\boldsymbol{S}}}$)

Thus, for the retrieval of emissions from MODIS measurements when a priori estimates are used, the following minimum of $2\Psi'$ can be expected:

$$\left\langle \left(2\Psi'\left(\hat{\boldsymbol{S}}\right)\right)_{\min}\right\rangle = \sum_{j=1}^{N_{\mathrm{obs}}}\left(\tau_j^* - \tau_j\left(\hat{\boldsymbol{S}}\right)\right)^2 + \gamma_{S_*^2}\sum_{j=1}^{N_{S^*}}\left(\boldsymbol{S}_j^* - \hat{\boldsymbol{S}}_j\right)^2 \approx N_\tau\varepsilon_\tau^2. \quad (58)$$

Here we assumed the simplest covariance matrices for uncertainties in MODIS measurements τ_j^* and a priori estimates $\boldsymbol{S}^*$: $\mathbf{C}_\tau = \mathbf{I}\varepsilon_\tau^2$ and $\mathbf{C}_{S^*} = \mathbf{I}\varepsilon_{S^*}^2$. Correspondingly, $\gamma_{S^*}^2 = \varepsilon_{S^*}^2/\varepsilon_\tau^2$.

Thus, Eq. (58) relates the value of the derived minimum $\left(2\Psi'\right)_{\min}$ with expected error variance ε_τ^2 of the measurements τ_j^*. Any significant difference between the derived $\left(2\Psi'\right)_{\min}$ and $N_\tau\varepsilon_\tau^2$ indicates inconsistency in some assumptions. Such inconsistencies could include inadequate prescription of the magnitudes and shapes of random error distributions in the measurements and a priori estimates. They also can indicate the existence of systematic inconsistencies (biases) between the model and observations that exceed the random error magnitude.

The estimate of ε_τ^2 in Eq.(58) is written as the product of N_{obs} measurement residuals $\left(\tau_j^* - \tau_j^*\left(\hat{\boldsymbol{S}}\right)\right)^2$ and $N_S{}^*$ residuals of fitting a priori estimates $\left(\boldsymbol{S}_j^* - \hat{\boldsymbol{S}}_j\right)^2$. Accordingly, if $N_{\mathrm{obs}} \ll N_{S^*}$, then the a priori residuals dominate the value of the minimized form $2\Psi'$, and the sensitivity of the Eq. (58) criterion to the measurement fitting accuracy becomes weak. Therefore, if we try to retrieve aerosol emissions from MODIS data (average τ_{fine} (550 nm) and τ_{coarse} (550 nm) for each atmospheric column horizontally resolved by the GOCART model) in exactly the same format as is assumed in GOCART model (i.e. 16 aerosol particle types/size with possible sources at different layers and allowing hourly emission variability), N_{obs} exceeds N_S by a factor of about a hundred or larger. So the residual value shown by Eq. (58) is dominated by the fitting of a priori estimates, And the retrieval is too dependent on the a priori assumptions. From a practical point of view, it is difficult to ensure appropriate sensitivity of the solution $\hat{\boldsymbol{S}}$ to the MODIS measurements.

Simplifying the forward model

The goal of our study is to develop a global emission retrieval that is highly sensitive to the satellite observations and minimally dependent on a priori assumptions. It is particularly appealing to explore the performance of an unsupervised retrieval that distributes the global aerosol emission based only on satellite observations and transport. Therefore, instead of correcting a large number of parameters describing aerosol emissions in GOCART with a relatively small number of MODIS observations, we consider constraining the retrieval by employing a simplified GOCART model, having a reduced number of parameters to be retrieved. We adopted the following simplifications (see illustration in Fig. 8).

- First, MODIS sees each location on the globe no more than once in 24 h (except at high latitudes). Therefore, to constrain the retrieval, the emission variability over 24 h was neglected.
- Second, MODIS characterizes each atmospheric column by only two parameters (τ_{fine} (550 nm) and τ_{coarse} (550 nm)). Therefore, we reduce the aerosol particle type/size discrimination to only two mono-size particle bins, one for fine mode and one for coarse mode aerosol.
- Third, we assume that all aerosol sources are located in the surface level. Thus, we have reduced the number of unknowns to only two parameters for each horizontally resolved atmospheric column, so the number of unknown N_S is comparable to the number of MODIS measurement N_{obs}.
- Fourth, the measurement set becomes well determined ($N_{\mathrm{obs}} > N_S$) if we further reinforce the constraints on emission variability, by assuming constant emission during several days. However, taking into account that typical aerosol lifetime is about a week, one might expect that assuming constant aerosol emission over 24 h might be sufficient for constraining the retrieval. Indeed, subsequent MODIS observations of aerosol diffusion in the region surrounding an aerosol source should be sufficient to constrain the location and magnitude of aerosol emissions from this source that occurred at the beginning of the week. It is important to understand that any simplifications of the forward model used in the retrieval algorithm can be justified only if those simplifications do not induce systematic errors (biases) exceeding the measurement uncertainty. This condition can be verified by numerical tests.

The possibility of non-unique solutions

In addition, it should be noted that our algorithm is based on steepest descent iterations (see Eqs. 39–41), so it will not collapse even for an ill-posed problem. Under these conditions, the non-unique solution would depend strongly on the

initial guess (e.g. Dubovik, 2004) in the sense that the algorithm would simply correct the emission initial guess as needed to improve agreement between the model and observations. Similarly, in some situations such non-unique solutions can still provide useful information, especially when only an identifiable sub-group of the retrieve parameters is non-unique, and the remaining sub-space of the retrieved solution is stable and independent of the initial guess. For example, in subsequent Sections we will consider the retrieval of emissions into several vertical layers. Since MODIS data do not have sensitivity to aerosol vertical distribution, MODIS data cannot appropriately constrain the retrieval of emission vertical variability.

Nevertheless, it is sometimes possible for the algorithm to place emissions in higher atmospheric layers when it helps to fit the measurements and, therefore, may avoid producing some strong biases. For example, if the aerosol is really emitted into higher altitudes (e.g. smoke from forest fires) and is transported according to the distribution of winds at higher altitudes, then modeling the transport of such an aerosol event may not be satisfactory if the emissions in the model are restricted to the lowest atmospheric layer. On the other hand, in other situations, non-stability of the retrieved aerosol emission vertical distribution may not be a serious issue, for example, if only the vertical distribution of the emission is uncertain, whereas the total emission into atmospheric column is stable. This retrieved emission field can be satisfactory, taking into account that the emissions vertical distribution is often not critical, since strong vertical mixing occurs in the planetary boundary layer.

To understand the potential of our approach for retrieving aerosol sources from MODIS observations, we will examine the impact of MODIS data limitations, test the effect of the proposed simplifications on modeling accuracy, and test retrieval performance for different scenarios: retrievals constrained by a priori data and/or reduced aerosol emission parameterizations, and unconstrained retrievals dependent on the initial guess.

3.3 Inverse algorithm testing

A series of numerical tests was performed to verify and illustrate how the algorithm inverts the modeled data in a "no error" environment, i.e., when the inverted aerosol fields are fully consistent with the model, and neither measurements nor model errors are present. We also tested the retrieval sensitivity to the presence of random measurement noise.

We tested the possibility of retrieving emissions from MODIS-like remote sensing data that lack information about aerosol vertical distribution and horizontal diurnal variability of aerosol amount and type. These tests are also used for evaluating the effects of reducing the number of aerosol types and limiting the temporal and spatial emission variability, adopted as a simplification of the GOCART model for constraining emission retrievals.

3.3.1 Sensitivity of the retrieval to columnar aerosol properties (no information on vertical aerosol distribution)

The algorithm was applied to the global "synthetic measurements" of $\tau(0.55)$ – optical thickness at $0.55\,\mu$m for a single aerosol type. Two scenarios were employed for restricting emission variability: (i) the emissions are constant over 24 h; (ii) the emissions are constant over one week. The tests were performed for the same time period as the one chosen for inverting actual observations, i.e. the meteorological fields corresponding to the last two weeks of August 2000 were used for these tests. For simplicity the aerosol was assumed in the test to be a single, BC component, and the "synthetic measurements" were simulated as BC optical thickness $\tau_{BC}(0.55)$. Aerosol chemical transformations included in GOCART model were neglected. The test aerosol emissions where assumed equal to the total of BC and OC emissions used in the GOCART model for the same two weeks. The optical thickness was modeled based on the GOCART total atmospheric column aerosol mass, by adopting a density of $1\,g/cm^3$), deriving the aerosol volume and assuming that the aerosol has the same optical properties as fine mode smoke from Zambian savanna burning (Dubovik et al., 2002). For constraining aerosol emission vertical variability, two scenarios were used: (i) aerosol sources were restricted to the near-surface layer; (ii) aerosol sources were allowed within the 10 lowest aerosol layers (i.e. approximately below 2 km, as suggested by recent analysis of lidar observations of biomass burning emissions by Labonne et al., 2007). For constraining horizontal variability of emissions we tested two scenarios: (i) aerosol sources were restricted to the land surface; (ii) aerosol sources were allowed everywhere over land and ocean. The retrieval was initialized by in the tests as follows: (i) initial guess for emission estimates was set to "zero emissions" (no sources). We chose this setting because if retrieval is non-unique in some situations, using zero emissions should not create any false emission. With this initialization, the algorithm is expected to create sources only where they are required to fit the observations. The same retrieval initialization is used in all subsequent tests, unless specified otherwise.

After conducting a large series of the tests, we selected the following algorithm setting for use in follow-up tests and applications to actual data:

1. The emissions are constant over 24 h. The retrieval with this assumption provided a better fit to $\tau(0.55)$ than when emissions were constrained over one week. Also, this assumption did not produce any false weekly variability in the retrieved emissions that we could detect.

2. The aerosol sources are allowed within the 10 lowest aerosol layers. The test results were not substantially different when the retrievals constrained to the lowest or the 10 lowest atmospheric layers. Nevertheless, we favor the second scenario, because it provides a slightly

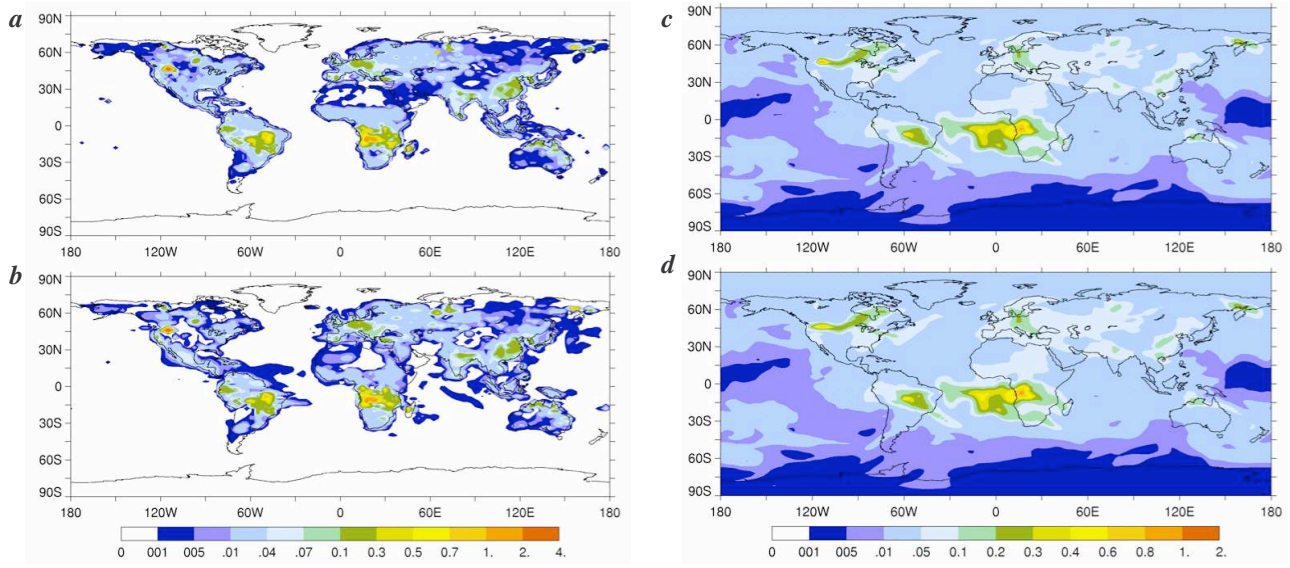

Fig. 9. Inversion test for retrieving daily emissions of fine-mode BC component from two weeks of simulated observations (August 18-30, 2000). Results are shown here only for one day, August 28, 2000. **(a)** Assumed BC aerosol source distribution (emissions units: 10^7 kg/day) **(b)** Retrieved BC aerosol source distribution. **(c)** Initial global BC aerosol optical depth distribution. **(d)** Global BC aerosol optical depth distribution simulated from retrieved sources. Optical depth is given as 0.55 micron total column values.

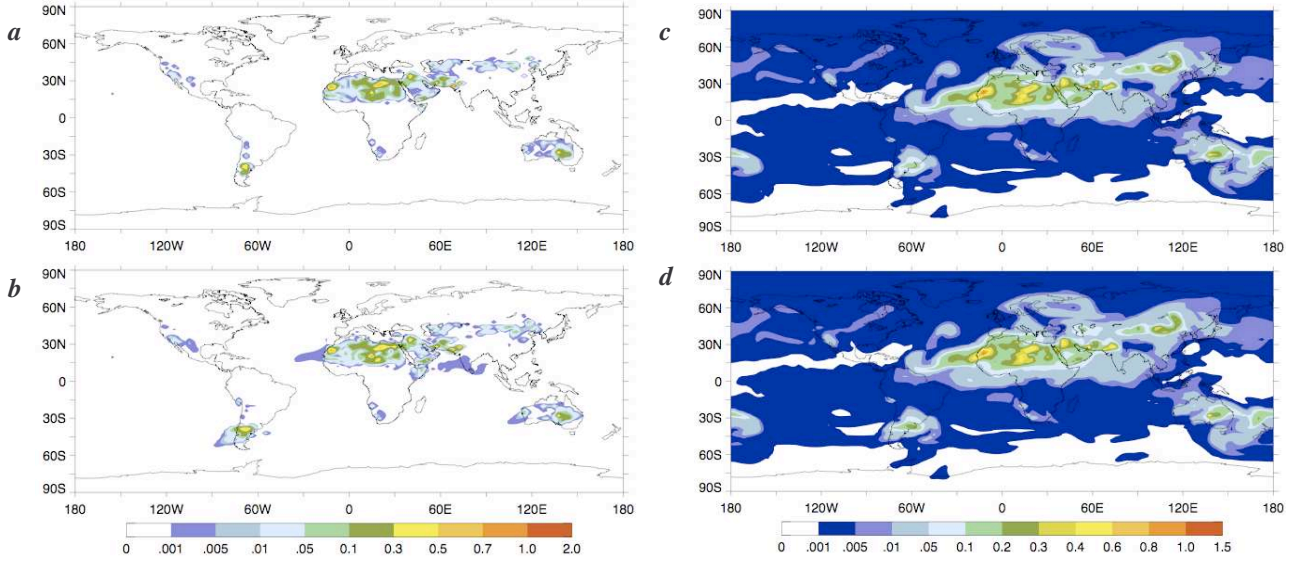

Fig. 10. Inversion test for retrieving daily emissions of a coarse-mode dust component from two weeks of simulated observations (August 20-28, 2000). Results are shown here only for one day, August 28, 2000. **(a)** Assumed dust aerosol sources (emissions units: 10^8 kg/day). **(b)** Retrieved dust aerosol source distribution. **(c)** Initial global dust aerosol optical depth distribution. **(d)** Global dust aerosol optical depth distribution simulated from retrieved sources. Optical depth is given as 0.55 micron total column values.

better fit to the $\tau(0.55)$ field, and also because this assumption provides additional freedom in vertical distribution of emissions without any negative impact on the horizontal distribution of sources.

3. The aerosol sources are allowed everywhere over land and ocean. All our tests showed that this assumption does not produce significant false sources over ocean when "synthetic measurements" were simulated using emissions over land only. On the other hand, the assumption seems to be more practical, since some types of aerosols have sources over ocean.

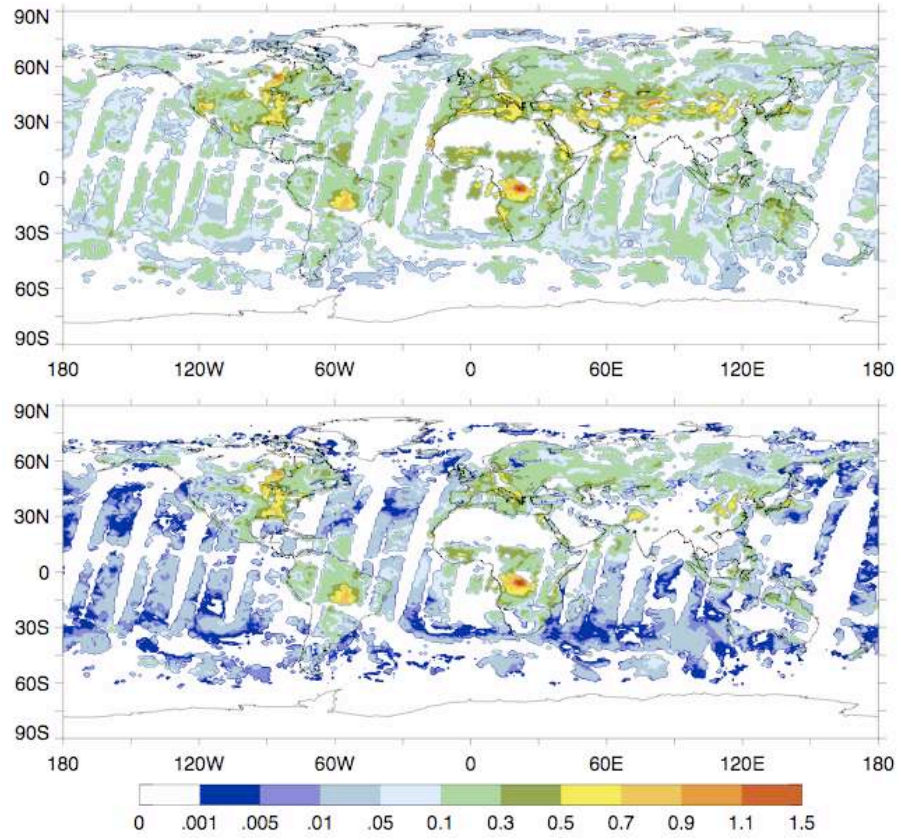

Fig. 11. Actual satellite sampling for August 28, 2000. (a) Global MODIS measurements of total optical depth. (b) MODIS-observed fine-mode optical depth.

Figure 9a–d illustrates the retrieval results for one specific day, 28 August. Figure 9a–b shows that the total emission produced in the first 10 layers agrees well with the "assumed" emissions. Figure 9c–d shows that this retrieved emission results in $\tau(0.55)$ fitting accurately the "synthetic measurements". Since the retrieval uses zero emissions as the emission estimate initial guess, the agreement is especially encouraging. There are minor differences between the prescribed and retrieved emissions in Fig. 9a–b, such as the appearance of minor aerosol sources over the ocean. Because the forward and backward aerosol mass and optical thickness were simulated under the same assumptions without adding any model error or other perturbations, these differences can only be explained by an insufficient number of retrieval iterations or by the effect of numerical errors and instabilities inherent in transport model simulations (Vukicevic et al., 2001).

The accuracy of optical thickness fitting for the entire period of the test is was characterized by two residual values:

$$\sigma_{\text{abs}} = \sqrt{\frac{1}{N_i} \sum_{(j=1,\ldots,N_j)} \left(\tau_i^* - \tau_i(\hat{\boldsymbol{S}})\right)^2}, \tag{59a}$$

$$\sigma_{\text{rel}} = 100 \sqrt{\frac{1}{N_k} \sum_{(k=1,\ldots,N_k)} \left(\frac{\tau_k^* - \tau_k(\hat{\boldsymbol{S}})}{\tau_k^*}\right)^2}, \quad (\tau_k^* \geq 0.05), \tag{59b}$$

where the absolute standard deviation σ_{abs} was simulated using all locations and times and the relative standard deviation σ_{rel} was simulated using only points where $\tau_{\text{BC}}(0.55)$ was not smaller than 0.05. The quantity σ_{abs} is relevant to the minimized quadratic form. Specifically, assuming the measurement covariance matrix is $\mathbf{C}_{\ldots}=\varepsilon_{\ldots}^2\ \mathbf{I}$, σ_{abs} corresponds to a first term in the quadratic form given by Eq. (58). This assumption was used in our inversion tests. The value σ_{rel} is introduced to characterize the accuracy with which aerosol events having high loading are fit. After 40 iterations the residuals were $\sigma_{\text{abs}}\approx 0.005$ and $\sigma_{\text{rel}}\approx 9\%$, i.e. the fitting accuracy achieved is below the expected MODIS measurement accuracy ($\Delta\tau=\pm 0.03\pm 0.05\tau$ over ocean and $\Delta\tau=\pm 0.05\pm 0.15\tau$ over land). It should be noted that Fig. 9a–b shows the total mass emitted into the first 10 atmospheric layers. Unfortunately, the exact vertical structure of the emissions was not reproduced well, because the observations of $\tau_{\text{BC}}(0.55)$ do not provide any vertical information.

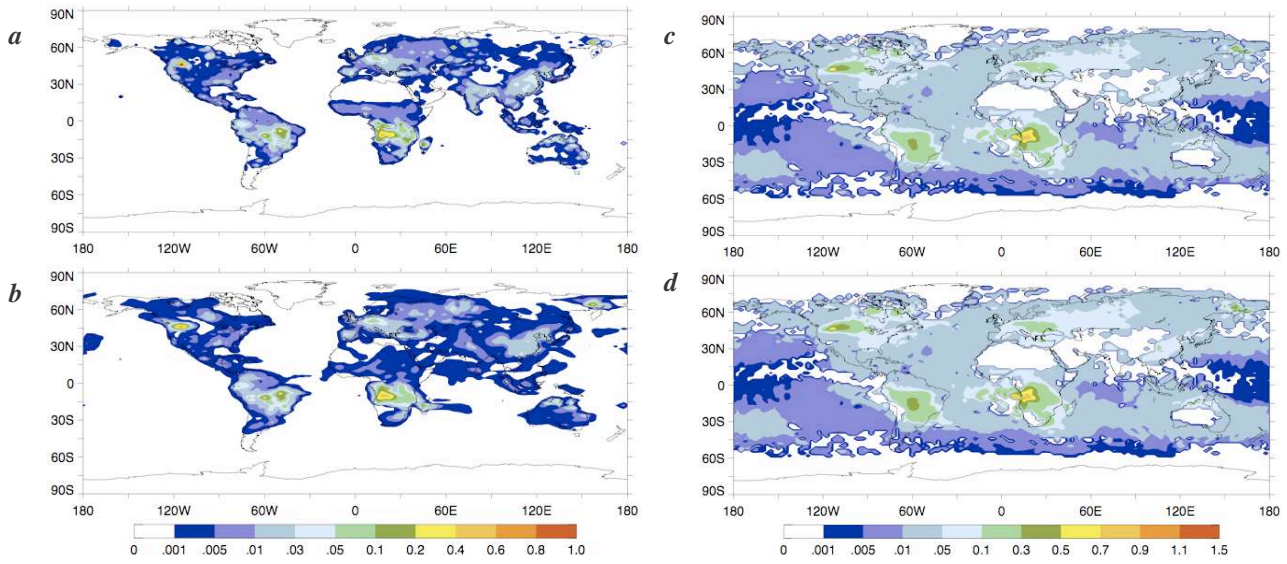

Fig. 12. Inversion test for retrieving daily emissions of a fine-mode BC component, using actual MODIS data sampling for August 18-30, 2000. Results are shown here for the two week average of retrieved daily emissions. (a) Two week average of assumed aerosol source distribution from GOCART model (emissions units: 10^7 kg/day). (b) Two week average of the retrieved aerosol source distribution for the same period. (c) Two week average of initial daily distributions of global aerosol optical depth. (d) Two week average of daily distributions of global aerosol optical depth simulated from retrieved sources.

Figure 10a–d shows the corresponding results for the coarse mode aerosol test, where desert dust emissions were retrieved from simulated measurements of the desert dust optical thickness $\tau_{\text{Dust}}(0.55)$. For simplicity, instead of seven size-resolved components usually used in GOCART model, desert dust was treated here as a single aerosol component with particle radius 2.3 μm. This size corresponds to the median size of the coarse mode climatological size distribution model derived by Dubovik et al. (2002) from multi-year AERONET observations of desert dust in Saudi Arabia. This optical model was also used for calculations of optical depth from the aerosol mass provided by the GOCART model. The test was conducted using the same meteorology and desert dust emission values assumed in the GOCART model for the two weeks in August 2000. The fitting residuals achieved after 40 iterations for this test were: $\sigma_{\text{abs}} \approx 0.01$ and $\sigma_{\text{rel}} \approx 15\%$. These values are slightly higher that those obtained for the BC aerosol test. This is can be explained by the fact the initial residuals were higher, since we used the same zero emissions initial guess, but the magnitude of the prescribed emissions for desert dust is higher than for BC (compare Figs. 9a and 10a). Some differences in the retrieval convergence can also be attributed to differences in aerosol removal processes, for example gravitational settling is higher for larger particles.

In a contrast to the BC and other fine mode aerosols that can be emitted within the boundary layer but above the surface, an additional constraint on the retrieval is that coarse mode desert dust and sea salt aerosol are emitted only into the surface level.

3.3.2 Retrieval sensitivity to input data time and space sub-sampling

In the tests illustrated by Figs. 9–10, the synthetic observations were available at each global location, although as mentioned earlier, this is not the case for actual MODIS data. The observations for any single day have gaps; Fig. 11 shows $\tau_{\text{fine}}(0.55)$ provided by MODIS for August 28, (which can be compared to Fig. 9c). To analyze the effects of gaps in the MODIS data on the global emission retrieval, another test was performed, where the synthetic measurements of $\tau(0.55)$ where sub-sampled in exactly the same way as the real MODIS data for same observation period.

Figure 12a–d illustrates the results of this test for the BC emissions retrieval. The test is analogous to the one shown on Fig. 9a–d, the only difference being that the $\tau_{\text{BC}}(0.55)$ used in Fig. 12a–d had exactly the same coverage as the actual MODIS observations collected during the same time period. The convergence of the retrieval process was slightly slower than for the test where $\tau_{\text{BC}}(0.55)$ had no gaps. After 40 iterations, the fitting residuals were $\sigma_{\text{abs}} \approx 0.009$ and $\sigma_{\text{rel}} \approx 12\%$. These numbers are still lower than the expected MODIS measurements accuracy and, as can be seen from Fig. 12c–d, the fitted $\tau_{\text{BC}}(0.55)$ reproduces the "synthetic measurements" of $\tau_{\text{BC}}(0.55)$ rather well. The retrieved emissions shown in Fig. 12b are also in good agreement with the assumed emission. However, the agreement is slightly poorer than the emission retrieval from $\tau_{\text{BC}}(0.55)$ with no gaps, illustrated in Fig. 9.

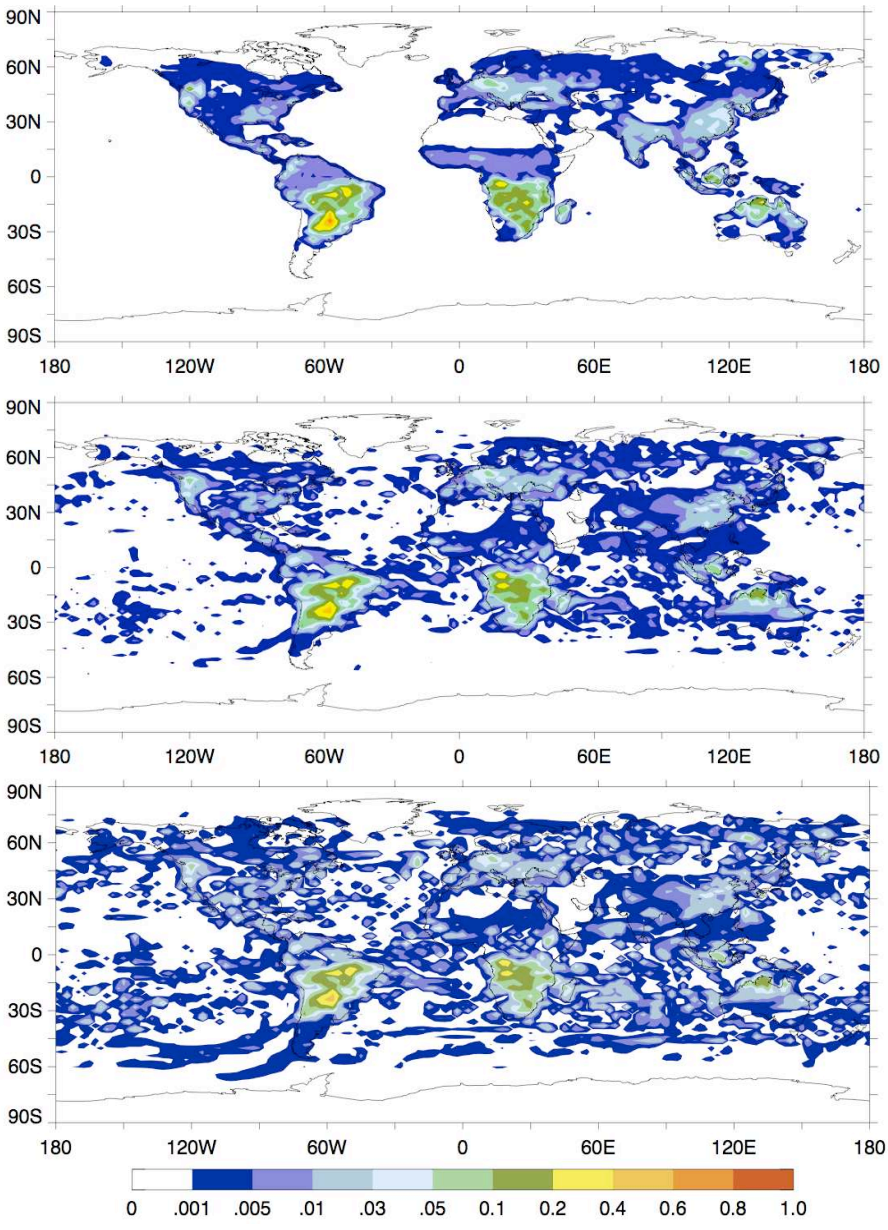

Fig. 13. Inversion test for retrieving daily emissions of a fine-mode BC component, using actual satellite sampling during period August 18-30, 2000, and perturbed by random noise. Results are shown here for the two week average of retrieved daily emissions. (a) Two week average of assumed daily aerosol source distributions assumed in the GOCART model. (b) Two week average of retrieved daily aerosol source distributions, when the variance of added random noise was $\sigma = 0.02 + 0.04\ \tau$. (c) Two week average of retrieved aerosol source distribution, when the variance of added random noise was $\sigma = 0.03 + 0.05\ \tau$ over ocean and $\sigma = 0.05 + 0.15\ \tau$ over land (emissions units: 10^7 kg/day).

Figure 13 shows the effect on the retrievals of perturbing the measurements with random noise. Comparing Fig. 13 with Figs. 9 and 12, one can see that the main pattern of the BC emissions distribution is recovered in the presence of significant (Fig. 13, middle panel) and even high (Fig. 13, lower panel) random noise. The retrieval is stable for major emission hotspots. However, minor sources are strongly contaminated by false emissions appearing (e.g. over ocean) as result of the propagation of random measurement error in the retrieval results. Fitting accuracy was also significantly affected by random noise: after 40 iterations, fitting residuals were $\sigma_{abs} \approx 0.024$ and $\sigma_{rel} \approx 36\%$ for the results shown in the middle panel and $\sigma_{abs} \approx 0.038$ and $\sigma_{rel} \approx 58\%$ for the results shown in the lower panel. This demonstrates that

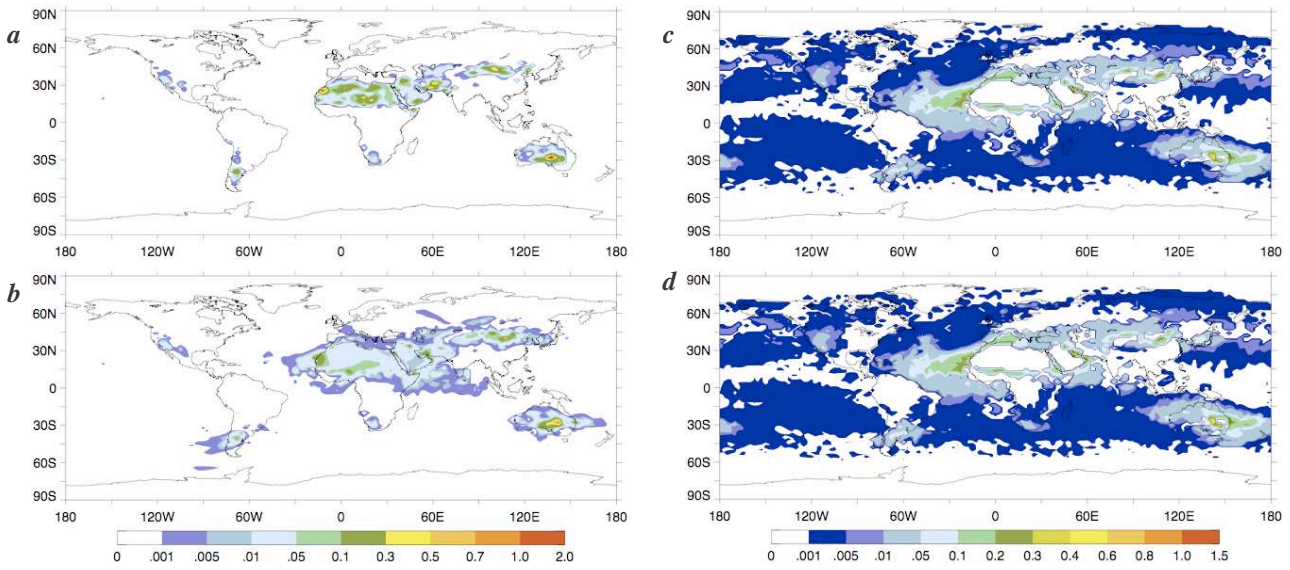

Fig. 14. Inversion test for retrieving daily emissions of a fine-mode BC component from two weeks of simulated observations (August 18-30, 2000). Results are shown here only for one day, August 28, 2000. (a) Assumed BC aerosol source distribution (emissions units: 10^7 kg/day). (b) Retrieved BC aerosol source distribution. (c) Initial global BC aerosol optical depth distribution. (d) Global BC aerosol optical depth distribution simulated from retrieved sources. Optical depth is given as 0.55 micron total column values.

the data contain sufficient redundancy to identify the random noise contribution to the observations. It should be noted that following the discussion in Sect. 3.2.2, the fitting residuals should be expected to be close to zero even in the presence of random noise because in our test, N_{obs} is equal to N_S (or is even smaller, in tests with the gaps). However, the both σ_{abs} and σ_{rel} are always noticeably larger than zero (even in "no error" tests). This can be explained by the fact that we allow only limited number of iterations and, therefore we do not reach the actual minimum of the residual. The significant increase of the σ_{abs} and σ_{rel} in the test with random noise indicates there is some redundancy in the data that does allow exact fitting of noisy data, as expected if $N_{obs}=N_S$. Since the model distributes the emitted aerosol using predetermined meteorological fields and processes, it cannot reproduce arbitrary distribution of aerosols.

The desert dust emission retrieval, from $\tau_{dust}(0.55)$ subsampled according to the MODIS observations, was less successful than BC retrieval. Figure 14a–b shows that the retrieval did not adequately reproduce the assumed "test desert dust" emissions. For example, some patterns of strong desert dust emissions in the Western Sahara do not appear correctly in the retrieved emission field. In this case, the $\tau_{dust}(0.55)$ observations do not sufficiently constrain the retrieval over the areas where the strongest desert dust sources are expected (Ginoux et al., 2001), because MODIS does not provide aerosol retrievals over bright desert surfaces (Kaufman et al. 1997, Remer et al., 2005; Hsu et al., 2004). As a result, the prescribed and retrieved desert dust emissions produce virtually identical $\tau_{dust}(0.55)$ (see Fig. 14c–d), despite some false desert dust source locations in the retrieval such as over the Atlantic Ocean near the western Africa coast). The values of the fitting residuals were $\sigma_{abs}\approx 0.006$ and $\sigma_{rel}\approx 12\%$. It should be noted that these values are actually lower than the residuals obtained in the tests with data that was not subsampled. The explanation is that the sub-sampled data set for the coarse aerosol is not sufficiently redundant, and therefore can be well fit with a coarse aerosol source distribution different from the real one. In another words, the solution is non-unique in this situation since $N_{obs} \ll N_S$.

3.3.3 Retrieval sensitivity to aerosol type

As explained earlier, to constrain the global emission retrieval with imperfect data, we simplified the GOCART model to include only two particle types: fine and coarse mode aerosols. Several numerical tests were performed to evaluate the consequences of this limitation.

We focus on inverting $\tau_{fine}(0.55)$, because fine mode aerosol covers a range of aerosol types. GOCART assigns sulfate, OC and BC aerosols exclusively to the fine mode, along with minor fractions of desert dust and sea salt. Figure 15a–d illustrates the results of the test where modeled $\tau_{fine}(0.55)$ was set to the sum of the optical thicknesses of sulfate, BC and OC aerosols simulated using original GOCART model. Then the modeled $\tau_{fine}(0.55)$ were inverted assuming a single fine mode aerosol. As before, aerosol mass was converted to optical thickness assuming a generic BC component, and we used the same aerosol removal process parameterization as GOCART uses for BC. Note that the

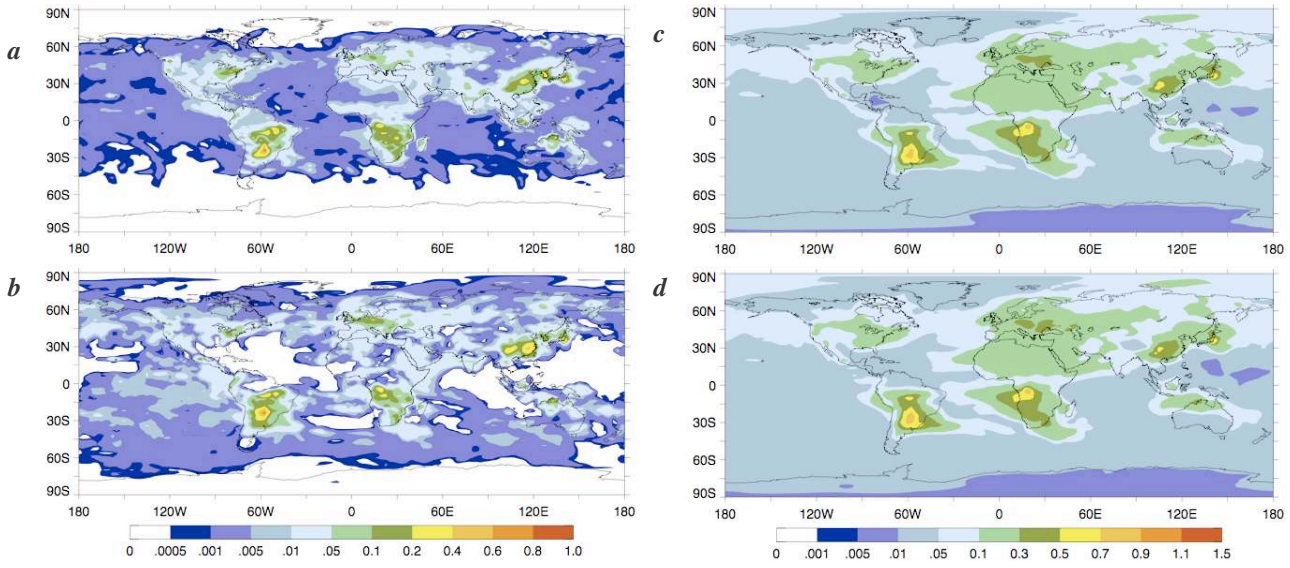

Fig. 15. Inversion test for for retrieving daily emissions of a three-component fine-mode aerosol distribution using a single, fine-mode aerosol model. Results are shown here for the two-week average of retrieved daily emissions (August 18-30, 2000). (a) Two week average of the assumed three-component fine-mode aerosol source distribution, from GOCART model simulations (emissions units: 10^7 kg/day). (b) Two week average of the retrieved single-component fine-mode aerosol source distribution for the same period. (c) Two week average of initial daily global aerosol optical depth distribution of three-component fine-mode aerosol $\tau(0.55)=\tau_{BC}(0.55)+\tau_{OC}(0.55)+\tau_{sulfates}(0.55)$. (d) Two week average daily global aerosol optical depth distribution, simulated from retrieved daily sources of the retrieved single-component fine-mode aerosol.

data were not sub-sampled for these tests, because such data have higher redundancy, making the inversion more sensitive to model simplification. As can be seen in Fig. 15a–d, the results of using a simplified single-fine-mode aerosol look encouraging. For example, the retrieved fine mode aerosol source retains all the major features of the original BC, OC and sulfate sources (Fig. 15a–b). Because sulfate is formed in the atmosphere via photochemical oxidation of its precursor gases, the sulfate source is not confined to land. The differences between the assumed and retrieved source magnitudes and shapes arise because the algorithm attempted to reproduce the combination of BC, OC and sulfates with optical properties for a single BC aerosol component. Nevertheless, as Fig. 15c–d shows, the simplified aerosol model performs well; the fitting errors for this test are $\sigma_{abs}\approx 0.02$ (and $\sigma_{rel}\approx 25\%$).

These values are higher than in previous tests, but still below the expected accuracy of aerosol optical thickness provided by MODIS retrievals. This suggests that even a simplified transport model using a single fine mode aerosol with no further aerosol type discrimination can reproduce the global observations of $\tau_{fine}(0.55)$ at the accuracy level of the MODIS observations. Note that the calculation of residuals (as well as the entire retrieval) was based on the distribution of τ_{fine} only, and the presence of a significant coarse mode fraction could not mask the possible difficulties in fitting τ_{fine}.

Another test (not illustrated here) was performed for evaluating the behavior of under-constrained retrievals using a version of the algorithm that includes all the GOCART chemical processes and retrieves emissions for hydrophilic and hydrophobic BC and OC aerosols simultaneously. As stated above, such retrievals represent an alternative to using the simplified single component aerosol model. However, an under-constrained retrieval does not yield a unique solution, so the value of such retrievals is limited. The numerical test showed that if the "synthetic measurements" of $\tau_{fine}(0.55)$ were composed of BC and OC only, this retrieval provides a better fit than the retrieval based on a single fine mode aerosol. The fitting errors were: $\sigma_{abs}\approx 0.005$ and $\sigma_{rel}\approx 15\%$, whereas the retrieval with a single fine mode aerosol produced fitting errors $\sigma_{abs}\approx 0.01$ and $\sigma_{rel}\approx 20\%$. However, as expected, the partitioning between BC and OC emissions was strongly dependent on the initial guess, so this approach is not clearly preferred to the retrieval based on the single component aerosol. Thus, these results show that if discrimination by particle type is required, the retrieval must include substantial a priori constraints in addition to $\tau_{fine}(0.55)$ measurements.

Retrieving coarse mode aerosol emissions is simpler, as, in the GOCART model, only desert dust and sea salt aerosol types are involved, and dust is emitted exclusively over land whereas sea salt is emitted over water. These properties can be used as natural constraints. Tests show that sea salt and desert dust emissions can be reasonably well discriminated

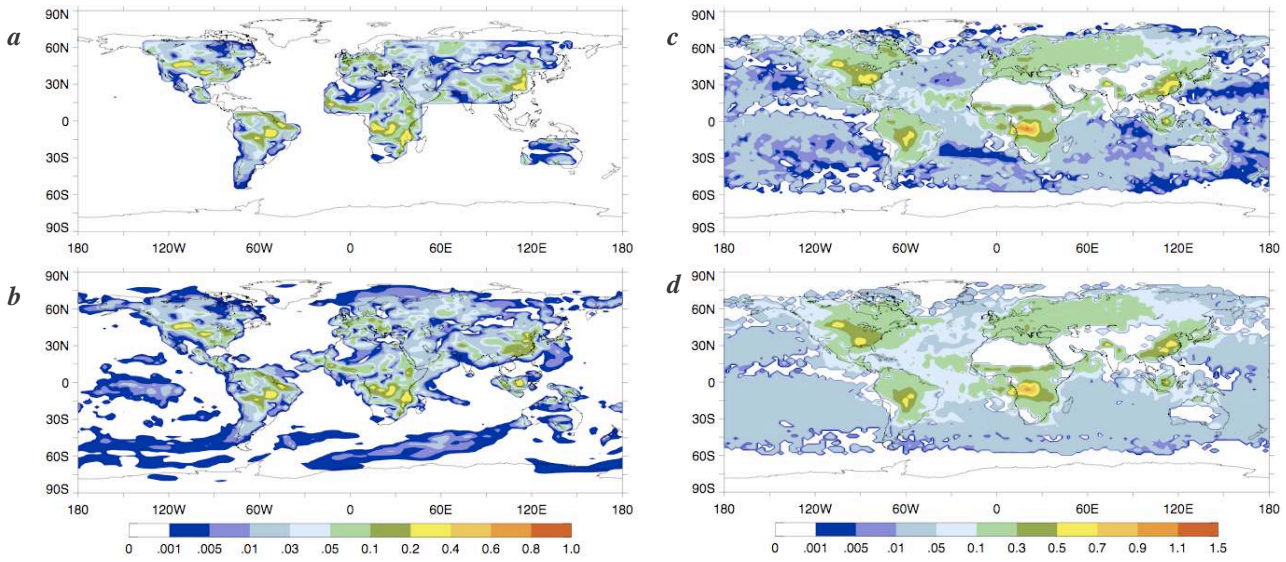

Fig. 16. Inversion of MODIS fine-mode optical depth $\tau_{\text{fine}}(0.55)$ measured during August 18-30 2000. Results are shown here for the two week average of retrieved daily emissions (August 18-30, 2000). (a) Two week average of retrieved aerosol source distribution with emissions constrained to land only. (b) Two week average of retrieved aerosol source distribution with emissions allowed over both land and water (emissions units: 10^7 kg/day). (c) Two week average of MODIS global measurements of $\tau_{\text{fine}}(0.55)$ observed during August 18-30, 2000. (d) Two week average of $\tau_{\text{fine}}(0.55)$ simulated using retrieved daily emissions retrieved over both land and water, assuming a single fine-mode aerosol model.

even if a single coarse mode aerosol is assumed. However, the sea salt retrieval appears very sensitive to uncertainties in the inverted data (e.g. the random noise) because the typical optical thickness of sea salt is very low.

We also performed tests verifying the improved results obtained using a priori constraints in the retrieval. However, the accuracy of the results depends upon the quality of the a priori estimates. Tests showed that constraints on emission time and space variability can be useful for inverting data that have coverage gaps. Some illustrations are given in next Section.

3.4 Inverting MODIS observations

We applied the inversion algorithm to global MODIS measurements of fine mode aerosol optical thickness $\tau_{\text{fine}}(0.55)$ between 18 and 30 August 2000. This is a period of high biomass burning activity, so performing the inversion was expected to illustrate how this approach can constrain BC and OC emissions, that are currently rather uncertain (Sato et al., 2003). For the test, the MODIS Level 3, Collection 4 $\tau_{\text{fine}}(0.55)$ global 1° by 1° data was rescaled to the 2° by 2.5° GOCART horizontal resolution. Where MODIS retrievals were available for the same day over more than 90% of a 2° by 2.5° GOCART grid cell, the average value of available $\tau_{\text{fine}}(0.55)$ was assigned to the entire 2° by 2.5° cell. Where MODIS data were available for less than 90% of a GOCART grid cell, the MODIS data in that cell were not used in the inversion.

3.4.1 Retrieval emissions of fine mode aerosol

Figure 16a–d shows the results of the retrieval assuming a single fine mode aerosol ($\tau_{\text{fine}}(0.55)$); the BC aerosol model was used, as was done in the numerical tests). The retrieved daily emissions shown in Fig. 16a–b were averaged over the entire time period considered. Figure 16c–d compares averaged MODIS $\tau_{\text{fine}}(0.55)$ observations with the retrieved values. The global instantaneous observation fitting accuracy, rescaled to GOCART resolution, was $\sigma_{\text{abs}}\approx 0.04$ and $\sigma_{\text{rel}}\approx 48\%$ after 40 iterations. Thus, $\tau_{\text{fine}}(0.55)$ simulated from the retrieved sources reproduces most of the spatial and temporal tendency in the MODIS observations. It is also important to note than even when the retrieved emissions are not restricted to the land surface, the distribution of main aerosol sources does not significantly change (compare Fig. 16a and b). The retrieved emissions (Fig. 16a–b) may be attributed largely to BC, OC and Sulfates. They may also include some fine mode component dust and possibly sea salt emissions, although these sources are likely to be small, and have rather predictable spatial distributions. Indeed, emissions retrieved over oceans have relatively small magnitudes and are most likely sea salt. (Figure 13 also suggests that random noise can produce small magnitude sources.)

The separation of BC, OC and Sulfate sources is particularly difficult because the emissions of these species are comparable in magnitude, and are often associated with the same biomass burning, fuel combustion and industrial sources (Chin et al., 2002). Therefore, to help in interpreting the

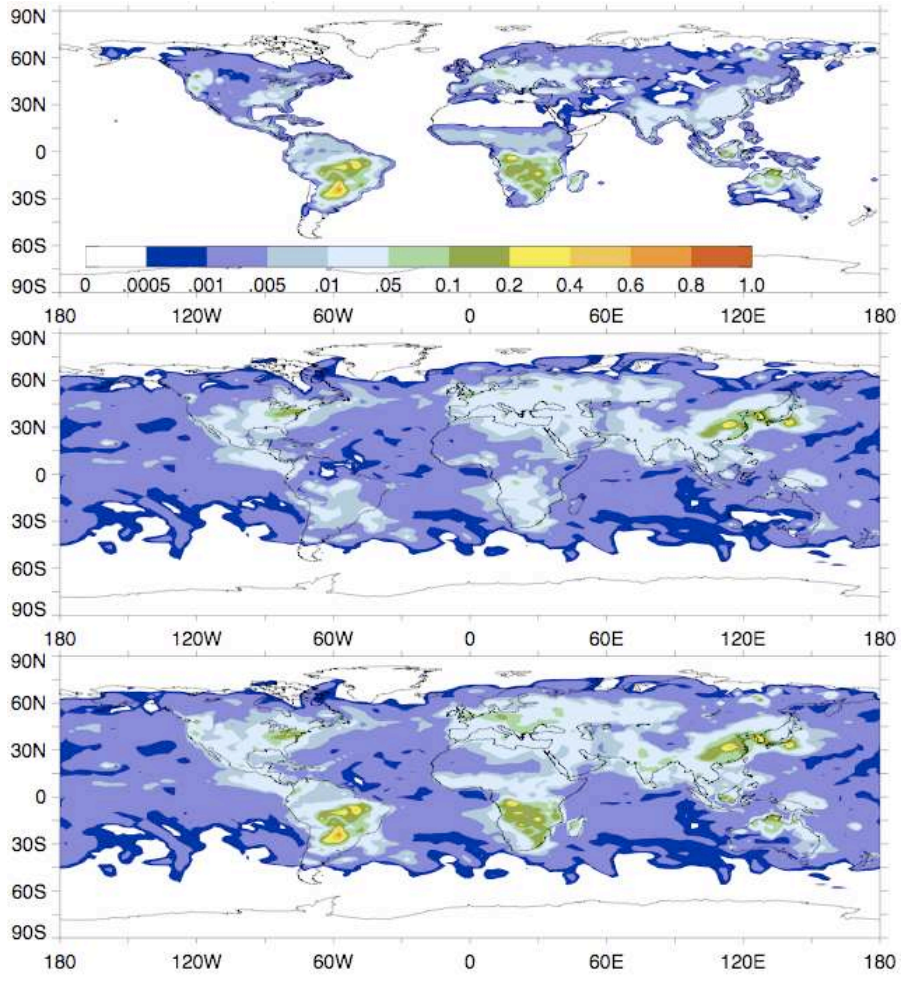

Fig. 17. GOCART simulated fine-mode components, averaged over August 18-30, 2000. (a) BC + OC emissions. (b) Sulfate emissions + atmospheric production. (c) BC + OC + sulfate sources (emissions units: 10^7 kg/day).

retrieval result, we display in Fig. 17 the BC, OC and Sulfate sources assumed in GOCART for August of 2000. In addition, Fig. 18 shows the total carbon emission, obtained from a combination of satellite data and biogeochemical modes by van der Werf et al. (2004), the same data upon which the default BC and OC sources in the GOCART model (shown at upper panel of Fig. 17) are based. Comparing Fig. 16a–b with Figs. 17–18, the global placement of the major fine aerosol sources in the retrieved emission field is in general agreement with known sources of carbon and sulfates. At the same time, the exact shapes and magnitudes of the main retrieved emission patterns (Fig. 16a–b) differ from those assumed in the GOCART model (Fig. 17). According to the main assumption of our approach, these differences should reflect primarily the mismatch between the global aerosol emissions used in modeling and the more realistic emissions distribution that is needed for better agreement with satellite observations, though other factors, including limitations of both measurements and modeling, may be involved (see further discussion of our approach limitation in Sect. 3.5). This assumption is recognized in the modeling community (e.g. Kinne et al., 2003), and many efforts are being made at using observations to improve emissions modeling accuracy. However, the satellite observations used contain information only about monthly variability, whereas the emission retrieval approach considered here could be used to derive emission on daily time scales. For example, standard output of the GOCART model, when the standard emission fields of BC, OC and sulfates are used, produces much higher residuals ($\sigma_{abs}\approx 0.12$ and $\sigma_{rel}\approx 170\%$)) than the output of the simplified single-fine-mode-aerosol GOCART version retrieval ($\sigma_{abs}\approx 0.04$ and $\sigma_{rel}\approx 48\%$).

3.4.2 Coarse mode aerosol emissions retrievals

Figure 19a–d show the results obtained from the MODIS coarse mode aerosol optical thickness $\tau_{coarse}(0.55)$ inversion. The retrievals were performed assuming a single aerosol

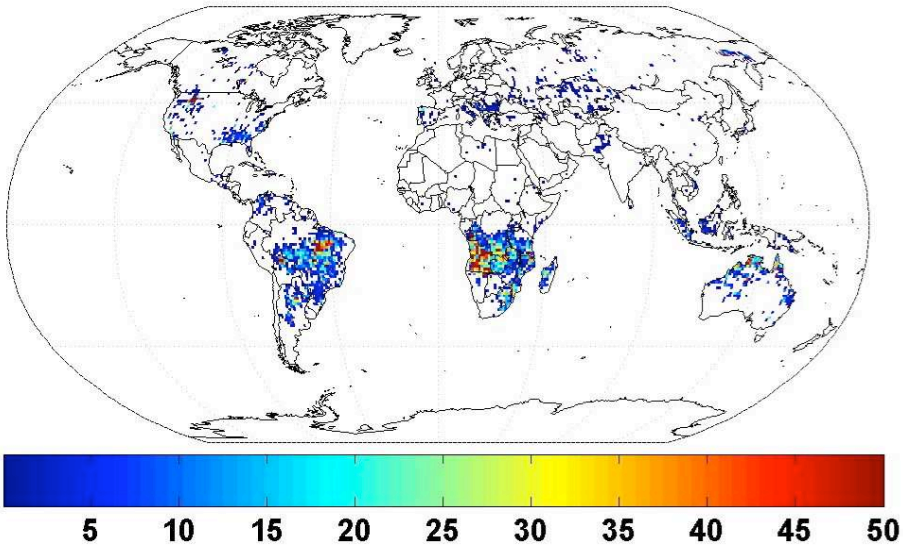

Fig. 18. Monthly carbon emissions (g/m^2) for August 2000 obtained by combining satellite hotspots and burned area with a biogeochemical model (Van der Werf et al., 2003).

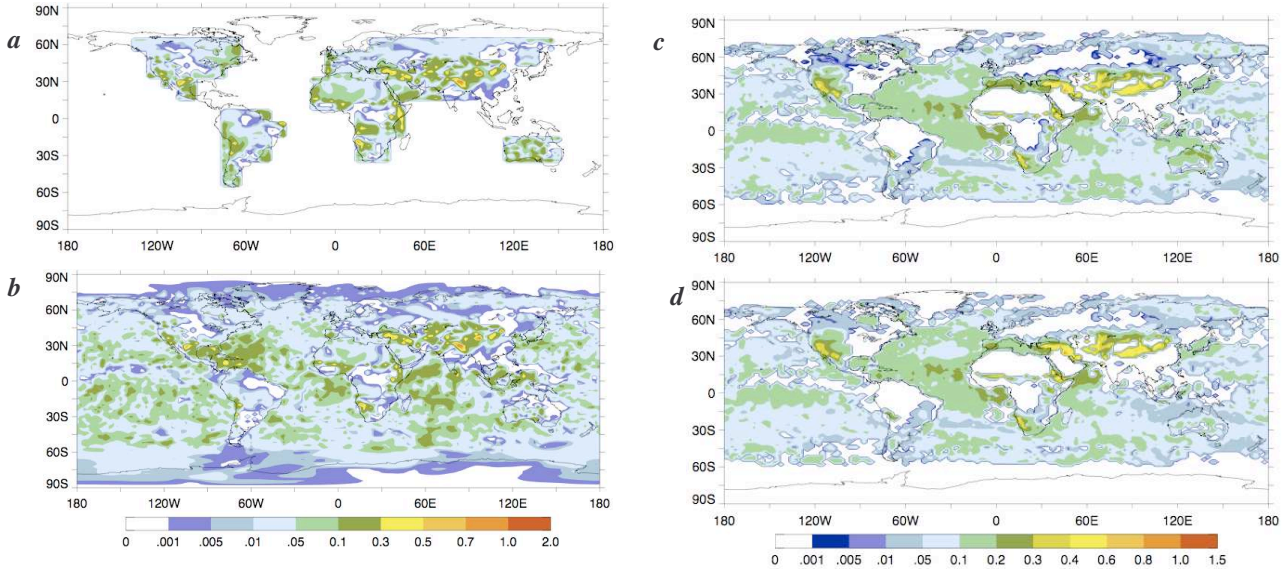

Fig. 19. Inversion of MODIS coarse-mode optical depth ($\tau_{coarse}(0.55)$) measured during August 18-30 2000. Results are shown here for the two week average of retrieved daily emissions (August 18-30, 2000). (a) Two week average of retrieved aerosol source distribution with emissions constrained to land only. (b) Two week average of retrieved aerosol source distribution with emissions allowed over both land and water (emissions units: 10^8 kg/day). (c) Two week average of MODIS global measurements of $\tau_{coarse}(0.55)$ observed during August 18-30, 2000. (d) Two week average of $\tau_{coarse}(0.55)$ simulated using retrieved daily emissions retrieved assuming a single coarse-mode aerosol model.

component, corresponding to the desert dust aerosol model. The desert dust and sea salt emission distributions used by GOCART for the same time period are shown in Fig. 20. Comparing Figs. 19a–b and 20, the major expected desert dust sources can be identified in the retrieved emission field. However, the most intense sources (e.g. over the Saharan Desert) seem to be underestimated, and are more spread out than expected. As discussed earlier, the main cause of these uncertainties is the lack of MODIS observations over deserts (see Fig. 19c). Therefore, even though the $\tau_{coarse}(0.55)$ fitting criteria have low values ($\sigma_{abs}\approx 0.04$ and $\sigma_{rel}\approx 48\%$ after 40 iterations), the desert dust retrieval is less robust than the $\tau_{coarse}(0.55)$ inversion. Also, some likely false sources appear over the Atlantic Ocean near the western coast of

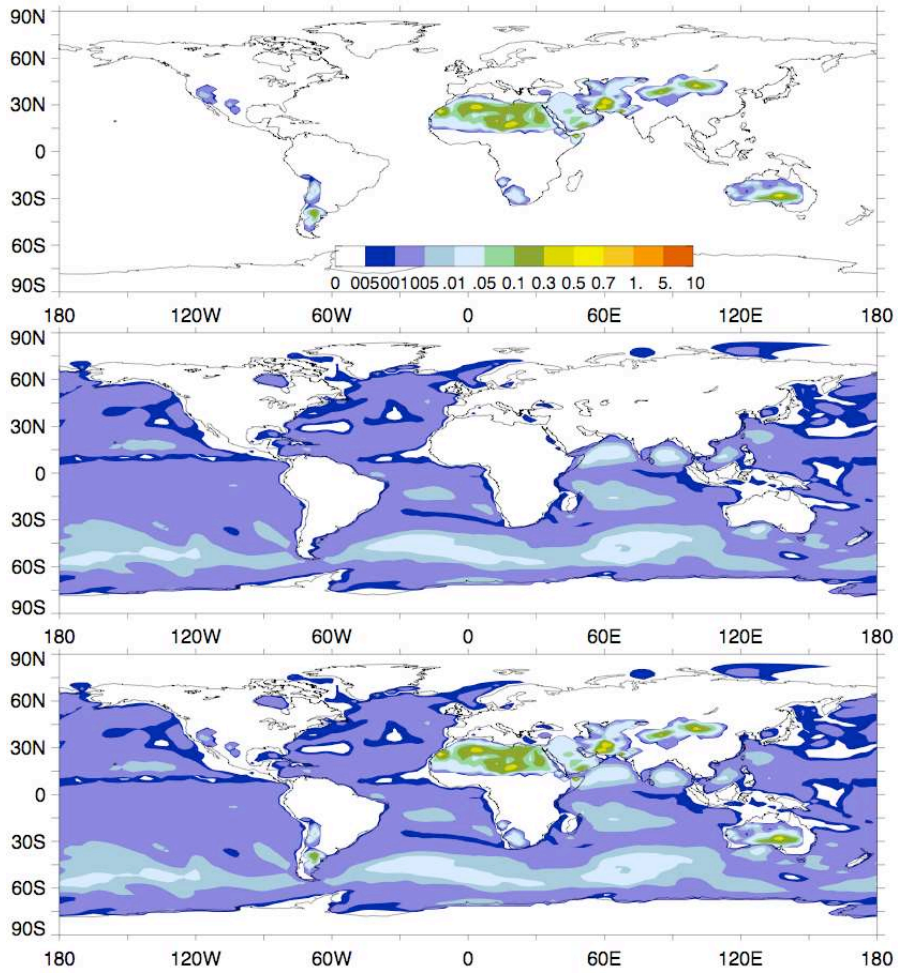

Fig. 20. Averaged (August 18-30, 2000) aerosol sources of desert dust and sea salt (emissions units: 10^8 kg/day) assumed in the GOCART model: Upper panel shows desert dust emissions; Middle panel shows sea salt emissions; Lower panel shows total desert dust + sea salt sources.

Africa. These sources show high $\tau_{coarse}(0.55)$ concentrations that cannot be explained by desert dust transport from Saharan sources. At the same time, high biomass burning aerosol concentrations were transported from Africa, as can be seen in Fig. 16. Given the bulk mode treatment of biomass burning aerosols by GOCART, one can speculate that $\tau_{coarse}(0.55)$ observed over that area could be at least partially attributed to coarse mode biomass burning aerosol. The apparent agreement between the retrieved sea salt source (we attribute all coarse mode aerosol emissions over ocean to sea salt) and that assumed by the GOCART model (Fig. 20) is rather dubious. It probably can be explained by typically low marine aerosol loading (Smirnov et al., 2003), that is often at the level of MODIS retrieval uncertainty. Also, any cloud contamination in the MODIS data is falsely identified by our retrieval as aerosol. Such contamination is common over the tropical oceans. At the same time, both the retrieval (Fig. 16b) and the GOCART assumptions (Fig. 20) show pronounced sea salt emissions over the roaring forties region in the Southern Hemisphere.

We also note that the retrieved emission patterns in Fig. 19b exhibits unrealistically high variability. Indeed, the coarse mode aerosol emission variability is usually related to wind speed variations (e.g. see Ginoux et al., 2001). Nevertheless, it is clear that the emissions shown in Fig. 19b are much more heterogeneous than those assumed in the GOCART model (Fig. 20). We therefore tested the possibility of constraining emission timing and horizontal variability by applying a priori limitations on the second derivatives of the retrieved emissions with respect to time horizontal coordinates x and y (the formulation is given in Sect. 2.6). Figure 21 shows that applying such a priori constraints, especially on spatial variability, helps to eliminate some unrealistically strong emissions over ocean. However, the fitting residuals are higher ($\sigma_{abs}\approx 0.06$ and $\sigma_{rel}\approx 61\%$ after 40 iterations) when the a priori constraints are applied.

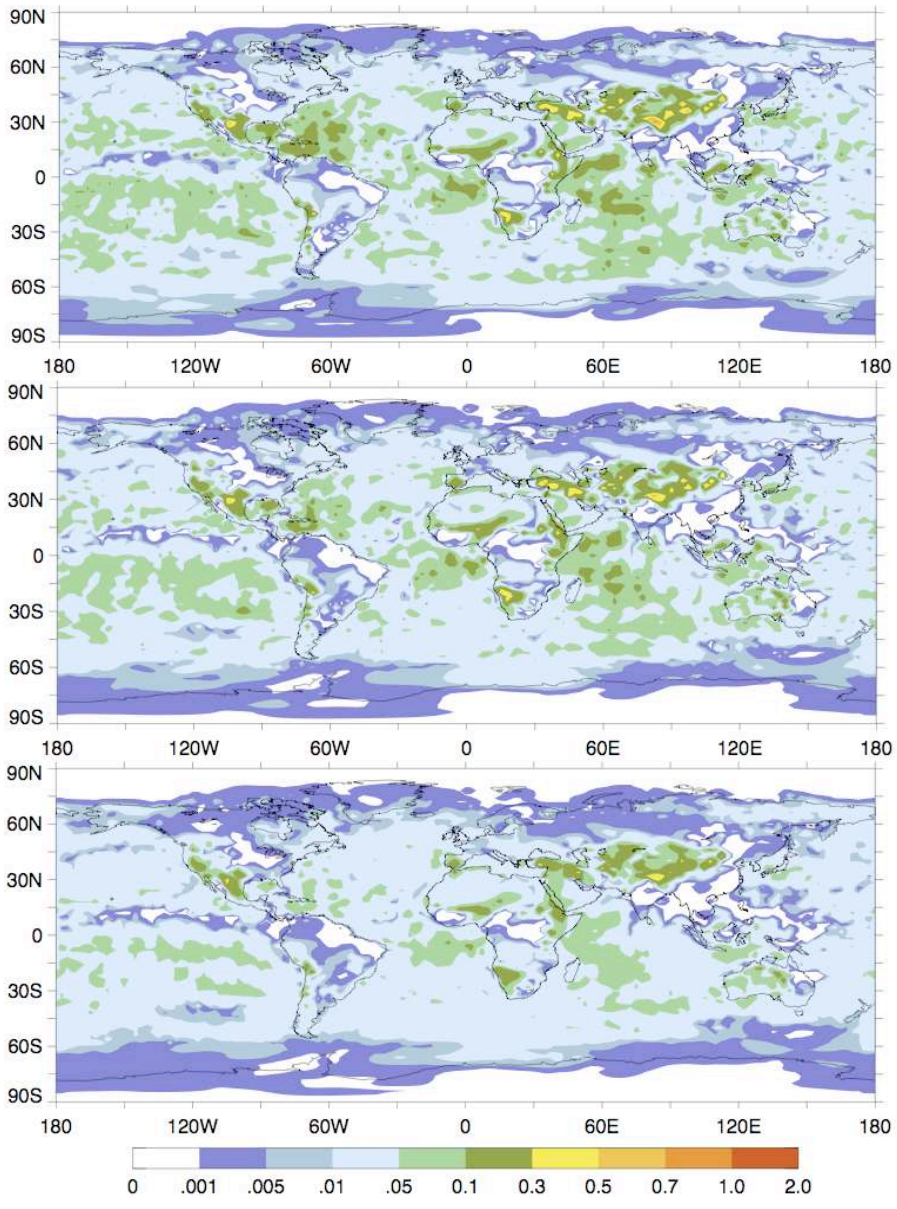

Fig. 21. Effects of time and horizontal continuity smoothness constraints on global coarse-mode source retrievals. **(a)** Results when only the time continuity constraints were applied. **(b)** Results when only the horizontal (xy) aerosol emission variability smoothness constraints were applied. **(c)** Results when both the time and horizontal variability constraints were applied (emissions units: 10^8 kg/day).

3.5 Issues and perspectives

We have shown through numerical tests and applications to real satellite data that the algorithm developed here can retrieve useful information about the global distribution of aerosol emissions without applying excessive a priori constraints on the location and strength of aerosol sources. Further, one can expect that applying the algorithm to long time series of satellite observations should provide a global aerosol emissions climatology that can help to improve our understanding of aerosol climate forcing. Although we have analyzed only two weeks of observations, the algorithm can easily be applied to longer sets of observations. We illustrate this with 6 months of $\tau_{\text{fine}}(0.55)$ MODIS data from 2001. Figure 22 shows the global distribution of fine mode aerosol emissions for February, May and July of that year. The dynamics of global aerosol emissions are apparent, for example, as higher Central and Southern African emissions during February and July, compared to May. This is a manifestation of biomass burning seasonality. Also, the retrievals show high emission over the Indian sub-continent in February, in agreement with known high levels of pollution in this

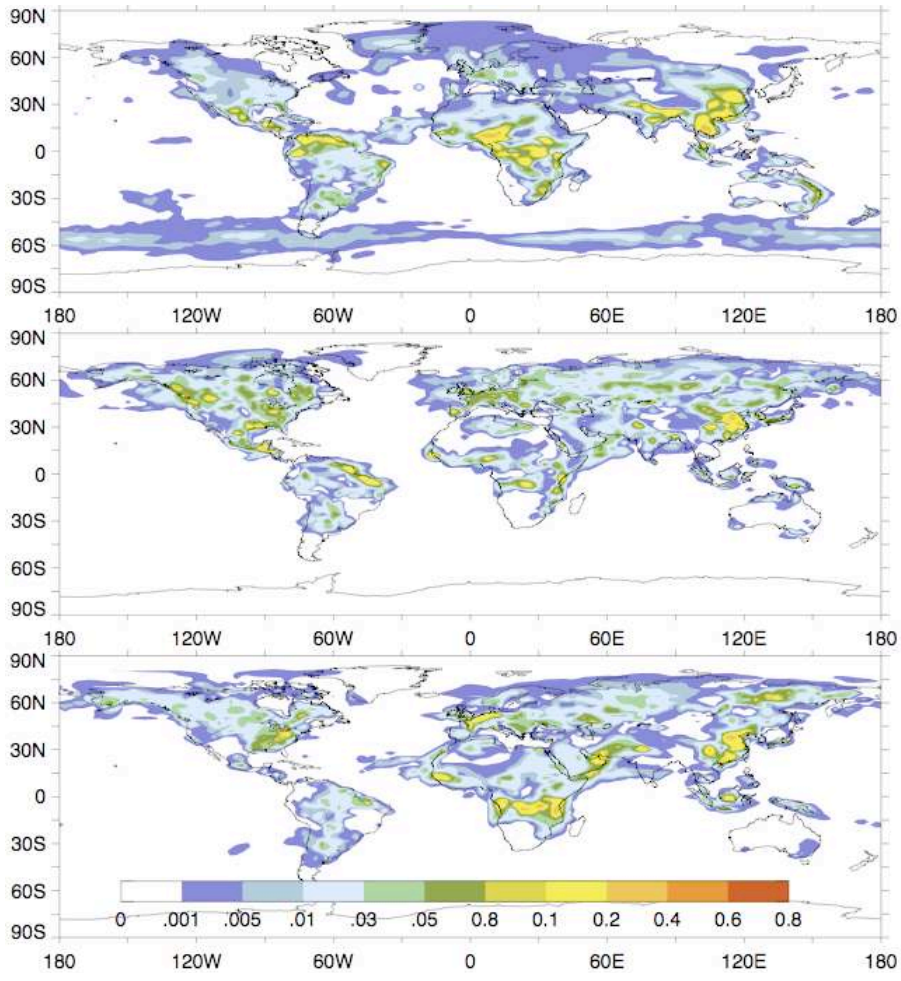

Fig. 22. Monthly averages of retrieved, daily fine-mode global aerosol sources. **(a)** February 2001. **(b)** May 2001. **(c)** July 2001 (emissions units: 10^7 kg/day).

region during winter (e.g., DiGirolamo et al., 2004). The retrievals in Fig. 22 are mainly illustrative; analysis of longer time series, and refinements to the algorithm, are planned.

3.5.1 Limitations of the satellite observations

Obviously the quality of these global emissions retrievals depends on MODIS aerosol data quality. Specifically, as discussed earlier, MODIS data do not provide information about aerosol property vertical variation, and do not distinguish aerosol types by chemical composition (e.g. BC, OC and sulfates). The lack of MODIS retrieval reliability over bright surfaces seriously limits the outcome of desert dust emission retrievals. The accuracy of MODIS optical depth retrievals overall also imposes limitations on global emission retrievals. For example, the inversion of $\tau_{\text{fine}}(0.55)$ produced measurement fitting errors ($\sigma_{\text{abs}}\approx 0.04$ and $\sigma_{\text{rel}}\approx 48\%$) that are as large as MODIS data accuracy, and random measurement noise is not the only limitation. Indeed, the emissions obtained from actual MODIS data (Fig. 16) look significantly more spatially homogeneous than the results of the test retrieval in Fig. 13 (lower panel). This suggests that, in addition to a component of more heterogeneous random noise, significant systematic differences between MODIS observations and GOCART modeling results exist (see also Section 3.5). Our numerical tests indicate that errors of this size would not allow us to benefit fully from atmospheric modeling refinements. Indeed, in the numerical tests shown in Fig. 15a–d, $\tau_{\text{fine}}(0.55)$ composed of BC, OC and sulfates in the GOCART model simulation was fitted with noticeably higher accuracy ($\sigma_{\text{abs}}\approx 0.02$ and $\sigma_{\text{rel}}\approx 25\%$) when a single fine mode aerosol was assumed, neglecting the differences in aerosol chemical composition. Note that the quoted MODIS aerosol optical depth uncertainty applies to the total column, whereas some decrease in accuracy may be expected when $\tau_{\text{total}}(0.55)$ is divided into the $\tau_{\text{coarse}}(0.55)$ and $\tau_{\text{fine}}(0.55)$ quantities used in the present study (Remer et al., 2005; Anderson et al., 2005). Fortunately, a number of improvements in the operational MODIS aerosol algorithm are under development (Remer et al., 2005), including more reliable retrievals over bright surfaces (Hsu et al., 2004).

The global emission retrieval can also be applied to data from other aerosol satellite sensors, such as, MISR (Diner et al., 1998; Kahn et al., 2005), POLDER (Deschamps et al., 1994; Deuze et al., 2001), APS (Mishchenko et al., 2004, 2007), and CALIPSO (Poole et al., 2003). For example, the MISR sensor has multi-angle measurement capability that generally allows a larger number of aerosol properties to be retrieved, including some information about aerosol size and shape, and also produces more robust optical thickness retrievals over bright surfaces (e.g. Kahn et al., 2005; Martonchik et al., 1998). Moreover, satellite instruments with multi-angular polarimetric capabilities have additional sensitivity to particle size distribution and the real part of the refractive index (Mishchenko et al., 2004, 2007). So satellite multi-angle polarimeters may provide aerosol source chemical composition discrimination to hope constrain the global retrieval. For example, the Aerosol Polarimetry Sensor (APS), which is part of currently planned Glory mission (Mishchenko et al., 2007) will perform sensitive polarimetric measurements (over a wide spectral range and with high angular resolution), and is expected to provide such detailed information as fine and coarse mode aerosol refractive index and particle shape. The data provided by space borne lidar on CLAIPSO satellite (Poole et al., 2003) constrain aerosol vertical distribution. Note, however that unlike the MODIS, MISR and POLDER passive imagers that provide two-dimensional global coverage, CLAIPSO and APS are limited to extremely narrow cross-track swaths only a single pixel wide, often referred to as jail-bar coverage.

To maximize the benefits from available satellite aerosol measurements, the global inversion should utilize both the detailed aerosol information provided by CLAIPSO and APS together with less accurate but more extensive observations by satellite imagers. Using data from multiple satellite sensors as input to the global inversion should provide better spatial and temporal data coverage compared to the coverage of any single sensor. For example, the MODIS instruments on the Terra and Aqua satellites (Remer et al., 2006) observe the same geographic locations in the late morning and early afternoon, offering the global inversion some information about aerosol emission diurnal variability.

The global inversion can also include ground-based network observations, such as those from the AERONET sunphotometer network (Holben et al., 1998), and the EARLINET lidar network (Boesenberg et al., 2003). Unlike satellite observations, these networks do not provide global coverage, but they offer more accurate, high-temporal-resolution information at a number of global locations that can be used in the global inversion. The benefits of these data combinations will probably be greatest when, unlike the examples given here, global inversions having high temporal and spatial resolution are required. Thus, the global aerosol inversion can also be considered a tool for synthesizing information from multiple measurement types.

3.5.2 Modeling accuracy limitations

As discussed in Sect. 2, our inversion approach finds an optimal solution using known measurement error statistics, under the assumption that the model errors are much smaller than those of the measurements, so the effect of the model errors can be neglected. This assumption is employed in most inverse methodologies, and it can easily be justified in many applications. However, it is not yet possible to perform global transport modeling with the same level of certainty as radiative modeling, for example, so the effects of the transport modeling uncertainties on the retrievals may not be negligible. Many factors may contribute to transport modeling uncertainty. For example, the atmospheric temperature, pressure, wind distributions, and other meteorological fields inputed to the models are known with limited accuracy. Significant uncertainty in the four-dimensional distribution of clouds affects the aerosol modeling, e.g. via uncertainties in aerosol transport by cloud convection. The aerosol removal processes employed by the models are also known with only limited accuracy, and aerosol-cloud interactions can be described only qualitatively at present. The temporal and spatial resolution achieved by models such as GOCART are insufficient to reproduce some local details of aerosol dynamics. In addition, all models suffer from numerical instabilities that can produce negative aerosol mass, failure to conserve aerosol mass in transport simulations, etc. All these uncertainties limit the accuracy of global aerosol modeling, as demonstrated in recent model inter-comparison studies (e.g., Kinne et al., 2003, 2006; Textor et al., 2006).

However, a number of efforts aim at addressing many of these modeling uncertainties in the near future. For example, four-dimensional meteorological wind fields and cloud distributions will probably be improved by assimilation of observations from multiple sources. There is also hope that model aerosol optical properties can be improved by remote sensing data. Like most models, GOCART assumes all coarse aerosols are either dust or sea salt. Satellite retrieval algorithms (e.g. Kaufman et al., 1997; Tanré et al., 1997; Remer et al., 2005, etc.) rely on aerosol type climatologies obtained from other remote sensing measurements. Unlike global models, such climatologies (e.g. Remer et al., 1997; Dubovik, 2002) indicate the presence of coarse particles in practically all types of aerosol including biomass burning and urban pollution. Also, satellite retrievals usually use complex indices of refraction derived from remote sensing observations to describe ambient aerosol, whereas models tend to calculate radiative properties of aerosol based on chemical composition, using indices of refraction obtained from in situ or laboratory measurements of each chemical component. This difference may be another source of inconsistency between remote sensing retrievals and models. For example, recent analysis of aerosol remote sensing retrieval results (Kaufman et al., 2001; Dubovik et al., 2002) showed that mineral dust is less absorbing than was previ-

ously thought, and comparisons with aerosol transport models showed systematic differences with remote sensing observations of aerosol absorption for situations dominated by mineral dust (e.g. Takemura et al., 2002; Sato et al., 2003). One possible way to address this inconsistency is to include these aspects of the satellite retrieval as part of inverse modeling. In this approach, the aerosol emissions will be retrieved directly from atmospheric radiances (e.g. using the mathematical formalism described in Sect. 2.8).

3.5.3 Limitations of the retrieval approach employed

The numerical inversion approach has some limitations in itself. As explained in Sect. 2.5, the algorithm is based on an iterative, steepest descent method that in general converges very slowly (e.g. Tarantolla, 1987). However, the fitting residual can decrease to a value much smaller than the expected noise level of inverted satellite data after 40 iterations (see Sects. 3.3–3.4), because the transport operator **T** (written in matrix form) is sparse. Since each iteration involves two GOCART model runs (forward and backward), inverting global data takes about 80 times longer than a GOCART forward simulation of the aerosol fields covering same time period. Nevertheless, taking into account the global nature of the problem and the increasing speed of modern computers, these requirements are probably acceptable. For this study, which used only moderately powerful computers, we tried to identify ways to reduce the number of iterations (e.g. Gill et al., 1982; Tarantolla, 1987). Theoretically, the method of conjugated gradients can accelerate convergence compared to steepest descents; it also uses only gradient vectors and can therefore be implemented with transport model adjoint operators (see Appendix C). However, in practice, rounding errors often cause the computed directions to loose conjugacy rapidly, and the method behaves more like an iterative method, making converge much slower than theoretically predicted (Gill et al., 1982). To test the possibility of accelerating inversion convergence, we implemented the method of conjugated gradients described in Appendix C. Convergence improved: after only 20 iterations, the method of conjugated gradients achieved the same residual size as the steepest descent method after 40 iterations. However, as can be seen from the derivation in Appendix C, our implementation of the method of conjugated gradients requires running the GOCART model four times for a single iteration, instead of the two runs needed for steepest descents. Therefore it took the same computing time to perform 20 iterations with method of conjugated gradients as 40 iterations with steepest descents, so conjugated gradients did not effectively accelerate the inversion. At the same time, implementing the method of conjugated gradients required greater effort, because the formal logic is significantly more complex. Nevertheless, there may be a need to explore convergence acceleration possibilities further, in anticipation of the inevitable increase in aerosol global model resolution. Similarly, the steepest decent method may converge more slowly if the satellite retrieval is included as part of the inverse modeling (discussed in Sect. 2.7), due to the complexity of the radiative transfer equations.

Using diverse a priori constrains is another direction requiting further exploration. In Sect. 2.6 we described the mathematical techniques for using a priori emissions estimates or a priori limitations on the derivatives of emissions variability in time and space. Using these a priori constraints can substantially extend the field of retrieved parameters. For example, with an a priori estimate of emissions, MODIS data can be used in an emissions retrieval that discriminates aerosol type, even though the MODIS data contain little information about aerosol type. As discussed in Sect. 3.3, to enhance the sensitivity of our retrieval to the satellite observations, we decided not to constrain the retrieval with a priori emissions estimates. However, to produce a unique retrieval without such a priori information, we eliminated the aerosol chemical composition discrimination in our retrieval. As a result, the retrieval provided by this algorithm has high sensitivity to spatial and temporal emission features that were observed by satellites, but were unexpected from a modeling viewpoint. However the retrieved emission fields are of limited value for modeling activities that require aerosol type discrimination. Future efforts to improve this approach will aim at including in the retrieval a priori constraints that retain aerosol type discrimination without over-constraining retrieved emission spatial and temporal variability.

This advance is imperative for the future development of global emission retrievals, because even the most advanced aerosol satellite sensors will not constrain all the parameters used to describe aerosol emission in chemical transport models. For example, we will attempt to constrain a priori only relative contributions from different aerosol components to the total aerosol optical thickness, without constraining its magnitudes of the emissions.

We also showed (Fig. 21) that by placing a priori limitations on the emission variability time and space derivatives, one can retain emission fields with physically realistic time and space continuity (with no sharp oscillations), even if the inverted measurements do not provide sufficient constraints. However, we found that due to the relatively coarse time (24 h) and space ($2.0^\circ \times 2.5^\circ$) resolution used in this study, these constraints generally do not seem critical to our retrievals. Nevertheless, in future studies that attempt retrievals at much higher resolutions, the importance of continuity constraints will likely increase.

4 Conclusions

This paper describes an algorithm that uses inverse modeling to retrieve global aerosol source emissions from satellite observations.

The numerical inversion in the algorithm was structured as a multi-term least-squares-type fit. This statistical optimization scheme allows for high flexibility in constraining the retrieval, such as using multiple constraints in a single inversion. We also discuss and demonstrate the possibility of applying a priori constraints on retrieved quantity partial derivatives, to improve retrieved global aerosol emission time and space variability. This approach is widely used in atmospheric remote sensing, though rarely in inverse modeling. The similarities and differences of our inversion scheme with the standard Kalman filter and 4-D-var assimilation inverse modeling and Phillips-Tikhonov-Twomey constrained inversion remote sensing approaches are discussed. To simultaneously retrieve global aerosol fields over extended periods of time at the space and time resolution of the model, the fitting was expressed in terms of a variational approach, using an adjoint of the aerosol transport in a convenient form for practical implementation of the inversion. We also discuss the application of various a priori constraints in the adjoint form of the retrieval algorithm.

The algorithm was implemented in a practical way with the GOCART aerosol transport model, retrieving global aerosol emissions at $2^\circ \times 2.5^\circ$ horizontal resolution from the global distribution of aerosol optical thickness. The numerical tests conducted showed that the algorithm accurately derives aerosol emissions when inverting the detailed, global aerosol mass distribution produced from a forward run of GOCART model. However, when MODIS data was used as input for the inversion, some extra constraints were needed to produce a unique retrieval, due to limitations in coverage and information content of the MODIS data. Specifically, emission variability over the diurnal cycle was neglected, and aerosol types were discriminated by particle size, but not by their chemical composition. Emissions of fine and coarse mode aerosols were retrieved from the MODIS fine and coarse mode aerosol optical thickness data, respectively.

From two weeks of daily MODIS observations during August 2000, the global placement of fine mode aerosol sources agreed with available independent knowledge, even though the inverse method did not use any a priori information about aerosol sources, and was initialized with a "zero aerosol emission" assumption. The retrieval reproduced two weeks of global, instantaneous MODIS observations, with an aerosol optical thickness standard deviation for the fit of $\sim$0.04. Aerosol optical thickness observations of 0.05 and higher were reproduced with a standard deviation of $\sim$48%. Such agreement between global modeling and observation is quite encouraging, given that the coherence between model and observations is limited by MODIS observation accuracy, aerosol variability that can be much higher than the model resolution, uncertainties in the wind fields, three-dimensional cloud distributions, and other meteorological data, the accuracy of assumptions made in modeling atmospheric processes, numerical instabilities, etc. As a result, model predictions can differ significantly from observations, even for monthly and yearly averaged regional aerosol properties (Kinne et al., 2003; Sato et al., 2003).

Retrieving coarse mode aerosol emissions was less successful, mainly because MODIS aerosol data over highly reflecting desert dust sources is lacking. This situation should be much improved by the most recent MODIS product version (Collection 5), which includes aerosol retrievals over deserts from the "deep blue" algorithm of Hsu et al. (2004).

The efficiency of using a priori constraints on values of the emissions or on their variability was also evaluated. The use of such a priori constraints is a clear alternative to a straightforward reduction of the number of retrieved parameters characterizing the aerosol emission. For example, in spite of limitations in the MODIS data, if the algorithm uses GOCART emission fields as a priori estimates, it can be set to retrieve aerosol emissions in the same format as those used in the GOCART model (i.e. fully discriminated by chemical composition and sizes). This strategy is usually employed in assimilation methods. However, in that retrieval approach, the derived emissions tend to be influenced by the a priori estimates to an extent that may devaluate the satellite observations. Therefore, further effort is needed to minimize the possible effect of over-constraining the solution. In the present paper, we did not use a priori estimates for inverting MODIS observations; instead, we applied constraints that limit the general temporal and spatial variability of the emissions, and demonstrated that such general constraints were useful for eliminating some unrealistic features in underdetermined retrievals, such as coarse mode aerosol emission retrievals from MODIS observations.

To illustrate algorithm performance when processing long satellite observation records, we inverted the fine mode aerosol optical thickness produced by MODIS, for the first six months of 2001. Realistic global aerosol emission geographic distribution patterns, and their seasonal variations, were derived.

Thus, the method developed can be a useful tool for improving global aerosol source characterization in chemical transport models. Nevertheless, this paper describes only the first phase of the effort, and further analysis is needed to realize the full potential of the method. Specifically, we plan to use MODIS data to generate climatological records of remote-sensing-driven aerosol emission fields; these are expected to provide improved global aerosol source descriptions as inputs to chemical transport models. We plan to refine and optimize the use of a priori constraints and the convergence of the method. Also, at least some of the satellite measurement limitations are expected to be addressed in future studies, for example, by planned MODIS aerosol algorithm improvements.

Moreover, the global emission retrievals can incorporate a combination of aerosol products from other satellite sensors, such as CALIPSO, MISR, PARASOL and APS, as well as aerosol measurements from ground-based networks, such as AERONET and EARLINET. Using data from multiple

satellite sensors as input should improve spatial and temporal constraints from the global emission retrievals, compared to using data exclusively from any single sensor. In addition, using data from multi-angle radiometers (MISR), polarimeters (PARASOL, APS), lidar (CALIPSO), and ground-based networks (AERONET and EARLINET), may make it possible to discriminate between emissions of different aerosol types in the retrievals. Thus, the global aerosol inversion can also be considered as a tool for integrating the information from models and remote sensing instruments, providing aerosol information of enhanced value for climate research.

Appendix A

The adjoint operator

According to its formal definition (e.g. Tarantola, 1987), the adjoint of $\mathbf{G}$, $\mathbf{G}^{\#}$, is a linear operator defined by the equality of scalar products:

$$< \mathbf{G}^{\#}\boldsymbol{x}, \boldsymbol{y} >=< \boldsymbol{x}, \ \mathbf{G}\ \boldsymbol{y} > . \tag{A1}$$

If scalar product is defined as follows:

$$< \boldsymbol{a}, \boldsymbol{b} >= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}, \tag{A2}$$

then the right side of Eq. (A1) is:

$$< \boldsymbol{x}, \ \mathbf{G}\ \boldsymbol{y} >= \boldsymbol{x}^{\mathrm{T}}\ \mathbf{G}\ \boldsymbol{y}. \tag{A3}$$

The left side of Eq. (A1) is:

$$< \mathbf{G}^{\#}\boldsymbol{x}, \boldsymbol{y} >= (\mathbf{G}^{\#}\boldsymbol{x})^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}(\mathbf{G}^{\#})^{\mathrm{T}}\boldsymbol{y}. \tag{A4}$$

Thus, in order to achieve the equality between Eqs. (A3) and (A4) we can write for $\mathbf{G}^{\#}$:

$$\mathbf{G}^{\#} = \mathbf{G}^{\mathrm{T}}. \tag{A5}$$

It should be noted that Tarantola (1987) gives a more general definition of the adjoint operator that is not used here.

Appendix B

Derivation of Eq. (41)

Equation (2) can be written in matrix form, for a single time step, as follows:

$$\boldsymbol{M}_n = \mathbf{T}_{n-1}(\boldsymbol{M}_{n-1} + \boldsymbol{S}_{n-1}), \tag{B1}$$

where subscripts " $n-1$ " and " n " are associated with times steps t_{n-1} and t_n=t_{n-1}+Δt, i.e. matrix $\mathbf{T}_{n-1}$ and vectors $\boldsymbol{S}_{n-1}$, $\boldsymbol{M}_{n-1}$ and $\boldsymbol{M}_n$ represent $T\ (t_{n-1}, \boldsymbol{x})$, $s\ (t_{n-1}, \boldsymbol{x})$, $m\ (t_{n-1}, \boldsymbol{x})$ and $m\ (t_{n-1}+\Delta t, \boldsymbol{x})$, respectively. For time steps t_{n-2}, t_{n-1} and t_n one can write:

$$\begin{aligned}\boldsymbol{M}_n &= \mathbf{T}_{n-1}(\boldsymbol{M}_{n-1} + \boldsymbol{S}_{n-1}) = \mathbf{T}_{n-1}(\mathbf{T}_{n-2}(\boldsymbol{M}_{n-2} + \boldsymbol{S}_{n-2}) + \boldsymbol{S}_{n-1}) = \\ &= \mathbf{T}_{n-1}\mathbf{T}_{n-2}\boldsymbol{M}_{n-2} + \mathbf{T}_{n-1}\mathbf{T}_{n-2}\boldsymbol{S}_{n-2} + \mathbf{T}_{n-1}\boldsymbol{S}_{n-1}.\end{aligned} \tag{B2}$$

Correspondingly, the mass transport for $t_0, t_1, \ldots, t_n$ can be expressed as:

$$\boldsymbol{M}_n = \left(\prod_{i=0}^{i=n-1}\mathbf{T}_i\right)\boldsymbol{M}_0 + \sum_{k=0}^{k=n-1}\left(\prod_{i=k}^{i=n-1}\mathbf{T}_i\right)\boldsymbol{S}_k \quad , \tag{B3}$$

where

$$\prod_{i=0}^{i=n-1}\mathbf{T}_i = \mathbf{T}_{n-1}\mathbf{T}_{n-2}...\mathbf{T}_2\mathbf{T}_1\mathbf{T}_0 \quad . \tag{B4}$$

Based on Eq. (B3), the entire matrix Eq. (5) for mass distribution during the time period from t_0 to t_n can be written as follows:

$$\begin{pmatrix}\boldsymbol{M}_n^* \\ \ldots \\ \boldsymbol{M}_3^* \\ \boldsymbol{M}_2^* \\ \boldsymbol{M}_1^*\end{pmatrix} = \begin{pmatrix} \mathbf{T}_{n-1}\prod_{i=n-2}^{i=n-1}\mathbf{T}_i & \ldots & \prod_{i=2}^{i=n-1}\mathbf{T}_i & \prod_{i=1}^{i=n-1}\mathbf{T}_i & \prod_{i=0}^{i=n-1}\mathbf{T}_i \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & \ldots & \mathbf{T}_2 & \prod_{i=1}^{i=2}\mathbf{T}_i & \prod_{i=0}^{i=2}\mathbf{T}_i \\ 0 & \ldots & 0 & \mathbf{T}_1 & \prod_{i=0}^{i=1}\mathbf{T}_i \\ 0 & \ldots & 0 & 0 & \mathbf{T}_0 \end{pmatrix} \begin{pmatrix}\boldsymbol{S}_{n-1} \\ \ldots \\ \boldsymbol{S}_2 \\ \boldsymbol{S}_1 \\ \boldsymbol{S}_0\end{pmatrix} + \begin{pmatrix}\prod_{i=0}^{i=n-1}\mathbf{T}_i \\ \ldots \\ \prod_{i=0}^{i=2}\mathbf{T}_i \\ \prod_{i=0}^{i=1}\mathbf{T}_i \\ \mathbf{T}_0\end{pmatrix}\boldsymbol{M}_0, \tag{B5}$$

where $\boldsymbol{M}_n^*$ denotes the mass distribution of aerosols emitted during time period from t_0 to t_n: i.e.

$$\boldsymbol{M}_n^* = \boldsymbol{M}_n - \left(\prod_{i=0}^{n-1}\mathbf{T}_i\right)\boldsymbol{M}_0. \tag{B6}$$

Thus, Eq. (B5) is equivalent to the generalized matrix expression:

$$\boldsymbol{M}^* = \boldsymbol{M}\ - \mathbf{T}\boldsymbol{M}_0 = \mathbf{T}\ \boldsymbol{S}, \tag{B7}$$

This is a slightly modified form of Eq. (5). Similarly, for the correction term $\Delta\hat{\boldsymbol{S}}^p$ of the steepest descent iterative solution given by Eq. (40) can be written as:

$$\Delta\hat{\boldsymbol{S}}^p = \mathbf{T}^{\mathrm{T}}\mathbf{C}_m^{-1}\Delta\boldsymbol{M}^p = \mathbf{T}^{\mathrm{T}}\Delta^p = \begin{pmatrix} \mathbf{T}_{n-1}^{\mathrm{T}} & \ldots & 0 & 0 & 0 \\ \prod_{i=n-2}^{i=n-1}\mathbf{T}_i^{\mathrm{T}} & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \prod_{i=n-1}^{i=2}\mathbf{T}_i^{\mathrm{T}} & \ldots & \mathbf{T}_2^{\mathrm{T}} & 0 & 0 \\ \prod_{i=n-1}^{i=1}\mathbf{T}_i^{\mathrm{T}} & \ldots & \prod_{i=2}^{i=1}\mathbf{T}_i^{\mathrm{T}} & \mathbf{T}_1^{\mathrm{T}} & 0 \\ \prod_{i=n-1}^{i=0}\mathbf{T}_i^{\mathrm{T}} & \ldots & \prod_{i=2}^{i=0}\mathbf{T}_i^{\mathrm{T}} & \prod_{i=1}^{i=0}\mathbf{T}_i^{\mathrm{T}} & \mathbf{T}_0^{\mathrm{T}} \end{pmatrix} \begin{pmatrix}\Delta_n^p \\ \Delta_{n-1}^p \\ \ldots \\ \Delta_3^p \\ \Delta_2^p \\ \Delta_1^p\end{pmatrix}, \tag{B8}$$

where $\Delta^p = \mathbf{C}_m^{-1}\Delta\boldsymbol{M}^p$, Δ_i^p denotes the component of vector Δ^p corresponding to time step t_i and the following identity for the transpose of the matrix multiplication product is used:

$$\left(\prod_{i=0}^{i=n}\mathbf{T}_i\right)^{\mathrm{T}} = (\mathbf{T}_n\mathbf{T}_{n-1}...\mathbf{T}_2\mathbf{T}_1\mathbf{T}_0)^{\mathrm{T}} = \mathbf{T}_0^{\mathrm{T}}\mathbf{T}_1^{\mathrm{T}}\mathbf{T}_2^{\mathrm{T}}...\mathbf{T}_{n-1}^{\mathrm{T}}\mathbf{T}_n^{\mathrm{T}} = \prod_{i=n}^{i=0}\mathbf{T}_i^{\mathrm{T}}. \quad \text{(B9)}$$

From Eq. (B8), $\Delta\boldsymbol{S}^p$ can be obtained as:

$$\Delta\boldsymbol{S}^p = \begin{pmatrix} \Delta\boldsymbol{S}_n^p \\ \Delta\boldsymbol{S}_{n-1}^p \\ ... \\ \Delta\boldsymbol{S}_2^p \\ \Delta\boldsymbol{S}_1^p \\ \Delta\boldsymbol{S}_0^p \end{pmatrix} = \begin{pmatrix} \mathbf{T}_{n-1}^{\mathrm{T}}\Delta_n^p \\ \mathbf{T}_{n-2}^{\mathrm{T}}\Delta_{n-1}^p + \mathbf{T}_{n-2}^{\mathrm{T}}\mathbf{T}_{n-1}^{\mathrm{T}}\Delta_n^p \\ ... \\ \sum_{k=3}^{k=n-1}\left(\prod_{i=n-1}^{i=k}\mathbf{T}_i^{\mathrm{T}}\Delta_k^p\right) \\ \sum_{k=2}^{k=n-1}\left(\prod_{i=n-1}^{i=k}\mathbf{T}_i^{\mathrm{T}}\Delta_k^p\right) \\ \sum_{k=1}^{k=n-1}\left(\prod_{i=n-1}^{i=k}\mathbf{T}_i^{\mathrm{T}}\Delta_k^p\right) \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{T}_{n-1}^{\mathrm{T}}\Delta_n^p \\ \mathbf{T}_{n-2}^{\mathrm{T}}\left(\Delta_{n-1}^p + \boldsymbol{S}_{n-1}^p\right) \\ ... \\ \mathbf{T}_2^{\mathrm{T}}\left(\Delta_3^p + \Delta\boldsymbol{S}_3^p\right) \\ \mathbf{T}_1^{\mathrm{T}}\left(\Delta_2^p + \Delta\boldsymbol{S}_2^p\right) \\ \mathbf{T}_0^{\mathrm{T}}\left(\Delta_1^p + \Delta\boldsymbol{S}_1^p\right) \end{pmatrix}. \quad \text{(B10)}$$

From this equation it can be seen that $\Delta\boldsymbol{S}_i^p$ can be calculated via the following sequence starting from $i{=}n$ as follows:

$$\Delta\boldsymbol{S}_{i-1}^p = \mathbf{T}_{i-1}^{\mathrm{T}}\left(\Delta_i^p + \Delta\boldsymbol{S}_i^p\right). \quad \text{(B11)}$$

The component Δ_i^p of vector Δ^p corresponding to time step t_i can be easily formulated if observational errors $\Delta\boldsymbol{M}^*$ are not correlated with respect to time, though they may have spatial correlations. In this case, $\mathbf{C}_m$ has the array structure:

$$\mathbf{C}_m = \begin{pmatrix} \mathbf{C}_{m_n} & ... & 0 & 0 \\ ... & ... & ... & ... \\ 0 & ... & \mathbf{C}_{m_2} & 0 \\ 0 & ... & 0 & \mathbf{C}_{m_1} \end{pmatrix}, \quad \text{(B12)}$$

and Δ^p can be decomposed:

$$\Delta^p = \begin{pmatrix} \Delta_n^p \\ ... \\ \Delta_2^p \\ \Delta_1^p \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{m_n}^{-1}\left(\boldsymbol{M}_n\left(\hat{\boldsymbol{S}}^p\right) - \boldsymbol{M}_n^*\right) \\ ... \\ \mathbf{C}_{m_2}^{-1}\left(\boldsymbol{M}_2\left(\hat{\boldsymbol{S}}^p\right) - \boldsymbol{M}_2^*\right) \\ \mathbf{C}_{m_1}^{-1}\left(\boldsymbol{M}_1\left(\hat{\boldsymbol{S}}^p\right) - \boldsymbol{M}_1^*\right) \end{pmatrix}. \quad \text{(B13)}$$

Thus, Eq. (B11) gives the relationship between $\Delta\boldsymbol{S}_{i-1}^p$ for time step t_{i-1} and $\Delta\boldsymbol{S}_i^p$ for time step $t_i{=}t_{i-1}{+}\Delta t$. As the time step approaches zero, i.e. $\Delta t{\to}0$, then Eq. (B11) can be rewritten as its integral equivalent:

$$\Delta\hat{s}^p(t,\boldsymbol{x}) = \int_t^{t_0} T^{\#}\left(t',\boldsymbol{x}\right)\left(\Delta\hat{s}^p\left(t',\boldsymbol{x}\right) + \Delta^p\left(t',\boldsymbol{x}\right)\right)(-dt'), \quad \text{(B14)}$$

where the function $\Delta^p(t,\boldsymbol{x})$ is the continuous analog of the vector Δ^p. It can be formulated via the weighting function $C^{-1}(t,\boldsymbol{x},\boldsymbol{x'})$ (from covariance function $C_t(t,\boldsymbol{x},\boldsymbol{x'})$), and performs a role analogous to that of matrix $\mathbf{C}_{m_i}^{-1}$ in the discrete representation, i.e.:

$$\left\{\Delta_i^p\right\} = \left\{\mathbf{C}_{m_i}^{-1}\Delta\boldsymbol{M}^p\right\}_l \to$$
$$\iiint_{x',y',z'} C_t^{-1}\left(t_i,x_j,y_k,z_m,x',y',z'\right)\Delta m^p\left(t_i,x',y',z'\right)dx'dy'dz'. \quad \text{(B15)}$$

$T^{\#}(t,\boldsymbol{x})$ is the transport adjoint operator $T(t,\boldsymbol{x})$, and is composed of adjoints $T_i^{\#}(t,\boldsymbol{x})$ of the component processes $T_i(t,\boldsymbol{x})$:

$$T^{\#}(t,\boldsymbol{x}) = T_1^{\#}T_2^{\#}T_3^{\#}...T_{n-1}^{\#}T_n^{\#}. \quad \text{(B16)}$$

If the observation errors are uncorrelated, i.e. the covariance matrix of measurements $\mathbf{C}_m$ is diagonal with the elements on diagonal equal to $\sigma^2\left(t_i,x_j,y_k,z_m\right)$, the elements of vector $\mathbf{C}_m^{-1}\Delta\boldsymbol{M}^p$ relate to the continuous function $\sigma^{-2}(t,\boldsymbol{x})\,\Delta m^p(t,\boldsymbol{x})$ in a straightforward way (see Eq. 42 in body text). Then Eq. (B14) can be written:

$$\Delta\hat{s}^p(t,\boldsymbol{x}) = \int_t^{t_0} T^{\#}\left(t',\boldsymbol{x}\right)\left(\Delta\hat{s}^p\left(t',\boldsymbol{x}\right) + \sigma^{-2}\left(t',\boldsymbol{x}\right)\Delta m^p\left(t',\boldsymbol{x}\right)\right)(-dt'). \quad \text{(B17)}$$

where

$$\Delta m^p(t,\boldsymbol{x}) = m^*(t,\boldsymbol{x}) - \int_{t_0}^{t} T\left(t',\boldsymbol{x}\right)\left(m\left(t',\boldsymbol{x}\right) + s^p\left(t',\boldsymbol{x}\right)\right)dt', \quad \text{(B18)}$$

The symbols $\Delta\hat{s}^p(\boldsymbol{x},t)$ and $\sigma^{-2}(t,\boldsymbol{x})\,\Delta m^p(t,\boldsymbol{x})$ denote function equivalents of the vectors $\Delta\hat{\boldsymbol{S}}^p$ and $\mathbf{C}_m^{-1}\Delta\boldsymbol{M}^p$, respectively.

Thus, the steepest descent iterative solution written in the form of matrix expression in Eq. (40) can be replaced by its integral equivalent (Eqs. B17–B18). Further discussion is given is Sect. 2.5.

Appendix C

Application of the conjugated gradient method to the inversion, based on the adjoint transformation of the forward transport model

C1 *Basic formulation of the conjugated gradient method*

Let us formally write the linear system as follows:

$$\mathbf{A}\boldsymbol{x} = \boldsymbol{y}^*. \quad \text{(C1)}$$

Then, the solution of this system in the conjugated gradient method is given by the following iterative process:

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha_k\boldsymbol{p}_k, \quad \text{(C2)}$$

where

$$\alpha_k = \frac{\boldsymbol{p}_k^{\mathrm{T}}\nabla_k}{\boldsymbol{p}_k^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_k} = \frac{\nabla_k^{\mathrm{T}}\nabla_k}{\boldsymbol{p}_k^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_k}, \tag{C3}$$

and the gradient ∇_k is:

$$\nabla_x(\mathbf{x}^k) = \mathbf{A}\boldsymbol{x}_k = \mathbf{A}\boldsymbol{x}^k - \boldsymbol{y}^*. \tag{C4}$$

The vector $\boldsymbol{p}_k$ is determined as follows:

$$\boldsymbol{p}_k = \nabla_k + \beta\,\boldsymbol{p}_{k-1}, \tag{C5}$$

where

$$\beta = \frac{\nabla_k^{\mathrm{T}}\nabla_k}{\nabla_{k-1}^{\mathrm{T}}\nabla_{k-1}}. \tag{C6}$$

The initial condition for the iterative process is:

$$\boldsymbol{p}_0 = \nabla_0. \tag{C7}$$

C2 *Application of the forward transport model to inversion, using adjoint operators*

For inverting the aerosol mass transport model $\boldsymbol{M}$=$\mathbf{T}$ ($\boldsymbol{S}$ + $\boldsymbol{M}_0$) with the basic LSM, the following equation should be solved:

$$(\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T})\,\boldsymbol{S} = \mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{M}^*, \tag{C8}$$

where $\boldsymbol{M}^* = \boldsymbol{M}^{\mathrm{meas}}$–$\mathbf{T}\boldsymbol{M}_0$.

Similarly, for applying the conjugated gradient method, we can determine the matrix $\mathbf{A}$ and vector $\boldsymbol{y}$* as:

$$\mathbf{A} = \mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T}, \tag{C9}$$

$$\boldsymbol{y}^* = \mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{M}^*, \tag{C10}$$

$$\begin{aligned}\nabla_k &= \mathbf{A}\boldsymbol{S}^k - \boldsymbol{y}^* = (\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T})\boldsymbol{S}_k - \mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{M}^* \\ &= \mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{T}\boldsymbol{S}_k - \boldsymbol{M}^*) = \mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\Delta\boldsymbol{M}_k,\end{aligned} \tag{C11}$$

$$\boldsymbol{p}_k^{\mathrm{T}}\mathbf{A}\boldsymbol{p}_k = (\mathbf{T}\boldsymbol{p}_k)^{\mathrm{T}}\,\mathbf{C}^{-1}(\mathbf{T}\boldsymbol{p}_k) = \boldsymbol{p}_k^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T}\boldsymbol{p}_k \tag{C12}$$

Finally, we have the following procedure:

$$\boldsymbol{p}_k = \nabla_k + \beta_{k-1}\boldsymbol{p}_{k-1}, \tag{C13}$$

$$\boldsymbol{p}_{-1} = \mathbf{0} \text{ and } \beta_{-1} = \mathbf{0} \tag{C14}$$

$$\nabla_k = \mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1} \quad \Delta\boldsymbol{M}_k, \tag{C15}$$

where $\Delta\boldsymbol{M}_k$=$\mathbf{T}\boldsymbol{S}_k$–$\boldsymbol{M}^*$. This equation can be used only for the first iteration (when ∇_k can be calculated using Eq. C20).

$$\beta_{k-1} = \frac{\nabla_k^{\mathrm{T}}\nabla_k}{\nabla_{k-1}^{\mathrm{T}}\nabla_{k-1}}. \tag{C16}$$

$$\boldsymbol{S}^{k+1} = \boldsymbol{S}^k - \alpha_k\,\boldsymbol{p}_k, \tag{C17}$$

where

$$\alpha_k = \frac{\nabla_k^{\mathrm{T}}\nabla_k}{\boldsymbol{p}_k^{\mathrm{T}}\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T}\boldsymbol{p}_k}, \tag{C18}$$

The vector $\boldsymbol{b}_k$=$\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T}\boldsymbol{p}_k$ can be calculated using adjoint transformations:

$$b_k(t,\boldsymbol{x}) = \int_t^{t_0} T^{\#}(t',\boldsymbol{x})\left(b_k(t',\boldsymbol{x}) + \sigma^{-2}(t',\boldsymbol{x})\,\Delta g_k(t',\boldsymbol{x})\right)(-dt') \tag{C19}$$

where $b_k(t,\boldsymbol{x})$ are the components of the vector $\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T}\boldsymbol{p}_k$, and $\Delta g_k(t,\boldsymbol{x})$ are the components of the vector $\mathbf{T}\boldsymbol{p}_k$.

In addition, we can use the following equation:

$$\begin{aligned}\nabla_{k+1} &= \mathbf{A}\boldsymbol{S}^{k+1} - \boldsymbol{y}^* = \mathbf{A}(\boldsymbol{S}^k - \alpha_k\boldsymbol{p}_k) - \boldsymbol{y}^* = \mathbf{A}(\boldsymbol{S}^k - \boldsymbol{y}^*) \\ &\quad -\alpha_k\,\mathbf{A}\boldsymbol{p}_k = \nabla_k - \alpha_k\,\mathbf{A}\boldsymbol{p}_k.\end{aligned} \tag{C20}$$

Thus, for implementing this method, we need to run the transport model twice to calculate $\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T}\boldsymbol{p}_k$: once forward for $\mathbf{T}\boldsymbol{p}_k$ and once backward $\mathbf{T}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{T}\boldsymbol{p}_k$ (where $\boldsymbol{S}$=$\mathbf{T}\boldsymbol{p}_k$). A problem may appear if the transport operator does not allow use of negative sources; in this case we should always carry two terms $\boldsymbol{S}_k$=$\mathbf{T}\boldsymbol{p}_k$=$\boldsymbol{S}_k^{(+)}$+$\boldsymbol{S}_k^{(-)}$, and $\mathbf{T}\boldsymbol{S}$=$\mathbf{T}$($\boldsymbol{S}_k^{(+)}$+$\boldsymbol{S}_k^{(-)}$)=$\mathbf{T}\boldsymbol{S}_k^{(+)}$–$\mathbf{T}$(–$\boldsymbol{S}_k^{(-)}$).

Appendix D

List of acronyms and abbreviations

AERONET	Aerosol Robotic Network
APS	Aerosol Polarimetry Sensor
BC	Black Carbon
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
GEOS DAS	Goddard Earth Observing System Datda Assimilation System
GOCART	Goddart Chemistry Aerosol Radiation and Transport model
EARLINET	European Lidar Network
LSM	Least Squares Method
MISR	Multiangle Imaging SpectroRadiometer
MML	Method of Maximum Likelihood
MODIS	Moderate Resolution Imaging Spectrometer
OC	Organic Carbon
PDF	Probabillity Density Function
POLDER	Polarization and Directionality of the Earth's Reflectance

Acknowledgements. We thank the EOS Project Science Office for the support. We acknowledge support from NASA Radiation Science Program managed by H. Maring. We thank R. Levy for the help with MODIS data analysis, G. van der Werf and J. Collatz for assistance with obtaining global carbon emission map. We also thank L. Bounoua, and J. Collatz, T. Eck and V. Shcherbakov for reading early version of the manuscript and providing comments.

Edited by: T. Peter

References

Anderson, T. L., Wu, Y., Chu, D. A., Schmid, B., Redemann, J., and Dubovik, O.: Testing the MODIS satellite retrieval of aerosol fine-mode fraction, J. Geophys. Res., 110, D18204, doi:10.1029/2005JD005978, 2005.

Balkanski, Y. J., Jacob, D. J., Gardener, G. M., Graustein, W. C., and Turekian, K. K.: Transport and residence times of tropospheric aerosols inferred from a global 3-dimensional simulations of PB-210, J. Geophys. Res., 98 , 20 573–20 586, 1993.

Brasseur, G. P., Orlando, J. J., and Tyndall, G. S.: Atmospheric Chemistry and Global Change, Oxford University Press, 1st ed., 654 pp., 1999.

Boesenberg, J., Matthias, V., Amodeo, A., et al., EARLINET: A European Aerosol Research Lidar Network to Establish an Aerosol Climotology, Report 348, Max Planck Institute for Meteorology, Hamburg, Germany, 2003.

Cacuci, D. G.: Sensitivity theory for non-linear systems. I: Nonlinear functional analysis approach, J. Math. Phys., 22, 2794–2802, 1981.

Chin, M., Rood, R. B., Lin, S. J., Muller, J. F., and Thompson, A. M.: Atmospheric sulfur cycle simulated in the global model GOCART: Model description and global properties, J. Geophys. Res., 105(D20), 24 671–24 687, 2000.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements, J. Atmos. Sci., 59, 461–483, 2002.

Chin, M., Chu, D. A., Levy, R., Remer, L. A., Kaufman, Y. J., Holben, B. N., Eck, T., and Ginoux, P.: Aerosol distribution in the northern hemisphere during ACE-Asia: Results from global model, satellite observations, and sunphotometer measurements, J. Geophys. Res., 109, D23S90, doi:10.1029/2004JD004829, 2004.

Collins, W. D., Rasch, P. J., Eaton, B. E., Khattatov, B. V., Lamarque, J. F., and Zender, C. S.: Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, J. Geophys. Res., 106, 7313–7336, 2001.

Collins, W. D., Rasch, P. J., Eaton, B. E., Fillmore, D. W., Kiehl, J. T., Beck, C. T., and Zender, C. S.: Simulation of aerosol distributions and radiative forcing for INDOEX: Regional climate impacts, J. Geophys. Res., 107, 8028, doi:10.1029/2000JD000032, 2002.

Courtier, P. and Talagrand, O.: Variational assimilation of meteorological observations with the adjoint of the vorticity equations: Part II. Numerical results, Q. J. Roy. Meteor. Soc., 113, 1311–1328, 1987.

Dee, D. P. and Da Silva, A. M.: Data assimilation in the presence of forecast bias, Q. J. Roy. Meteor. Soc., 124(545), 269–295, Part A, 1998.

Deschamps, P. Y., Breon, F. M., Leroy, M., Podaire, A., Bricaud, A., Buries, J. C., and Seze, G.: The POLDER mission: Instrument characteristics and scientific objectives, IEEE Trans. Geosci. Remote Sens., 32, 598–615, 1994.

Deuzé, J. M., Breon, F. M., Devaux, C., Goloub, P., Herman, M., Lafrance, B., Maignan, F., Marchand, A., Nadal, F., Perry, G., and Tanré, D.: Remote sensing of aerosols over land surfaces from POLDER-ADEOS-1 polarized measurements, J. Geophys. Res., 106, 4913–4926, 2001.

Di Girolamo, L., Bond, T. C., Bramer D., Diner, D.J., Fettinger F., Kahn, R. A., Martonchik, J. V., Ramana, M. V., Ramanathan, V., and Rasch, P. J.: Analysis of Multi-angle Imaging SpectroRadiometer (MISR) aerosol optical depths over greater India during winter 2001–2004, Geophys. Res. Lett., 31, L23115, doi:10.1029/2004GL021273, 2004.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R. A., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Muller, J. P., Myneni, R. B., Sellers, P. J., Pinty, B., and Verstraete, M.: Multi-angle Imaging SpectroRadiometer (MISR) instrument description and experiment overview IEEE Trans. Geosci. Remote Sens., 36, 1072–1087, 1998.

Dubovik, O. and King, M. D.: A flexible inversion algorithm for retrieval of aerosol optical properties from sun and sky radiance measurements, J. Geophys. Res., 105, 20 673–20 696, 2000.

Dubovik, O., Holben, B. N., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanré, D., and Slutsker, I.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, J. Atmos. Sci., 59, 590–608, 2002.

Dubovik, O.: Optimization of Numerical Inversion in Photopolarimetric Remote Sensing, in: Photopolarimetry in Remote Sensing, edited by: Videen, G., Yatskiv, Y., and Mishchenko, M., Kluwer Academic Publishers, Dordrecht, Netherlands, 65–106, 2004.

Edie, W. T., Dryard, D., James, F. E., Roos, M., and Sadoulet, B.: Statistical Methods in Experimental Physics, North-Holland Publishing Company, Amsterdam, 155 pp., 1971.

Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modeling, J. Geophys. Res., 102, 15 967–15 985, 1997.

Elbern, H., Schmidt, H., Talagrand, O., and Ebel, A.: 4D-variational data assimilation with an adjoint air quality model for emission analysis, Environ. Modell. Software, 15, 539–548, 2000.

Elbern, H. and Schmidt, H.: Ozone episode analysis by four-dimensional variational chemistry data assimilation, J. Geophys. Res., 106, 3569–3590, 2001.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749-3769, 2007

Enting, I. G., Trudinger, C. M., and Francey, R. J.: A synthesis inversion of the concentration and d13C of atmospheric CO2, Tellus B, 47, 35–52, 1995.

Gill, P. E., Murray, W., and Wright, M. E.: Practical optimization, Academic Press, London, p. 401, 1982.

Ginoux, P., Chin, M., Tegen, I., Prospero, J., Holben, B. N., Dubovik, O., and Lin, S. J.: Sources and distributions of dust

aerosols simulated with the GOCART model, J. Geophys. Res., 106, 20 255–20 274, 2001.

Ghan, S., Easter, R., Chapman, E., Abdul-Razzak, H., Zhang, Y., Leung, L., Laulainen, N., Saylor, R., and Zaveri, R.: A physically based estimate of radiative forcing by anthropogenic sulfate aerosol, J. Geophys. Res., 106, 5279–5293, 2001a.

Ghan, S., Laulainen, N., Easter, R., Wagener, R., Nemesure, S., Chapman, E., Zhang, Y., and Leung, R.: Evaluation of aerosol direct forcing in MIRAGE, J. Geophys. Res., 106, 5295–5316, 2001b.

Hakami, A., Henze, D. K., Seinfeld, J. H., Chai, T., Tang, Y., Carmichael, G. R., and Sandu, A.: Adjoint inverse modeling of black carbon during the Asian Pacific Regional Aerosol Characterization Experiment, J. Geophys. Res., 110, D14301, doi:10.1029/2004JD005671, 2005.

Hartley, D. and Prinn, R.: Feasibility of determining surface emissions of trace gases using an inverse method in a 3-dimentional chemical-transport model, J. Geophys. Res., 98, 5183–5197, 1993.

Hsu, N. C., Tsay, S. C., King, M. D., and Herman, J. R.: Aerosol properties over bright-reflecting source regions, IEEE Trans. Geosci. Remote Sens., 42, 557–569, 2004.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, 1998.

Hourdin, F. and Talagrand, O.: Eulerian backtracking of atmospheric traces: I Adjoint derivation and parametrisation of subgrid-scale transport, Quart. J. Am. Meteorol. Soc., 132, 567–583, 2006.

Houweling, S., Breon, F.-M., Aben, I., Rödenbeck, C., Gloor, M., Heimann, M., and Ciais, P.: Inverse modeling of CO_2 sources and sinks using satellite data: a synthetic inter-comparison of measurement techniques and their performance as a function of space and time, Atmos. Chem. Phys., 4, 423–538, 2004, http://www.atmos-chem-phys.net/4/423/2004/.

Jacob, D. J.: Introduction to Atmospheric Chemistry, Princeton University Press, 266 pp., 1999.

Kahn, R. A., Gaitley, B. J., Martonchik, J. V., Diner, D. J., and Crean, K. A.: Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations, J. Geophys. Res., 110, D10S04, doi:10.1029/2004JD004706, 2005.

Kaminski, T., Heimann, M., and Giering, R.: A coarse grid three dimensional global inverse model of the atmospheric transport, 1, Adjoint model and Jacobian matrix, J. Geophys. Res., 104, 18 535–18 553, 1999a.

Kaminski, T., Heimann, M., and Giering, R.: A coarse grid threedimensional global inverse model of the atmospheric transport, 2, Inversion of the transport of CO2 in the 1980s, J. Geophys. Res., 104, 18 555–18 581, 1999b.

Kasibhatla, P. S., Heimann, M., Rayner, P., Mahowald, N., Prinn, R. G., Hartley, D. E.: Inverse Methods in Global Biogeochemical Cycles, American Geophysical Union, 324 pp., 2000.

Kalman, R. E.: A New approach to linear filtering and prediction problems, J. Basic. Eng., 82, 35–40, 1960.

Kaufman, Y. J., Tanré, D., Remer, L. A., et al.: Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer, J. Geophys. Res., 102, 17 051–17 067, 1997.

Kaufman, Y. J., Tanré, D., and Boucher, O.: A satellite view of aerosols in the climate system, Nature, 419(6903), 215–223, 2002.

King, M. D., Byrne, D. M., Herman, B. M., and Reagan, J. A.: Aerosol size distributions ob-tained by inversion of spectral optical depth measurements, J. Atmos. Sci., 21, 2153–2167, 1978.

King, M. D., Kaufman, Y. J., Tanré, D., and Nakajima, T.: Remote sensing of tropospheric aerosols from space: Past, present, and future, B. Am. Meteorol. Soc., 80, 2229–2259, 1999.

Kinne, S., Lohmann, U., Feichter, J., Schulz, M., Timmreck, C., Ghan, S., Easter, R., Chin, M., Ginoux, P., Takemura, T., Tagen, I., Koch, D., Herzog, M., Penner, J., Pitari, G., Holben, B., Eck, T., Smirnov, A., Dubovik, O., Slutsker, I., Tanré, D., Torres, O., Mishchenko, M., Geogdzhayev, I., Chu, D. A., and Kaufman, Y.: Monthly averages of aerosol properties: A global comparison among models, satellite data and AERONET ground data, J. Geophys. Res., 108, 4634, doi:10.1029/2001JD001253, 2003.

Kinne, S., Schulz, M., Textor, C., et al.: An AeroCom initial assessment – optical properties in aerosol component modules of global models, Atmos. Chem. Phys., 6, 1815–1834, 2006, http://www.atmos-chem-phys.net/6/1815/2006/.

Khattatov, B., Lamarque, J.-F., Lyjak, L. V., Menard, R., Levelt, P., Tie, X., Brasseur, G. P., and Gille, J. C.: Assimilation of satellite observations of long-lived chemical species in global chemistry-transport models, J. Geophys. Res., 105, 29 135–29 144, 2000.

Koch, D.: The transport and direct radiative forcing of carbonaceous and sulfate aerosol in the GISS GCM, J. Geophys. Res., 106, 20 311–20 332, 2001.

Koch, D., Jacob, D., Tegen, I., Rind, D., and Chin, M.: Tropospheric sulfur simulation and sulfate direct forcing in the Goddard Institute for Space Studies (GISS) general circulation model, J. Geophys. Res., 104, 23 799–23 822, 1999.

Labonne, M., Breon, F-M., and Chevallier, F.: Injection height of biomass burning aerosols as seen from a spaceborne lidar, Geophys. Res. Lett., 34, L11806, doi:10.1029/2007GL029311, 2007.

Lahoz, W. A., Errera, Q., Swinbank, R., and Fonteyn, D.: Data assimilation of stratospheric constituents: a review, Atmos. Chem. Phys., 7, 5745-5773, 2007

Le Dimet, F. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, Tellus A, 38, 97–110, 1986.

Lin, S.-J. and Rood, R. B.: Multidimensional flux-form semi-Lagrangian transport schemes, Mon. Wea. Rev., 124, 2046–2070, 1996.

Marchuk, G. I.: Method of numerical mathematics, Nauka, Moscow, 1977 (in Russian).

Marchuk, G.: Mathematical Models in Environmental Problems, Elsevier, New York, 1986.

Menut, L., Vautard, R., Beekmann, M., and Honoré, C.: Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model, J. Geophys. Res., 105, 15 379–15 402, 2000.

Menut, L.: Adjoint modeling for atmospheric pollution process sensitivity at regional scale, J. Geophys. Res., 108, 8562, doi:10.1029/2002JD002549, 2003.

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.

Mishchenko, M. I., Cairns, B., Hansen, J. E., Travis, L. D., Burg, R., Kaufman, Y. J., Martins, J. V., and Shettle, E. P.: Monitoring of aerosol forcing of climate from space: analysis of measurement requirements, J. Quant. Spetrosc. Radiat. Transfer, 88, 149–161, 2004.

Mishchenko, M. I., Cairns, B., Kopp, G., Schueler, C. F., Fafaul, B. A., Hansen, J. E., Hooker, R. J., Itchkawich, T., Merimg, H. B., and Travis, L. D.: Accurate monitoring of terrestrial aerosola and total solar irradiance: introducing the Glory, Mission, B. Am. Meteorol. Soc., 88, N5, 677–691, 2007

Nakajima, T., Tonna, G., Rao, R., Boi, P., Kaufman, Y., and Holben, B.: Use of sky bright-ness measurements from ground for remote sensing of particulate polydisper-sions, Appl. Opt., 35, 2672–2686, 1996.

Navon, I. M.: Practical and theoretical aspects for adjoint parameter estimation and identifiability in meteorology and oceanography, Dyn. Atmos. Oceans, 27, 55–79, 1997.

Patra, P. K., Maksyutov, S., Sasano, Y., Nakajima, H., Inoue, G., and Nakazawa, T.: An evaluation of CO2 observations with Solar Occultation FTS for Inclined-Orbit Satellite sensor for surface source inversion, J. Geophys. Res., 108, 4759, doi:10.1029/2003JD003661, 2003.

Phillips, B. L.: A technique for numerical solution of certain integral equation of first kind, J. Assoc. Comp. Mach., 9, 84–97, 1962.

Poole, L. R., Winker, D. M., Pelon, J. R., and McCormick, M. P.: CALIPSO: global aerosol and cloud observations from lidar and passive instruments, Proc. SPIE 4881, 419–426, 2003.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes in FORTRAN. The art of Scientific Computing, Cambridge University Press, 965 pp., 1992.

Rao, C. R.: Linear Statistical Inference and Its Applications, Wiley, New York, 500 pp., 1965.

Reddy, M. S. and Boucher, O.: A study of the global cycle of carbonaceous aerosols in the LMDZT general circulation model, J. Geophys. Res., 109, D14202, doi:10.1029/2003JD004048, 2004

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products and Validation, J. Atmos. Sci., 62, 947–973, 2005.

Remer, L. A., Kaufman, Y. J., and Kleidman, R. G.: Comparison of Three Years of Terra and Aqua MODIS Aerosol Optical Thickness Over the Global Oceans, IEEE Geosci. Remote Sens. Lett., 3, 537–540, 2006.

Rodenbeck, C., Houweling, S., Gloor, M., and Heimann, M.: CO2 flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric trace transport, Atmos. Chem. Phys., 3, 1919–1964, 2003, http://www.atmos-chem-phys.net/3/1919/2003/.

Rodgers, C. D.: Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation, Rev. Geophys. Space Phys., 14, 609–624, 1976.

Roeckner, E., Arpe, K., Bengtsson, L., Christoph, M., Claussen, M., Duemenil, L., Esch, M., Giorgetta, M., Schlese, U., and Schulzweida, U.: The atmospheric general circulation model ECHAM-4: Model description and simulation of present-day climate, Tech. Rep. 218, Max-Planck-Inst. für Meteorol., Hamburg, Germany, 1996.

Sato, M., Hansen, J., Koch, D., Lacis, A., Ruedy, R., Dubovik, O., Holben, B., Chin, M., and Novakov, T.: Global atmospheric black carbon inferred from AERONET, Proc. Nat. Acad. Sci., 100(11), 6319–6324, 2003.

Schmidt, H. and Martin, D.: Adjoint sensitivity of episodic ozone in Paris area to emissions on the continental scale, J. Geophys. Res., 108, 8561, doi:10.1029/2001JD001583, 2003.

Smirnov A., Holben, B. N., Dubovik, O., Frouin, R., Eck, T. F., and Slutsker, I.: Maritime component in aerosol optical models derived from Aerosol Robotic Network data, J. Geophys. Res., 108, 4033, doi:10.1029/2002JD002701, 2003.

Strand, O. N. and Westwater, E. R.: Statistical estimation of the numerical solution of a Fredholm internal equation of the first kind, J. Assoc. Comput. Mach., 15, 104–114, 1968.

Takemura, T., Okamoto, H., Maruyama, Y., Numaguti, A., Hiragushi, A., and Nakajima, T.: Global three dimensional simulation of aerosol optical thickness distribution of various origins, J. Geophys. Res., 105, 17 853–17 873, 2000.

Takemura, T., Nakajima, T., Dubovik, O., Holben, B., and Kinne, S.: Single scattering albedo and radiative forcing of various aerosol species with a global three-dimensional model, J. Climate, 4, 333–352, 2002.

Thacker, W. C.: Fitting models to inadequate data by enforcing spatial and temporal smoothness, J. Geophys. Res., 93, 10 655–10 665, 1988.

Thacker, W. C. and Long, R. B.: Fitting dynamics to data, J. Geophys. Res., 93(C2), 1227–1240, 1988.

Talagrand, O.: A study of the dynamics of four dimensional data assimilation, Tellus, 33, 43–60, 1981a.

Talagrand, O.: On the mathematics of data assimilation, Tellus, 33, 321–339, 1981b.

Talagrand, O. and Courtier, P.: Variational assimilation of meteorological observations with the adjoint of the vorticity equations: Part I., Theory, Q. J. Roy. Meteor. Soc., 113, 1311–1328, 1987.

Tarantola, A.: Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation, Elsevier, Amsterdam, 614 pp., 1987.

Tanré, D., Herman, M., and Kaufman, Y. J.: Information on aerosol size distribution contained in solar reflected spectral radiances, J. Geophys. Res., 101, 19 043–19 060, 1996.

Tanré, D., Kaufman, Y. J., Herman, M., and Mattoo, S.: Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances, J. Geophys. Res., 102, 16 971–16 988, 1997.

Tegen, I., Hollrig, P., Chin, M., Fung, I., Jacob, D., and Penner, J.: Contribution of different aerosol species to the global aerosol extinction optical thickness: Estimates from model results, J. Geophys. Res., 102, 23 895–23 915, 1997.

Tegen, I., Koch, D., Lacis, A., and Sato, M.: Trends in tropospheric aerosol loads and corresponding impact on direct radiative forcing between 1950 and 1990: A model study, J. Geophys. Res., 105, 26 971–26 989, 2000.

Textor, C., Schulz, M., Guibert, S., et al.: Analysis and quantification of the diversities of aerosol life cycles within AeroCom, Atmos. Chem. Phys., 6, 1777–1813, 2006, http://www.atmos-chem-phys.net/6/1777/2006/.

Tikhonov, A. N.: On the solution of incorrectly stated problems and a method of regu-larization, Dokl. Akad. Nauk SSSR, 151, 501–

504, 1963.

Twomey, S.: On the numerical solution of Fredholm integral equations of the first kind by the inversion of the linear system produced by quadrature, J. Assoc. Comp. Mach., 10, 97–101, 1963.

Van der Werf, G. R., Randerson, J. T., Collatz, G. J., Giglio, L., Kasibhatla, P. S., Arellano, A. F., Olsen, S. C., and Kasischke, E. S.: Continental-scale partitioning of fire emissions during the 1997 to 2001 El Nino/La Nina period, Science, 303(5654), 73–76, 2004.

Vautard, R., Beekmann, M., and Menut, L.: Applications of adjoint modelling in atmospheric chemistry: Sensitivity and inverse modeling, Environ. Model. Software, 15, 703–709, 2000.

Vukicevic, T. and Hess, P.: Analysis of tropospheric transport in the Pacific Basin using the adjoint technique, J. Geophys. Res., 105, 7213–7230, 2000.

Vukicevic, T., Steyskal, M., and Hecht, M.: Properties of advection algorithms in the context of variational data assimilation, Mon. Wea. Rev., 129(5), 1221–1231, 2001.

Weaver, C., da Silva, A., Chin, M., Ginoux, P., Dubovik, O., Flittner, D., Zia, A., Remer, L., Holben, B. N., and Gregg, W.: Direct Insertion of MODIS Radiances in a Global Aerosol Transport Model, J. Atmos. Sci., 64(3), 808–827, 2007.

Yaremchuk, M. L., D. A. Nechaev, K.R. Thompson, Seasonal variation of the North Atlantic Current, J. Geophys. Res., 106, 6835–6852, doi:101029/2000JC900166, 2001

Yaremchuk, M. L. and Maximenko, N. A.: A dynamically consistent analysis of the mesoscale eddy field at western North Pacific Subarctic Front, J. Geophys. Res., 107, 3223, doi:10.1029/2002JC001379, 2002