J. Geophys. Res., 104 (D15), 18535-18553, 1999

A coarse grid three-dimensional global inverse model of the atmospheric transport
1. Adjoint model and Jacobian matrix

Thomas Kaminski and Martin Heimann

Max-Planck-Institut für Meteorologie, Hamburg, Germany

Ralf Giering[1]

Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge

Abstract. TM2 is a global three-dimensional model of the atmospheric transport of passive tracers. The adjoint of TM2 is a model that allows the efficient evaluation of derivatives of the simulated tracer concentration at observational locations with respect to the tracer's sources and sinks. We describe the generation of the adjoint model by applying the Tangent linear and Adjoint Model Compiler in the reverse mode of automatic differentiation to the code of TM2. Using CO_2 as an example of a chemically inert tracer, the simulated concentration at observational locations is linear in the surface exchange fluxes, and thus the transport can be represented by the model's Jacobian matrix. In many current inverse modeling studies, such a matrix has been computed by multiple runs of a transport model for a set of prescribed surface flux patterns. The computational cost has been proportional to the number of patterns. In contrast, for differentiation in reverse mode, the cost is independent of the number of flux components. Hence, by a single run of the adjoint model, the Jacobian for the approximately 8° latitude by 10° longitude horizontal resolution of TM2 could be computed efficiently. We quantify this efficiency by comparison with the conventional forward modeling approach. For some prominent observational sites, we present visualizations of the Jacobian matrix by series of illustrative global maps quantifying the impact of potential emissions on the concentration in particular months. Furthermore, we demonstrate how the Jacobian matrix is employed to completely analyze a transport model run: A simulated monthly mean value at a particular station is decomposed into the contributions to this value by all flux components, i.e., the fluxes into every surface model grid cell and month. This technique also results in a series of global maps.

1. Introduction

The radiative balance of the terrestrial atmosphere is sensitive to the concentrations of a number of trace gases. Enhanced concentrations of these greenhouse gases may thus lead to climate change. This sensitivity of climate to perturbations in the concentrations of greenhouse gases is being estimated by means of complex General Circulation Models [*Watson et al.*, 1995]. For predictions of climate change and its impacts, these models use the greenhouse gas concentrations as boundary condition. To control the temporal development of these concentrations, in turn, the sources and sinks of the respective gases have to be predicted over

[1] Now at Jet Propulsion Laboratory, Pasadena, California.

the time period of interest. Hence reliable models of the underlying source and sink processes are urgently needed to determine the feedbacks of future climate changes on the concentration of the gases. Improving our knowledge about the past and current source and sink magnitudes would help to improve and verify these process models.

At present, however, for many greenhouse gases such as carbon dioxide (CO_2), carbon monoxide (CO), methane (CH_4), or nitrous oxide (N_2O) not even the current magnitudes of the natural as well as the anthropogenic sources and sinks can be quantified with sufficient accuracy [*Houghton et al.*, 1995]. Especially for CO_2 and CH_4, there have been considerable efforts to measure directly the exchange fluxes between the atmosphere and different source reservoirs (over oceans, e.g., by global ship campaigns or over land by means of eddy correlation flux measurements). Although this "bottom up" approach locally yields important information on the relevant processes, large uncertainties are induced by the necessary assumptions for extrapolation to regional or global scales.

During the last decades, an observational network of increasing density is being established to monitor the composition of the atmosphere. Space-borne observations are also becoming available, as well as measurements on board of ships and air planes. In contrast to local flux measurements, if carefully selected, the atmospheric concentration observations are representative for larger spatial scales. Hence such observations provide a means of estimating the sources and sinks on larger scales. Yet, in order to link the surface fluxes to the atmospheric concentration observations, a more or less sophisticated model of the atmospheric transport is needed, possibly in conjunction with a module of the relevant atmospheric chemistry for the species under consideration. The systematic search for spatiotemporal flux fields that, in combination with an atmospheric transport model, yield modeled concentrations close to observations is called inverse modeling of the atmospheric transport.

In order to alleviate future climate change, international negotiations are currently underway to curb the emissions of several of the greenhouse gases In this context, another perspective for inverse modeling is to derive regional estimates of the source fluxes to monitor the success of these attempts.

A number of research groups have investigated the feasibility of inversion of the atmospheric transport. The challenge consists in using the information from a spatially sparse observational network in an optimal way to derive regional flux estimates together with an estimated range of confidence. Technically, this constitutes an underdetermined or "ill-posed" inverse problem: A unique solution can only be derived by use of additional assumptions (regularization of the inverse problem). The validity of these assumptions as well as the reliability of the transport model are crucial for the quality of the resulting estimates. Recently, a number of studies have been carried out to quantify the magnitude of the sources and sinks of CO_2 [*Enting and Mansbridge*, 1989; *Enting et al.*, 1995; *Ciais et al.*, 1995; *Haas-Laursen*, 1997; *Rayner et al.*, 1999; *Bousquet*, 1997; *Law*, 1999], CH_4 [*Brown*, 1993; *Hein and Heimann*, 1994; *Brown*, 1995; *Hein et al.*, 1996], and halocarbons [*Brown*, 1993; *Hartley and Prinn*, 1993]. Differences among these studies mainly consist in the resolution of the transport models (two dimensional or three dimensional) and in the kind of assumptions for regularization, which is formally reflected by different inversion techniques [see, e.g., *Enting*, 1999].

Most of the relevant long-lived trace gases are either not (CO_2) or only weakly (CH_4, N_2O, halocarbons) coupled to tropospheric chemistry and thus, in a good approximation, can be inverted with a linearized representation of the transport. The transport then can be taken into account in the following way: The surface flux field is decomposed into prescribed spatiotemporal patterns ("source" or "flux" components) with unknown scaling coefficients. The transport model is separately run with each of the source components, and the contributions to the concentration signal at each of the monitoring sites and times are recorded. These contributions can be interpreted as a discretized "impulse response" or "Greens function" that quantifies the response of the modeled concentration at the observational sites and time periods to unit changes in the magnitude of each source component.

Formally, this impulse response or Greens function is the Jacobian matrix representing the first derivative of the modeled concentration at the observational sites and dates with respect to the coefficients of the source components. Computationally, for n_f source components, n_f model runs (or a single n_f tracer run transporting emissions from each source component separately) have to be performed to determine the n_f differential quotients constituting the columns of the Jacobian matrix. Hence the number of source components that can be considered is essentially limited by the computational cost of the necessary transport model runs, i.e., by the model's complexity in terms of both the representation of the transport and possibly chemical processes and their spatiotemporal resolution. The additional assumption that the flux fields can be represented by a few patterns is thus inherent in this approach and, in part, determines the result of the inversion, because the internal shape of these patterns cannot be altered by the inversion. It is evident, though, that for many trace gases, such a restricted representation does not take account of the full spatiotemporal variability in an appropriate way. Further, in combination with inhomogeneous sampling (which for sparse networks is inavoidable), this low resolution in the space of unknowns may lead to biased

estimates as recently investigated by *Trampert and Snieder* [1996]: For example, to adjust a well-observed fraction of a fixed prescribed pattern, the inversion algorithm must adjust the pattern's scaling factor, and the accompanying change in the pattern's less well-observed fraction can spoil the estimate of the pattern's integrated emissions.

Here we present an alternative approach employing the adjoint of the three-dimensional transport model TM2. By means of the Tangent linear and Adjoint Model Compiler (TAMC, R. Giering, 1997, available at http://puddle.mit.edu/~ralf/tamc) this numerical module has been constructed automatically from the TM2 source code in the "reverse mode" of automatic differentiation. The principles of adjoint code generation and the adjoint model are introduced in section 3. Unlike most adjoint models applied in geosciences, which are constructed for iterative minimization of scalar valued functions, the adjoint of TM2 computes the derivative of a vector valued function. Hence, by a single run of the adjoint model the exact Jacobian is efficiently computed row by row, for which the cost is proportional to the number of observations and nearly independent of the number of flux components. Hence, defining the flux patterns as the model grid cells, we are able to determine the Jacobian for the horizontal TM2 resolution of approximately $8^o \times 10^o$ and monthly temporal resolution.

The Jacobian is computed for the simulation of the quasi-stationary seasonal cycle of CO_2, which is carried out in a cyclostationary setup of TM2 described in section 2. The rows of the Jacobian quantify the sensitivity of the modeled concentration at a particular station and month to the fluxes into every surface layer grid cell at every month. A visualization results in instructive maps of the potential influence of the flux components for the respective months on a particular observable. Prescribing for each grid cell the relative distribution of the fluxes over the year (e.g., constant flux), the information on potential influence can be condensed to one map for each monthly mean concentration. On the other hand, it is possible to derive the sensitivity of any particular feature that can be computed from the monthly mean concentrations (e.g., the yearly mean concentration, or the magnitude of the seasonal cycle). For linear combinations of the monthly mean concentrations, in addition to compute potential influence areas, it is possible to decompose the feature as modeled in a particular run according to the contributions resulting from the respective flux components. Besides these sensitivity studies, the Jacobian can be applied for tracer simulations instead of TM2 [*Knorr*, 1997], as long as the setup the matrix has been derived for is appropriate for the problem at hand. In a companion paper [*Kaminski et al.*, this issue], we present a Bayesian inversion on the TM2 grid, in which we combine the Jacobian with both atmospheric CO_2 observations and a priori information on the fluxes.

In summary, the outline is as follows: In section 2 we give a description of the transport model and the setup for which we derive the matrix representation. The principles of adjoint code generation and the adjoint model are introduced in section 3. Section 4 discusses the Jacobian and its use to compute sensitivities of particular features. Section 5 contains concluding remarks.

2. Model of the Quasi–Stationary Seasonal Cycle

A statistical analysis of the observed atmospheric CO_2 concentrations as performed, e.g., by *Keeling et al.* [1989] points out that, on timescales of a few years, the concept of a quasi-stationary seasonal cycle is appropriate to describe the prevailing features in the records. This quasi-stationary seasonal cycle component in the concentration, which essentially is composed of a global trend and a spatially varying seasonal cycle, can be extracted from the observations as well as be simulated by atmospheric transport models. Since these transport models use CO_2 surface exchange flux fields as boundary condition, comparison of the observed and the simulated quasi-stationary seasonal cycles provides a way to constrain these fluxes. In this section we briefly introduce our transport model TM2, give a formal definition of the quasi-stationary seasonal cycle, and describe an appropriate setup of TM2 for simulation of the quasi-stationary seasonal cycle. The adjoint model, which is described in section 3, then evaluates the derivative of the function that is defined by this particular setup. Comparison of simulated concentrations to observations is deferred to section 7 of *Kaminski et al.* [this issue].

TM2 is a three-dimensional atmospheric transport model, which solves the continuity equation for an arbitrary number of atmospheric tracers on an Eulerian grid spanning the entire globe [*Heimann*, 1995]. It is driven by stored meteorological fields derived from analyses of a weather forecast model or from output of an atmospheric general circulation model. Tracer advection is calculated using the "slopes scheme" of *Russel and Lerner* [1981]. Vertical transport due to convective clouds is computed using the cloud mass flux scheme of *Tiedtke* [1989]. Turbulent vertical transport is calculated by stability dependent vertical diffusion according to the scheme by *Louis* [1979]. Numerically, in each base time step the model calculates the source and sink processes affecting each tracer, followed by the calculation of the transport processes.

The spatial structure of the model is a regular latitude-longitude grid and a sigma coordinate system in the vertical.

The base "coarse grid" version of the model uses a horizontal resolution of approximately 8° latitude by 10° longitude (the horizontal dimension of the grid is $n_g = 36 \times 24$) and 9 layers in the vertical dimension. The numerical time step of this model version is 4 hours.

We apply TM2 to simulate the quasi-stationary seasonal cycle component in the CO_2 concentration at particular observational sites. Therefore, prescribing the same monthly mean surface exchange flux fields f each year (cyclostationarity), and starting from zero initial concentration, TM2 is run by repeatedly cycling through the same meteorological fields of the year 1987 derived from analyses of the European Center for Medium Range Weather Forecast (ECMWF), which are available to the model every 12 hours. These meteorological fields have been adjusted in order to guarantee air mass conservation. This adjustment is also applied when switching from the fields of December 31 to January 1 [*Heimann*, 1995]. We use monthly mean values of the simulated concentration for comparison with observations, because for shorter averaging periods the influence of synoptic events, whose interannual variations are not resolved, would become too important. To extract time series of concentrations c_S at particular sites S, we first compute monthly means and then perform a bilinear interpolation in the horizontal from the TM2 grid to the exact location of S.

With periodic boundary conditions and periodic transport, at every site, the simulated concentration as well tends towards a periodic state c_p. For a flux field with nonzero global annual mean, however, a linear trend is superimposed on the cyclostationary concentrations. The spatial variation of the magnitude of the annual mean flux as well as the effect of covarying seasonal cycles of fluxes and transport (rectifier effect) described, e.g., by *Pearman and Hyson* [1980], *Heimann et al.* [1986], *Heimann and Keeling* [1989], and *Denning et al.* [1995] result in a spatially varying offset in c_p. Formally, at the ith month, the simulated concentration $c_{S,i}$ can be composed as

$$c_{S,i} \quad = \quad S_{S,i} + b \cdot t_i + a_S + R_{S,i}, \tag{1}$$

where the single terms have the following meaning: The periodic component has been split up into a function $S_{S,i}$ with yearly period ($S_{S,i+12} = S_{S,i}$) and zero annual mean denoting the seasonal cycle as well as the spatial gradient contribution a_S. The long-term global linear trend b is related to the global annual mean flux $\bar{f}$ by

$$b = \alpha \cdot \bar{f}, \tag{2}$$

where $\alpha = 0.476$ ppmv GtC^{-1} is the conversion factor from mass to concentration for instantaneous global mixing as used by the transport model. The length of the time interval from the beginning of the simulation to the middle of the ith month t_i is given by

$$t_i := \frac{(i - {}^1\!/_2)}{12}\, a. \tag{3}$$

The residuum $R_{S,i}$ tends to zero as the length of the time series increases.

We define the quasi-stationary seasonal cycle as

$$c_{S,i} - R_{S,i} = S_{S,i} + b \cdot t_i + a_S = c_{p,S,i} + b \cdot t_i. \tag{4}$$

To represent the quasi-stationary seasonal cycle, in addition to the global linear trend, 12 numbers per site are needed to quantify c_p: 11 numbers for S_S (the 12th monthly value is determined by the other 11, because the sum must be 0) and 1 number for a_S. As soon as $R_{S,i}$ is close enough to zero to be neglected, the quasi-stationary seasonal cycle can be extracted from our modeled time series. *Heimann and Keeling* [1989] found that for tropospheric sites a spin up period of 3 years is sufficient to achieve an appropriate degree of convergence in (4). The rate of convergence reflects the model's timescales of mixing. These timescales are commonly quantified in terms of exchange times [*Rayner and Law*, 1995; *Law et al.*, 1996]. More precisely, the rate of convergence is determined by the longest exchange time, which, in the troposphere, is associated to the interhemispheric transport. Using the radioactive tracer ^{85}Kr, *Jacob et al.* [1987] found an interhemispheric exchange time of 1.1 years for a similar transport model, and *Heimann and Keeling* [1989] found 1.3 years for TM2. Similar to *Heimann and Keeling* [1989] as "standard setup" of TM2, we choose to perform a four year run, of which we extract the monthly mean concentrations in the last year. Together with the global annual mean flux, these 12 values per site determine the trend and the periodic component representing the quasi-stationary seasonal cycle:

$$\begin{aligned} b &= \alpha \cdot \bar{f} \\ c_{p,S,i} &= c_{S,i+3\cdot 12} - t_{i+3\cdot 12} \cdot \alpha \cdot \bar{f} \qquad (i = 1, 12). \end{aligned} \tag{5}$$

In the terminology of linear algebra, the standard setup includes the choice of a basis (and its order) for the space of fluxes, i.e., a set of $n_f = 12 \times n_g$ vectors spanning the space, and $f \in I\!R^{n_f}$ is a representation of a particular flux vector by its components with respect to that basis. The components of f quantify the 12 monthly mean fluxes into each surface grid cell. In particular, the basis defines the physical units of the fluxes. Similarly, with respect to a basis in the space of concentrations, the output $c \in I\!R^{n_c}$ is a vector of $n_c = 12 \times n_s$ components for the modeled monthly mean concentration at n_s observational sites. Since, in addition, every step in the simulation is linear, in the standard setup TM2 can be represented by a real $n_c \times n_f$ matrix T, and the application of the model to a flux field f can be written as

$$c = Tf. \tag{6}$$

Using this matrix notation, the model of the quasi-stationary seasonal cycle in (5) reads

$$\begin{aligned} b &= \alpha \cdot \bar{f} \\ c_p &= Tf - t \cdot \alpha \cdot \bar{f}, \end{aligned} \tag{7}$$

where the vector t contains the values of t_i.

Concatenating b and c_p to one vector c_{qsc}, these equations define a single matrix M:

$$c_{qsc} =: Mf. \tag{8}$$

Since our model neglects interannual variations in the transport as well as in the fluxes, a careful interpretation of c_{qsc} is necessary: If it was interpreted as the quasi-stationary seasonal cycle of 1987, the year of the meteorological data, c_{qsc} would be subject to both sources of error: For the spin-up years the difference in the meteorologies to 1987 as well as the differences in the fluxes to 1987 would be neglected. Instead, as in the study of *Hein et al.* [1996], c_{qsc} should be interpreted as a mean quasi-stationary seasonal cycle over a target period of a few years: Prescribing the mean flux over the whole target period, the error caused by the cyclo-stationary flux assumption decreases with increasing length of the target period. The error induced by using the meteorology of a particular year to simulate the whole target period still remains. One might argue that a climatology, i.e., the meteorology of a mean year, should be used instead. In order not to underestimate the transport, however, TM2 needs the synoptic scale variation, which is partly removed by the averaging procedure yielding the climatology. Hence, instead of using a mean meteorology, c_{qsc} is interpreted as one particular element of the ensemble of modeled concentrations that would result from using the same mean fluxes but the meteorologies from the particular years of the target period. This model error has to be taken into account when comparing c_{qsc} to the mean quasi-stationary seasonal cycle extracted from observations. Recent studies indicate that for monthly mean concentrations, this error is not too large: *Knorr and Heimann* [1995] investigated the impact of the meteorological data by comparing the seasonal cycle of the monthly mean concentration simulated with TM2 in the standard setup driven by the meteorology either of 1986 or 1987. In their study they obtain only a minor difference. With a different model, *Law and Simmonds* [1996] explored the sensitivity of fluxes resulting from an inversion to the year of the meteorological fields. They also found small differences. In section 9 of *Kaminski et al.* [this issue] these results are confirmed by a comparison of the flux fields inferred from two inversions that we perform on the basis of meteorological data from 1986 and 1987.

3. The Adjoint Model

As explained in section 2, for the standard setup, TM2 can be represented by a $n_c \times n_f$ matrix T. For given surface fluxes f, by a model run, we are able to compute the resulting concentrations at the station locations $c_{mod} = Tf$. The matrix T itself is yet to be determined.

Following, e.g., *Enting et al.* [1995], by applying TM2 subsequently to the n_f standard basis vectors

$$e_1 = (1, 0, ..., 0), ..., e_{n_f} = (0, ..., 0, 1)$$

spanning $\mathbb{R}^{n_f}$, the matrix T could be computed column by column. This can be looked upon as a special case of approximating the Jacobian matrix that represents the first derivative of a function by differential quotients: Owing to linearity of the model (1) differential quotients are not merely an approximation of the Jacobian, and (2) the Jacobian of T is equal to T. A disadvantage of this approach is that it requires n_f runs of TM2, and thus is only feasible for a small number of flux components. In this section we introduce an alternative and for our matrix much more efficient approach: By the model adjoint to TM2 in the standard setup the Jacobian matrix is computed row by row in reverse mode. Here the computational cost depends on the number of rows, i.e., on n_c, rather than on the number of columns, i.e., on n_f. This kind of an adjoint model is uncommon in geosciences: Usually, rather than vector valued functions, scalar valued functions are being differentiated.

As will be sketched in section 3.1, for the implementation of an adjoint model there are alternative strategies. Following the concept of differentiation of algorithms, the adjoint of TM2 has been directly derived from the model code. For automatic generation of this adjoint code, the Tangent linear and Adjoint Model Compiler (TAMC, R. Giering, 1997, available at http://puddle.mit.edu/∼ralf/tamc) has been applied. Briefly summarizing earlier work [*Giering and Kaminski*, 1998], section 3.2 introduces the concept of differentiation of algorithms. Finally, section 3.3 describes how TM2's adjoint has been generated.

3.1. Adjoint Code Construction

In the following we briefly sketch three approaches to adjoint code construction whose essential difference is the level on which the adjoint operators are constructed.

1. Traditionally, as demonstrated, e.g., by *Marchuk* [1995] for various dynamical systems, adjoint models have been derived from the description of the system by a state function of space and time, being the solution of what Marchuk refers to as the main problem. Typically, the main problem consists of a set of differential equations together with initial and boundary conditions that, in the terminology of

functional analysis, define a (potentially non linear) differential operator T in an appropriate space of functions H. Spaces of this type are examples of Hilbert spaces, vector spaces furnished with an inner product $\langle \cdot , \cdot \rangle$. For the atmospheric transport of a passive tracer, the main problem consists of the continuity equation, together with a prescribed initial concentration field and a prescribed source sink distribution. Each observable quantity is represented by a linear functional on the Hilbert space. The control variables, i.e., functions that characterize the system such as initial or boundary conditions or parameters in the formulation of T, are also elements of appropriate Hilbert spaces. The sensitivity of a quantity to a change in the control variables is then the Hilbert space or continuous analogue of the familiar first derivative in finite dimensional spaces, which will be discussed in section 3.2. Applying first-order perturbation theory to the particular problem at hand, a Hilbert space analogue of the chain rule is derived: The sensitivity of the functional's value to a change in the control variables can be composed of the sensitivity of the functional's value to a change in the state function and the sensitivity of the state function to a change in the control variables. As can be shown, this sensitivity of the state function with respect to a change in the control variables can be obtained as the solution of the adjoint problem, being defined by the adjoint T^* of the differential operator T. The adjoint operator can be defined by

$$\langle T\phi , \psi \rangle = \langle \phi , T^*\psi \rangle \tag{9}$$

for each $\psi \in D(T^*) \subset H$ and $\phi \in D(T) \subset H$, whenever the domain $D(T)$ of T is 'large enough'. (If T is non linear, i.e., it depends on the state of the system, or it depends in a direct way on the control variables, an additional term quantifies this contribution to the sensitivity of the functional to the control variables. This is a continuous analogue to the product rule.)

In most practical applications the main problem is so complex that it has to be tackled numerically: First, a discretization scheme for the main equations is chosen, and then a numerical model for integration of the discrete equations is coded. Since, in general, the adjoint problem is as complex as the main problem, it is solved numerically as well. The resulting implementation is called adjoint model. The solution of the adjoint problem is then used to evaluate the discretized expression of the sensitivity.

2. Besides the cumbersome analysis that for a particular problem is necessary to rigorously define T and T^* and to derive an expression for the sensitivity, approach 1 has a distinct disadvantage: There is no unique choice of a discretization scheme for the adjoint problem, and a priori it is not clear which choice will result in a discrete version that is adjoint to the discretization of the main problem. In particular, the appropriate discretization scheme for the adjoint problem can be different from that for the main problem, i.e., as operators, building the adjoint and discretization do not interchange [*Griewank*, 1989]. Owing to inappropriate discretization, thus the sensitivity computed by the adjoint model differs from the sensitivity of the numerical model of the main problem. As is examined by, e.g., *Shah* [1991] and remarked by *Talagrand and Courtier* [1987], therefore it is favorable to develop the adjoint model from the discretization of the main problem: The adjoint operator is derived for the discretized form of T, operating in a finite dimensional space. Implicitly, this adjoint operator also defines the discretization scheme for the adjoint problem. As in the approach 1, eventually an adjoint model solving the discrete adjoint problem has to be implemented, and the solution is used to evaluate the discretized expression of the sensitivity. This approach has been applied to weather forecast models, e.g., by *Talagrand and Courtier* [1987], *Courtier and Talagrand* [1987] or to ocean circulation models, e.g., by *Thacker and Long* [1988].

3. A more direct approach for adjoint code generation uses the code of the main model as starting point: The composition of the main model with some functionals characterizing the quantities of interest is considered as an algorithm mapping a finite representation of the control variables onto the values of the functionals. As described below, by applying systematically the chain rule of differentiation to every single step in the model code in reverse mode, a model for the sensitivity is constructed. In the terminology introduced above, this model is the composition of the adjoint model with the implementation of the functional's first derivative. Using the model code as starting point for adjoint code construction, however, this distinction is no longer important, so that we slightly change our terminology and refer to this composition as adjoint model in the following. In section 3.2 we demonstrate that essentially, the adjoint model performs subsequent multiplications in reverse order of the adjoints of the Jacobians corresponding to the single steps in the model code. The main advantage of this approach is that, on the level of the single steps in the model code, the adjoints can be constructed according to simple rules [*Giering and Kaminski*, 1998]. Thus this task can be handled automatically [*Juedes*, 1991] (R. Giering, 1997, available at http://puddle.mit.edu/∼ralf/tamc) without any knowledge of the nature of the main problem and the system that is integrated by the model. For applications to geosciences, see, e.g., *Talagrand* [1991] and *Thacker* [1991]. The concept of applying systematically the chain rule to differentiate a numerical code is known as "differentiation of algorithms," "computational differentiation," or "automatic dif-

ferentiation" [*Griewank*, 1989], and adjoint code construction is merely one of its applications. For an overview, see, e.g., *Iri* [1991] or *Corliss and Rall* [1996].

3.2. Differentiation of Algorithms

In the following we describe how a function that is composed of elementary functions can be differentiated by use of the chain rule. When talking about elementary functions the reader should have in mind the single statements of the TM2 code, although the same mathematical formalism can be applied, if the elementary functions are considered to be related, e.g., to basic physical processes such as advection or diffusion. For automatic generation of derivative computing code, however, it is crucial that the Jacobians of the single steps can be constructed according to simple rules. Let

$$\begin{array}{rcl} \mathcal{H} : I\!R^n & \rightarrow & I\!R^m \\ X & \mapsto & Y \end{array}$$

be a function that is composed

$$\mathcal{H} \;=\; \mathcal{H}^K \circ \ldots \circ \mathcal{H}^1 \;=:\; \bigodot_{l=1}^{K} \mathcal{H}^l \tag{10}$$

of K differentiable elementary functions:

$$\begin{array}{rcll} \mathcal{H}^l : I\!R^{n_{l-1}} & \rightarrow & I\!R^{n_l} & \qquad (l = 1, ..., K) \\ Z^{l-1} & \mapsto & Z^l \;. & \end{array}$$

Even if $\mathcal{H}$ is not given symbolically, i.e., by a formula, but by a numerical algorithm such as TM2, the Jacobian matrix representing the first derivative of $\mathcal{H}$

$$\frac{\partial \mathcal{H}(X)}{\partial X} := \begin{pmatrix} \frac{\partial \mathcal{H}_1(X)}{\partial X_1} & \cdots & \frac{\partial \mathcal{H}_1(X)}{\partial X_n} \\ \vdots & & \vdots \\ \frac{\partial \mathcal{H}_m(X)}{\partial X_1} & \cdots & \frac{\partial \mathcal{H}_m(X)}{\partial X_n} \end{pmatrix}$$

can be computed using the chain rule of differentiation from the Jacobians of the elementary functions

$$\left.\frac{\partial \mathcal{H}(X)}{\partial X}\right|_{X=X_0} = \left.\frac{\partial \mathcal{H}^K}{\partial Z^{K-1}}\right|_{Z^{K-1}=Z_0^{K-1}} \cdot \ldots \cdot \left.\frac{\partial \mathcal{H}^1}{\partial Z^0}\right|_{Z^0=X_0} \;. \tag{11}$$

We have used

$$Z_0^l \;:=\; \mathcal{H}^l \circ \ldots \circ \mathcal{H}^1(X_0) \qquad (1 \leq l \leq K)$$

to denote the intermediate results, through which the derivatives of the elementary functions depend on X_0.

For evaluating the multiple matrix product in (11) there are many possibilities. Depending on the size of the elementary matrices they differ in the number of operations that have to be performed and in the size of the matrices containing the intermediate derivatives. For an algorithm tackling the evaluation of this multiple matrix product, the most obvious strategies are the forward and the reverse mode, where forward and reverse refer to the order of operations imposed by the composition: Operating in forward mode, the product is evaluated from the right to the left, which means that the product is computed in the same order as for evaluation of $\mathcal{H}$ in (10). Alternatively, the product can be evaluated from the left to the right, which is denoted as reverse mode, because the order is opposite to the order for evaluation of $\mathcal{H}$ in (10). In this evaluation procedure, the intermediate matrices at the lth step contain

$$\left.\frac{\partial(\mathcal{H}^l \circ \ldots \circ \mathcal{H}^1)(X)}{\partial X}\right|_{X=X_0}$$

in forward mode and

$$\left.\frac{\partial(\mathcal{H}^K \circ \ldots \circ \mathcal{H}^{l+1})(Z^l)}{\partial Z^l}\right|_{Z^l=Z_0^l}$$

in reverse mode. Thus forward and reverse refer to the directions in which the intermediate derivatives are propagated by the respective algorithm for evaluation of (11). According to (11), the forward mode step corresponding to the lth step of the composition (10) is

$$\left.\frac{\partial(\mathcal{H}^l \circ \ldots \circ \mathcal{H}^1)(X)}{\partial X}\right|_{X=X_0} = \left.\frac{\partial \mathcal{H}^l}{\partial Z^{l-1}}\right|_{Z^{l-1}=Z_0^{l-1}} \cdot \left.\frac{\partial(\mathcal{H}^{l-1} \circ \ldots \circ \mathcal{H}^1)(X)}{\partial X}\right|_{X=X_0} \;. \tag{12}$$

With respect to the standard inner product the adjoint matrix of

$$\frac{\partial \mathcal{H}(X)}{\partial X}$$

is simply the transposed matrix. Thus (11) can be written in the form

$$\left.\frac{\partial \mathcal{H}(X)}{\partial X}\right|_{X=X_0}^{*} = \left.\frac{\partial \mathcal{H}^1}{\partial Z^0}\right|_{Z^0=X_0}^{*} \cdot \ldots \cdot \left.\frac{\partial \mathcal{H}^K}{\partial Z^{K-1}}\right|_{Z^{K-1}=Z_0^{K-1}}^{*} \;. \tag{13}$$

This means, the reverse mode step corresponding to the lth step of the composition (10) is performed by multiplying the intermediate matrix

$$\left.\frac{\partial(\mathcal{H}^K \circ \ldots \circ \mathcal{H}^{l+1})(Z^l)}{\partial Z^l}\right|_{Z^l=Z_0^l}$$

Forward mode

$$
\begin{aligned}
& \begin{bmatrix} x & x & x \end{bmatrix} \begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \begin{bmatrix} x & x \\ x & x \\ x & x \end{bmatrix} \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \end{bmatrix} \\
= \; & \begin{bmatrix} x & x & x \end{bmatrix} \begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{bmatrix} \\
= \; & \begin{bmatrix} x & x & x \end{bmatrix} \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{bmatrix} \\
= \; & \begin{bmatrix} x & x & x & x & x \end{bmatrix}
\end{aligned}
$$

Reverse mode

$$
\begin{aligned}
& \begin{bmatrix} x & x & x \end{bmatrix} \begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \begin{bmatrix} x & x \\ x & x \\ x & x \end{bmatrix} \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \end{bmatrix} \\
= \; & \begin{bmatrix} x & x & x \end{bmatrix} \begin{bmatrix} x & x \\ x & x \\ x & x \end{bmatrix} \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \end{bmatrix} \\
= \; & \begin{bmatrix} x & x \end{bmatrix} \begin{bmatrix} x & x & x & x & x \\ x & x & x & x & x \end{bmatrix} \\
= \; & \begin{bmatrix} x & x & x & x & x \end{bmatrix}
\end{aligned}
$$

Figure 1. Example of forward and reverse modes illustrating the differences in the storage requirements and in the number of operations: The same matrix product, whose result has one row and five columns, is evaluated in forward mode, i.e., from right to left (top), and in reverse mode, i.e., from left to right (bottom). In forward mode the matrices holding the intermediate results have five columns, while in reverse mode they have one row.

by the adjoint of

$$\left.\frac{\partial \mathcal{H}^l}{\partial Z^{l-1}}\right|_{Z^{l-1}=Z_0^{l-1}},$$

which yields

$$\left.\frac{\partial(\mathcal{H}^K \circ \ldots \circ \mathcal{H}^l)(Z^{l-1})}{\partial Z^{l-1}}\right|_{Z^{l-1}=Z_0^{l-1}}^{*} = \left.\frac{\partial \mathcal{H}^l}{\partial Z^{l-1}}\right|_{Z^{l-1}=Z_0^{l-1}}^{*} \cdot \left.\frac{\partial(\mathcal{H}^K \circ \ldots \circ \mathcal{H}^{l+1})(Z^l)}{\partial Z^l}\right|_{Z^l=Z_0^l}^{*}. \tag{14}$$

Therefore the reverse mode is also called adjoint mode.

As illustrated by Figure 1 for a scalar valued function ($m = 1$) of $n = 5$ variables, in the forward mode all matrices containing intermediate derivatives have n columns, whereas in the reverse mode they have m rows. Therefore, in forward mode the number of operations as well as the storage requirements are proportional to n, whereas in reverse mode both is proportional to m.

In general, the intermediate results Z_0^l of the preceding step are required for the evaluation of the derivatives of the elementary functions (see (11)). While in the forward mode the intermediate results are required in the same order as computed, in the reverse mode they are required in reverse order. Thus providing of the intermediate results is more complicated in reverse mode and in general causes extra operations or extra storage requirements [*Giering and Kaminski*, 1998], which has to be taken into account when comparing the efficiency of reverse and forward mode for a particular function $\mathcal{H}$ (see section 3.3).

The Tangent linear and Adjoint Model Compiler (R. Giering, 1997, available at http://puddle.mit.edu/∼ralf/tamc) is a tool that automatically generates code for evaluation of first derivatives. The TAMC is a source to source translator that accepts essentially FORTRAN 77 code for the evaluation of a function and generates code for evaluation of its Jacobian. As requested by the user, the generated code operates either in forward or reverse mode. The schemes for forward or reverse mode are practically implementations of the general rules (12) and (14), respectively. Of course, this implementation is not unique: The scheme chosen for the TAMC is based on a few principles [*Giering and Kaminski*, 1998], which essentially have been suggested by *Talagrand* [1991]. Rigorous application of these principles yields rules for differentiating the single statements a code is composed of. These simple rules can be applied automatically by source to source translators like TAMC or Odyssée [*Rostaing et al.*, 1993].

3.3. Generation of the Adjoint Model

By the TAMC the model adjoint to TM2 in the standard setup has been generated automatically. To ensure an accurate interpretation by the TAMC the structure of the model code had to be slightly rearranged.

As is obvious from (14), the intermediate results Z_0^l (required variables) have to be provided for the adjoint run. Unlike many other adjoint applications in meteorology and oceanography, in transport models many of the required variables quantify the dynamic state of the atmosphere. These required variables do not depend on the control variables, i.e., the sources and sinks. In the terminology of adjoint code construction they are called passive variables. Hence, in principle, they could be computed and stored once and then be read during each adjoint run. Since this would require disk space of about 1.3 gigawords (GW), (at least on a Cray C90) it is more efficient to recompute the required values during every adjoint run. In order to reduce

these storage requirements during the adjoint run it is favorable to include a so-called checkpointing scheme [*Griewank*, 1992] in the adjoint model: In a first integration of TM2 the state of the model is saved at checkpoints in weekly intervals on disk. During the adjoint run the checkpoints are used as starting points for recomputation and storing of required values for the whole week in a second file. Finally, for the adjoint computations these stored values are read. The storage requirements are reduced considerably at the cost of an additional model integration. This checkpointing scheme also is implemented automatically by the TAMC.

In Table 1 the adjoint model's CPU and memory requirements are compared to computation of the Jacobian by differential quotients. The numbers refer to a Cray C90 supercomputer. For the standard setup with $n_c = 1$, the adjoint model needs the CPU time of about 3.5 TM2 runs and about the same amount of memory as TM2. The Jacobian for 27 stations, including the stations in Figure 2, has been computed in two separate runs in order not to allocate more memory than 32 megawords (MW). In total, the CPU time of about 85 TM2 runs has been used. While the memory requirements are proportional to the number of output values n_c, the CPU time per value decreases with increasing n_c for two reasons: First, for our function T, the cost of providing the required variables is independent of n_c. Thus, for higher n_c, there is no additional cost. Second, by the TAMC the adjoint code is arranged to achieve a vector lengths of n_c; for vectorized loops of the transport model, advanced compilers are even capable to enlarge vector dimensions by a factor of n_c. On a vector machine like the C90, this yields a considerable speedup, because the computations for the individual vector components are independent of each other. For the same reason, a similar speedup could be achieved on a parallel machine. In contrast, from the difference of runs with one and two tracers, one can estimate a CPU time of 7460 TM2 runs for the computation of the full Jacobian by an n_f tracer run. By rearranging the TM2 code, so that the tracer dimension n_f is used for vectorization instead of the dimension of the zonal grid (36), a speedup could be achieved, too. Yet this speedup is limited by the maximum vector length, which is 128 on the C90. In addition, this multitracer run would need more memory than is available on most machines (429 MW), so that it had to be split up to a couple of runs with less tracers. For a linear function like T, the Jacobian that is computed by differential quotients is free from truncation error. In that respect, the forward mode is not superior to differential quotients. Nor is the forward mode superior in terms of computational efficiency, because it includes an additional function evaluation, so that for small n_f the forward mode would be slightly slower, and for large n_f the efficiency would be comparable to differential quotients. Hence there is no need to include explicit numbers for the forward mode in this comparison. The last row in Table 1 explores the feasibility of the computation of a Jacobian for an observational network of 90 stations, which is about the extend of the Globalview CO_2 monitoring network [*Globalview–CO2*, 1996]. By scaling up the CPU time for an 18 station run (row 6) row 7 quantifies the requirements for computation of the Jacobian in reverse mode in 5 separate runs. Even for this extended network the cost of about 275 forward runs or 14.3 hours is small compared to the cost of about 20 days for the forward approach.

4. The Matrix Representation

In section 2 we have defined a standard setup of our transport model to simulate the quasi-stationary seasonal cycle at particular observational sites. Section 3 then has introduced the adjoint of the transport model and has discussed the computational benefit of applying the adjoint to derive a representation of the model by its Jacobian matrix T, which in the work of *Kaminski et al.* [this issue] is used for an inversion of the atmospheric transport of CO_2. Besides its use for inversions, the Jacobian by itself is an interesting object to study, because it quantifies how the transport relates a given flux field to the quasi-stationary seasonal cycle at the observational sites. In this section, we first visualize and discuss parts of the full Jacobian and then give examples of collapsing the matrix to compress or summarize its information. Also we demonstrate how the matrix is applied to analyze transport model runs by decomposing the simulated values with respect to the contributions of the fluxes into all grid cells in all months.

4.1. Visualization of Atmospheric Transport: Potential Impact

In the following we discuss the Jacobian matrix T derived for $n_s = 25$ locations of stations from the NOAA/CMDL global observational network (see Figure 2 and Table 2), whose data we use for our inversion example of *Kaminski et al.* [this issue]. A row of T consists of the sensitivity of the modeled concentration at a particular station and month to the fluxes into each of the $n_g = 36 \times 24$ TM2 surface layer grid cells at each month. The columns of T quantify the impact of a particular flux component on the modeled concentration at each station and month. The sensitivity or the impact are defined as the change in the concentration resulting from a change in the flux, which formally is represented by the derivative of the concentration with respect to the flux and has the unit of a concentration divided by a flux.

For comparison of the respective entries, direct visualization of the Jacobian is not very instructive: According to the

Table 1. Comparison of Efficiency in the Computation of the Jacobian Between Adjoint Model and Differential Quotients for a Cray C90

Run		CPU Time,			Memory,	
		s	h/d	Relative	MW	Relative
1	forward 1 tracer	186		1	0.933	1
2	forward 2 tracers	320		1.72	0.974	1.04
	10368 tracers (from 1 and 2)	1389364	16 d	7460	429.090	460
	10368 × 1 tracer	1928448	22 d	10368	0.933	1
3	adjoint, n_c =1	660		3.5	1.092	1.2
4	adjoint, n_c =24 (2 stations)	3045		16.4	3.999	4.3
5	adjoint, n_c =108 (9 stations)	5560		30	15.797	16.9
6	adjoint, n_c =216 (18 stations)	10260		55	30.962	33.2
	sum of 5 and 6	15820	4.4 h	85		
7	5 times 6 (90 stations)	51300	14.3 h	275	30.962	33.2

Columns: number and description of run, CPU time in seconds and multiples of the CPU time for a simple forward run, memory requirements in MW and in multiples of the memory required by a simple forward run. The numbers for 10368 tracers are computed from scaling up the differences between the 1 and 2 tracer runs (the forward model does not vectorize over the tracer dimension). The given numbers for a 90 station Jacobian correspond to a computation by five runs with 18 stations per run.

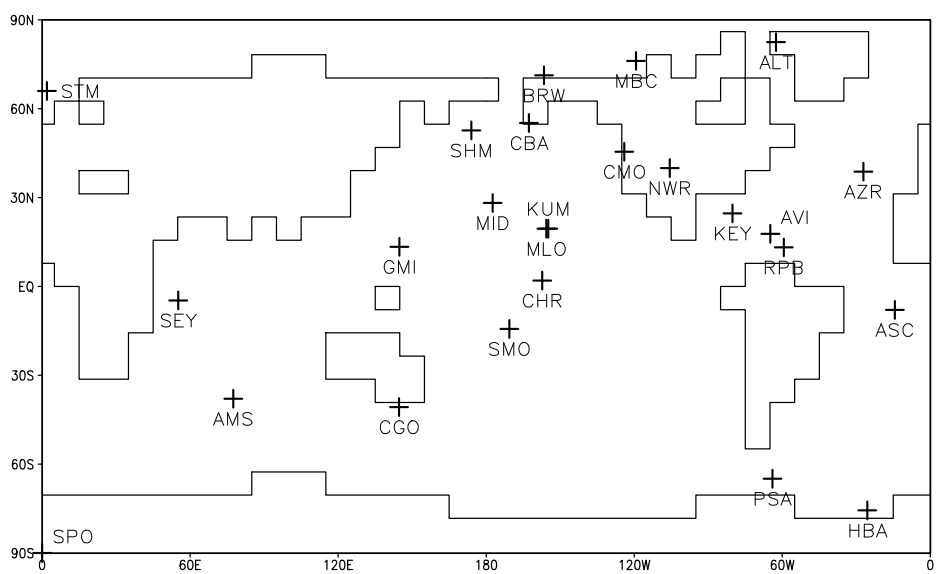


Figure 2. Twenty-five NOAA/CMDL monitoring stations whose observational data we use in our inversion example.

definition of our standard setup, the single entries quantify the concentration change that results from switching on a uniform flux for a particular month in a particular grid cell in every year of the four year simulation period. Hence, in addition to the properties of the atmospheric transport model, the matrix also reflects features determined by our setup, such as (1) the lengths of the spin up period, (2) whether the month the concentration refers to is earlier in the year than

Table 2. Twenty-Five NOAA/CMDL Monitoring Stations Whose Observational Data We Use in Our Inversion Example

Identifier	Description	Country	Latitude	Longitude	Elevation
ALT	Alert, N.W.T.	Canada	82 27'N	62 31'W	210
MBC	Mould Bay, N.W.T.	Canada	76 14'N	119 20'W	15
BRW	Point Barrow, Alaska	United States	71 19'N	156 36'W	11
STM	Ocean Station "M"	Norway	66 00'N	2 00'E	6
CBA	Cold Bay, Alaska	United States	55 12'N	162 43'W	25
SHM	Shemya Island	United States	52 43'N	174 06'E	40
CMO	Cape Meares, Oregon	United States	45 29'N	124 00'W	30
AZR	Azores (Terceira Island)	Portugal	38 45'N	27 05'W	30
NWR	Niwot Ridge, Colorado	United States	40 03'N	105 38'W	3749
MID	Sand Island, Midway	United States	28 13'N	177 22'W	4
KEY	Key Biscayne, Florida	United States	24 40'N	80 12'W	3
MLO	Mauna Loa, Hawaii	United States	19 32'N	155 35'W	3397
KUM	Cape Kumukahi, Hawaii	United States	19 31'N	154 49'W	3
GMI	Guam	U.S. Territory	13 26'N	144 47'E	2
AVI	St. Croix, Virgin Islands	United States	17 45'N	64 45 W	3
RPB	Ragged Point	Barbados	13 10'N	59 26'W	3
CHR	Christmas Island	Kiribati	2 00'N	157 19'W	3
SEY	Seychelles (Mahe Island)	Seychelles	4 40'S	55 10'E	3
ASC	Ascension Island	United Kingdom	7 55'S	14 25'W	54
SMO	American Samoa	U.S. Territory	14 15'S	170 34'W	30
AMS	Amsterdam Island	France	37 57'S	77 32'E	150
CGO	Cape Grim, Tasmania	Australia	40 41'S	144 41'E	94
PSA	Palmer Station (Anvers Island)	Antarctica	64 55'S	64 00'W	10
HBA	Halley Bay	Antarctica	75 40'S	25 30'W	10
SPO	Amundsen Scott (South Pole)	Antarctica	89 59'S	24 48'W	2810

the month the flux refers to, and (3) the lengths of the month the flux refers to. Feature (3) can be easily removed from the Jacobian by changing units from concentration per flux to concentration per yearly emission rate, i.e., ppmv GtC^{-1} a. To get rid of features (1) and (2), rather than the Jacobian itself, we plot its difference from an appropriate reference matrix. In (7), we already made use of such a reference matrix, namely the matrix whose entries quantify the changes in the global linear trend contributions to the respective concentration components that result from changes of the respective flux components. With this reference matrix, we get rid of feature (1) but not of feature (2), because the entries in the matrix are the same regardless of the month the flux belongs to. Yet this choice of a reference matrix is appropriate to visualize a column of the Jacobian, because within one column of T all entries refer to the same flux component, and its impact on all the concentration components can be compared. With respect to this reference matrix, plots of the columns, according to (7), show the impact of a particular flux component on the periodic contributions to each of the concentration components.

For visualization of the Jacobian's rows as in Figures 3 – 5 discussed below, in contrast, we choose a reference matrix that removes features (1) and (2), namely the Jacobian that our standard setup would yield, if global mixing was instantaneous. In other words, the reference matrix is derived from a one box model that behaves like TM2 with infinitely fast diffusion, i.e., it also uses $\alpha = 0.476$ ppmv GtC^{-1} to convert mass into concentrations. Since a row corresponds to the concentration at a particular station and month, it yields 12 global maps, each of which is quantifying this concentration's sensitivity to the mean surface exchange fluxes in a particular month at any location on the globe. A positive value on the map for any month quantifies a sensitivity to an emission at the corresponding grid cell and the respective months that is enhanced compared to instantaneous global mixing: a value of x ppmv GtC^{-1} y^{-1} means that a yearly emission of 1 GtC, which is uniformly distributed over the respective grid cell and month, in a TM2 run yields a monthly mean concentration at the station and month that is enhanced by x ppmv. Note that for stations in the lower model layers, the average of these sensitivities with respect to all flux components, in general, will be higher than zero. This is simply because we deal with surface fluxes, while our reference is derived for a homogeneous distribution in the entire atmosphere. In contrast, for observations in the stratosphere this average would be lower than zero.

As an example, in Figure 3 the second half of the matrix row corresponding to the November mean concentration at the station on Ascension Island (ASC: 7^o55'S, 14^o25'W, 54 m) is displayed. November emissions in the ocean region ranging from the south of Africa (30° south) to the equator at the longitude of ASC would have the highest impact (more than 10 ppmv GtC^{-1}). Going one month back to October emissions, the area of highest impact is shifting to the east, now covering the southern half of Africa. Still the impact of this region is at least as high as for November emissions. Interestingly, at the latitude of ASC in the Pacific Ocean and part of the Indian Ocean, the impact of emissions in November or even in October is smaller than for instantaneous global mixing. This demonstrates the disadvantages of using the mean concentration at a monitoring station in a two-dimensional inversion to constrain the fluxes at a latitude band around the respective station on a monthly timescale. In the maps quantifying the impact of emissions earlier in the year (not shown), the predominant structure is a division between both hemispheres. Compared to instantaneous global mixing the impact of the northern hemisphere is about 0.5 ppmv GtC^{-1} smaller, whereas the impact of the southern hemisphere is larger by the same amount. This feature is clearly caused by the relatively slow interhemispheric mixing across the Hadley cell. Quantitatively, the fact that the impact of October emissions north of 30° is more than 0.5 ppmv GtC^{-1} smaller as compared to instantaneous global mixing shows that not even the emissions of the previous year have been transported to ASC at an amount comparable to instantaneous global mixing (0.476 ppmv GtC^{-1}). This reflects the fact that in TM2 the transport needs more than one year to achieve a globally well mixed atmosphere (see section 2).

For comparison, maps for two stations and months are displayed, where the shape of the areas with high potential impact compared to instantaneous global mixing is more zonal than for ASC. Figure 4 shows the potential impact of emissions in the first half of the year to the May mean concentration at the station on the mountain Mauna Loa, Hawaii (MLO: 19^o32'N, 155^o35'W, 3397 m). The potential impact is highest for May emissions around of the station. The absolute peak values are lower than those for ASC (less than 10 ppmv GtC^{-1}) because the emission is diluted before reaching the mountain location. As another example, in Figure 5 we display the impact of emissions in the first half of the year on the June mean concentration at the Point Barrow station in Alaska (BRW: 66^o00'N, 2^o00'E, 6 m). Here the area of highest impact is well focussed near the station with high peak values of up to 70 ppmv GtC^{-1}. The information on potential impact can be compressed on the flux side, or on the concentration side, or both: Prescribing the shape of the seasonal cycle of the emissions into every surface grid cell, each matrix row can be projected to a single map of the potential impact of a yearly flux on the respective monthly mean concentration. On the concentration side, for all fea-

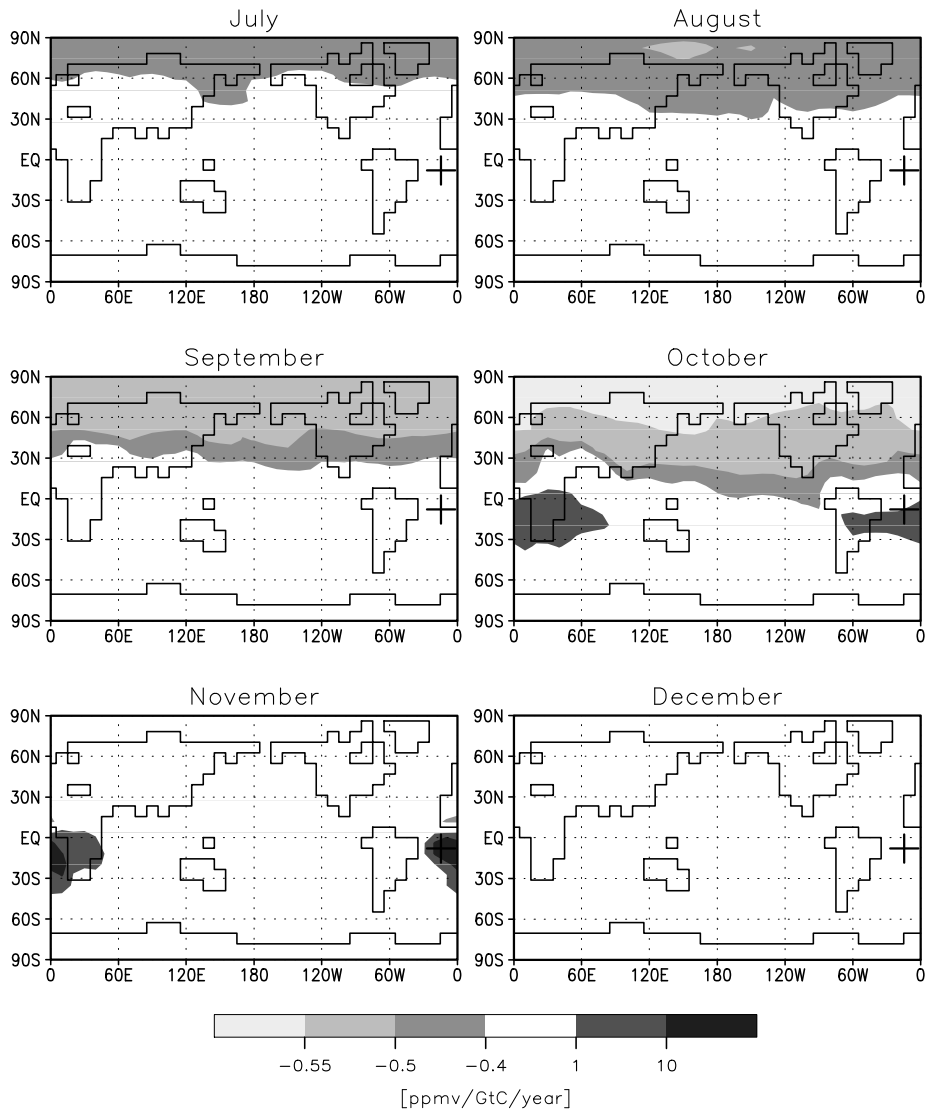


Figure 3. The second half of the Jacobian's row corresponding to the November mean concentration at the station on Ascension Island (ASC: 7°55'S, 14°25'W, 54 m). For our cyclostationary model setup, each global map shows the concentration's sensitivity to a periodical yearly emission, which is uniformly distributed over a particular month. Reference is instantaneous global mixing, i.e., negative numbers quantify sensitivities that are reduced owing to transport. The cross indicates the station location.

tures that can be derived from the monthly mean concentrations at the stations, the sensitivities with respect to monthly or yearly emissions (in combination with prescribed temporal shape) can be easily computed from the matrix. As an example, in Figure 6 we show the sensitivity of the annual mean concentration at ASC, MLO, and BRW, respectively, to fluxes that are constant in time over the whole year. Compared to the monthly maps the peak of the potential impact is lower, slightly more widespread but still in the same regions. This indicates that for uniform emissions throughout the year, at these stations the modeled concentration is not very sensitive to the seasonality of the transport.

Another way of looking at the maps is in terms of the size of surface areas that are "observed" by the respective stations: On the monthly timescale all three stations are most influenced by an area of only a few grid cells. On the annual

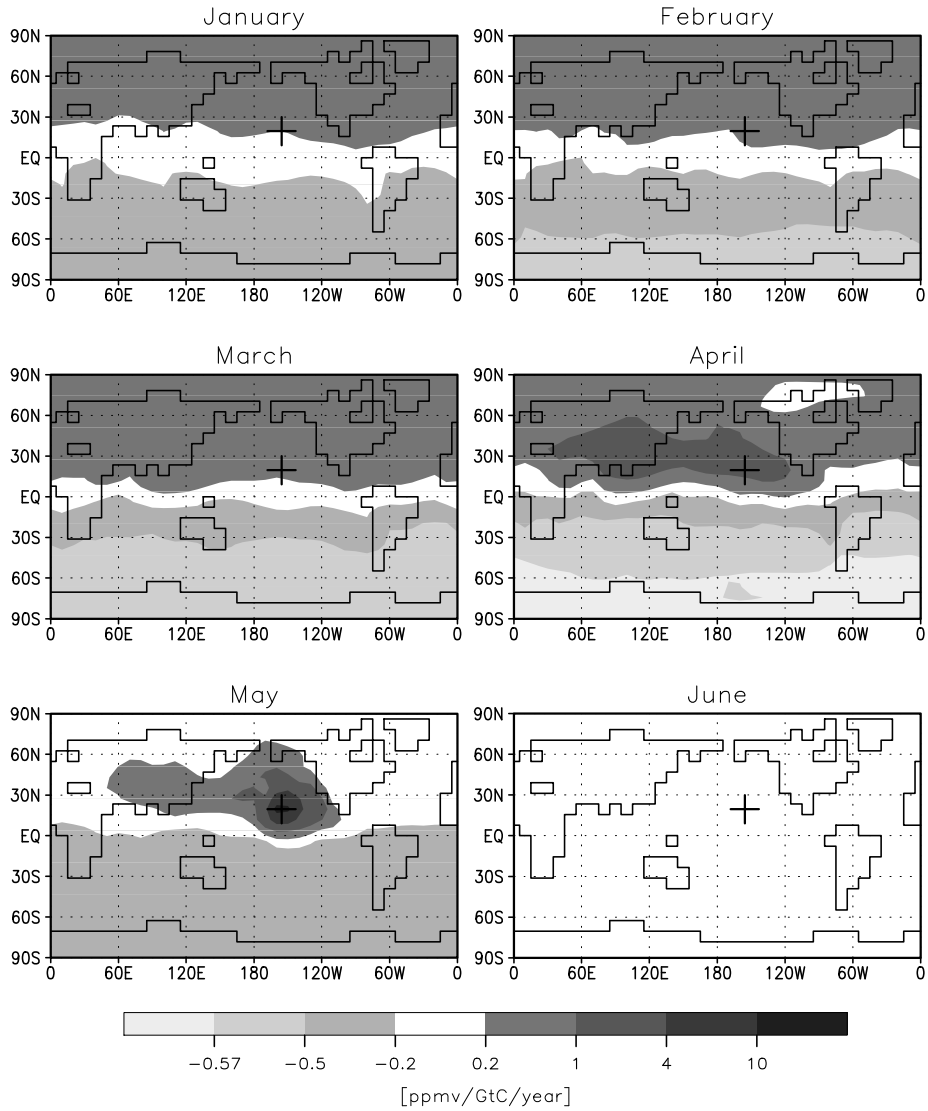


Figure 4. The first half of the Jacobian's row corresponding to the May mean concentration at the station on the mountain Mauna Loa, Hawaii (MLO: 19°32'N, 155°35'W, 3397 m). For our cyclostationary model setup, each global map shows the concentration's sensitivity to a periodical yearly emission, which is uniformly distributed over a particular month. Reference is instantaneous global mixing, i.e., negative numbers quantify sensitivities that are reduced owing to transport. The cross indicates the station location.

timescale there are differences among the stations: While ASC still observes only a small area, BRW is representative for the northern high latitudes, and MLO is strongly influenced by the entire northern hemisphere. When investigating a particular scientific question these transport characteristics, of course, are merely a fraction of the features that determine the importance of a monitoring location. Other features are the specific source/sink characteristics of the tracer of interest.

4.2. Combining Atmospheric Transport and Flux Fields: Simulated Impact

We discussed the potential impact quantified by the Jacobian. If a particular flux field f is prescribed, according to (8) by a matrix multiplication with the Jacobian this potential impact can be used to simulate the resulting quasi-stationary

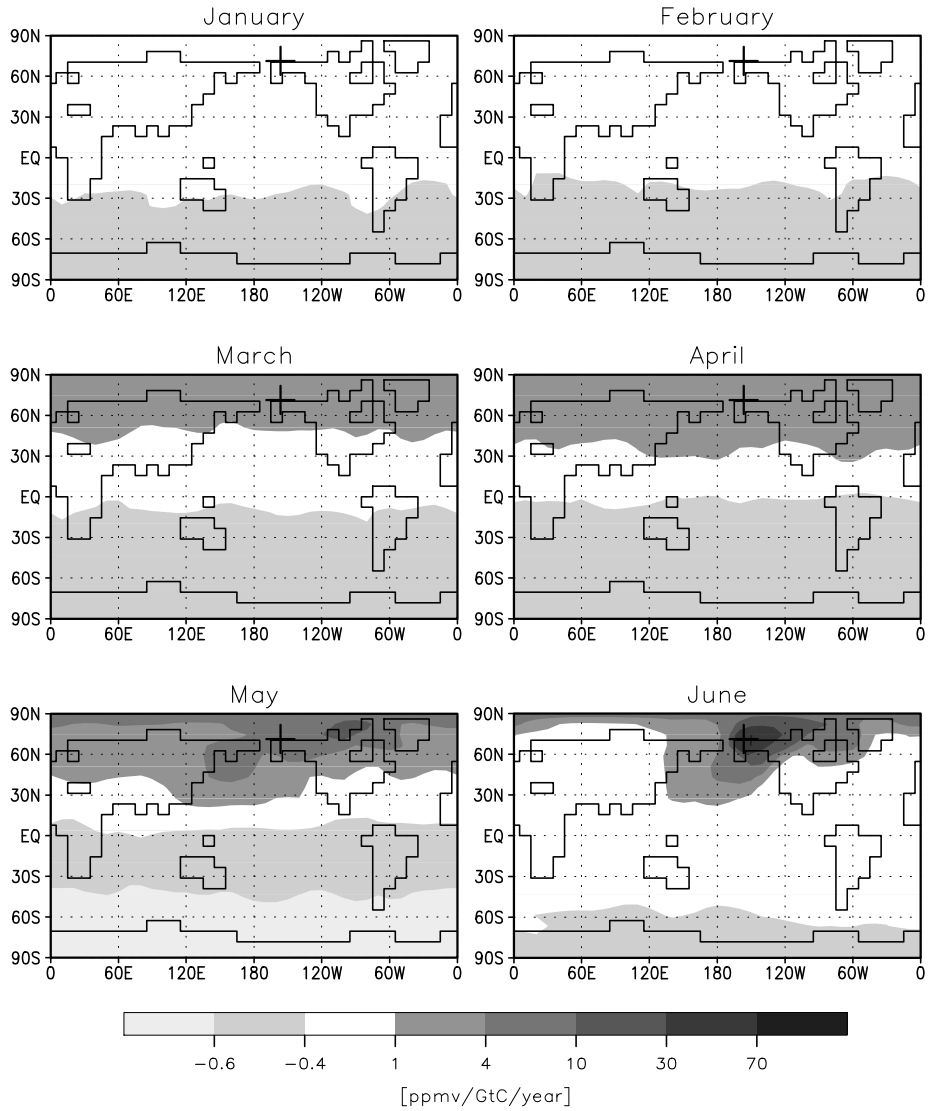


Figure 5. The first half of the Jacobian's row corresponding to the June mean concentration at the Point Barrow station in Alaska (BRW: 66°00'N, 2°00'E, 6 m). For our cyclostationary model setup, each global map shows the concentration's sensitivity to a periodical yearly emission, which is uniformly distributed over a particular month. Reference is instantaneous global mixing, i.e., negative numbers quantify sensitivities that are reduced owing to transport. The cross indicates the station location.

seasonal cycle at the station locations. Hence the Jacobian is an extremely efficient transport model by itself. Once the Jacobian has been computed, for the simulation of the quasi-stationary seasonal cycle at the stations, there is no need to run TM2 again, as long as the setup (including the location of the stations) is still appropriate for the tracer of interest.

For an example, we employed the a posteriori CO_2 fluxes inferred in an inversion of the atmospheric transport [*Kaminski et al.*, this issue]. These fluxes are the sum of the fossil fuel component and the biospheric and oceanic components depicted in Figure (9) of *Kaminski et al.* [this issue]. Figure 7 shows the simulated periodic component of the quasi-stationary seasonal cycle at Mauna Loa, which has been computed according to (8).

Using the matrix does not only reduce the computational cost of a simulation to the cost of a simple matrix multipli-

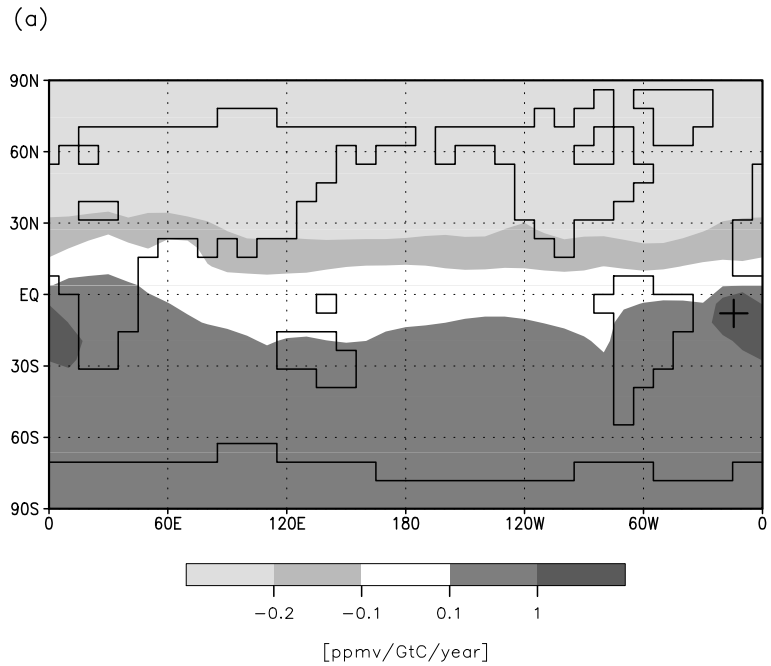
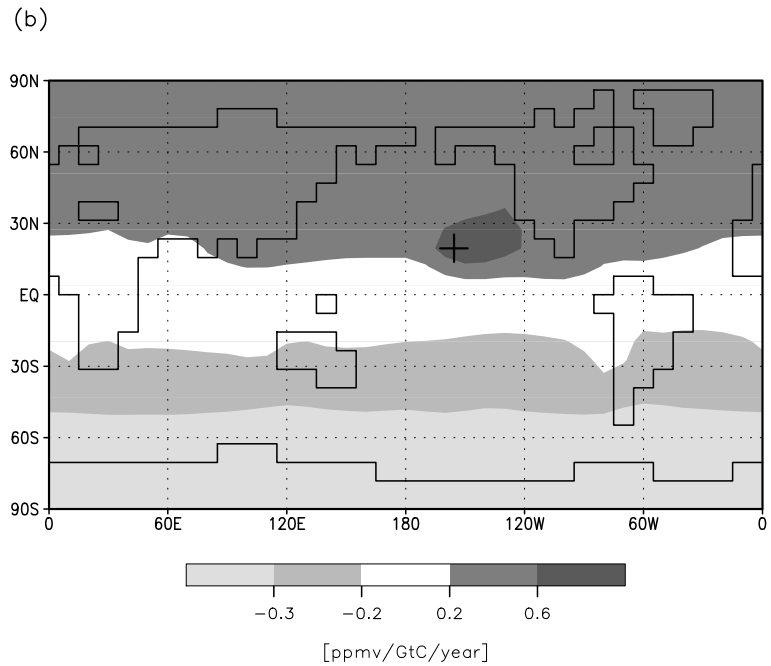


Figure 6. Collapsed Jacobian's rows corresponding to the 12 monthly mean concentrations at the stations (a) on Ascension Island (ASC: 7°55'S, 14°25'W, 54 m), (b) on the mountain Mauna Loa, Hawaii (MLO: 19°32'N, 155°35'W, 3397 m) and (c) at Point Barrow in Alaska (BRW: 66°00'N, 2°00'E, 6 m). The annual mean concentration's sensitivity to a periodical yearly emission, which is constant in time, in our cyclostationary model setup. Reference is instantaneous global mixing, i.e., negative numbers quantify sensitivities that are reduced owing to transport. The cross indicates the station location.



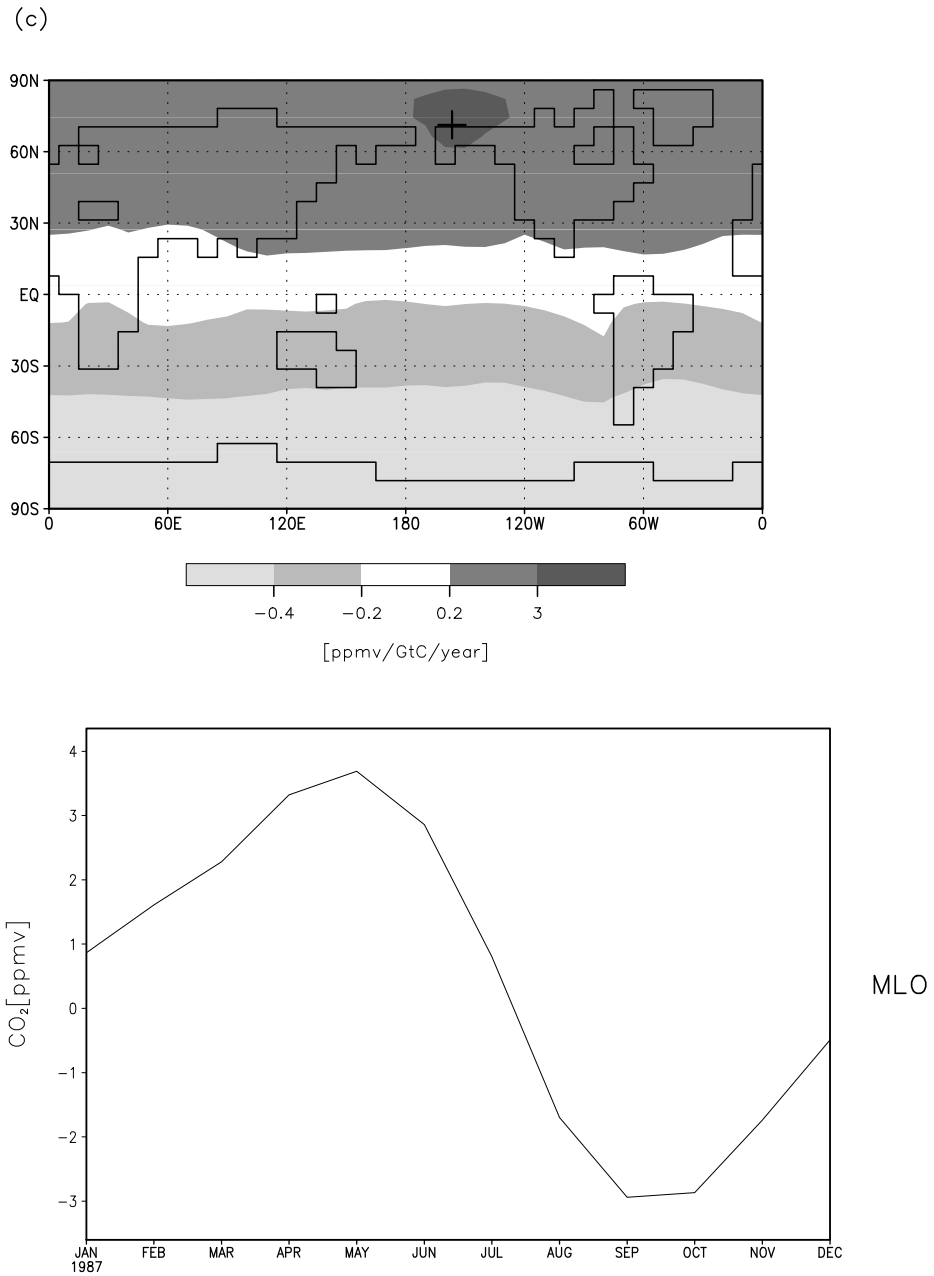


Figure 7. Simulated quasi-stationary seasonal cycle at Mauna Loa based on the flux field inferred in the inversion of *Kaminski et al.* [this issue].

cation but also the amount of required disk space. While the meteorological fields to drive TM2 for one year occupy about 30 MW, the matrix just needs $36 \times 24 \times 12 \times 27 \times 12$ W ≈ 3 MW. Thus, among other applications, as transport model the Jacobian represents a valuable tool for sensitivity tests: *Knorr* [1997] investigated the response of the atmo-

spheric CO_2 concentration at the NOAA/CMDL stations to exchange flux fields computed by a large number of different formulations of his terrestrial biosphere model.

In addition to quantifying potential impact and to perform transport simulations, by means of the Jacobian it is easy to analyze the simulation in terms of the simulated impact of all components of a prescribed flux vector: Writing (8) in the form (and dropping the index *qsc* for convenience)

$$c_i = \sum_{j=1}^{n_f} M_{i,j} f_j \quad , \tag{15}$$

each concentration component c_i is decomposed into the contributions $c_{i,j} := M_{i,j} f_j$ by the respective flux component f_j. The quantity $c_{i,j}/c_i$ is then the portion of c_i resulting from the flux component j in the simulation and like the potential impact can be conveniently displayed on 12 maps per concentration component.

As an example, we analyzed the simulation of the quasi-stationary seasonal cycle at Mauna Loa, which was based on the flux field described above. Figure 8 shows the decomposition of the May mean in the periodic component of the quasi-stationary seasonal cycle, which is depicted in Figure 7. Recall that according to the definition of this periodic component in (15), e.g., for an emission, the impact on the trend (reference term) is subtracted from the impact quantified by the product of the emission with the respective column of the Jacobian (Jacobian term). Discussion of the maps is complicated by the different signs of both terms, i.e., their relative magnitude is important. On the northern hemisphere where the station is located, in general, the reference term is smaller than the transport term, so that the interpretation is rather straight forward: In months where fluxes into the atmosphere are positive, grid cells tend to have a positive contribution to the May concentration at Mauna Loa (i.e., the May component in the quasi-stationary seasonal cycleat Mauna Loa). In winter, this is the case for most of the terrestrial grid cells, i.e., most of Asia, Europe and North America. In contrast, whenever there are large fluxes from the atmosphere into the ocean or the biosphere, the respective grid cells have a negative contribution, i.e., those fluxes reduce the May concentration at Mauna Loa. This is the case for the North Atlantic sink. Of course, according to (15), this is weighted by the effect of the transport: For example, although the absolute value of the May fluxes into the North Atlantic is smaller than that of the terrestrial uptake in June at the temperate latitudes over Asia, the contribution of the North Atlantic sink in May to the May concentration at Mauna Loa is much larger. These different weighting factors are reflected in Figure 4. In contrast, for fluxes on the southern (and remote) hemisphere, the reference term in (15) tends to dominate the Jacobian term, which is small due to slow interhemispheric exchange. For this reason sources in South America and the southern part of Africa have a negative impact on the May concentration at Mauna Loa: Such a source flattens the north-south gradient, and the reduction of the annual mean part of the periodic component at Mauna Loa (due to subtraction of the trend) overcompensates the increase due to the Jacobian term. On the other hand, for example in December, the fluxes from the atmosphere into the Southern Ocean have a positive contribution to the May concentration at Mauna Loa: The north south gradient is increased, so that the annual mean in the periodic component at MLO increases, too. And this increase overcompensates the decrease of the May mean concentration due to the Jacobian term.

Again, as for the potential impact, the information can be compressed on the flux side, the concentration side, or both sides. For example, in the work of *Kaminski et al.* [1996] we analyzed a TM2 run using the fluxes derived by a biosphere model [SDBM, *Knorr and Heimann*, 1995]: On the flux side we prescribed the shape of the SDBM fluxes, and on the concentration side we projected on the simulated seasonal cycle. We thus decomposed the magnitude of the modeled seasonal cycle at particular observational sites with respect to the contributions by the respective grid cells, which yields one map per station. For this study we had to run the adjoint model once per station. By means of the Jacobian this kind of decomposition is easily performed without the adjoint model.

5. Concluding Remarks

We demonstrated the benefit of the adjoint approach for the computation of the Jacobian matrix representing a three dimensional atmospheric transport model. This matrix maps flux fields on the model's approximately 8° by 10° horizontal grid onto the simulated concentrations at 27 observational sites. For this setup the computational efficiency of the adjoint was about 100 times higher as compared to conventional forward modeling.

The adjoint model has been automatically generated from the transport model code by the TAMC. To ensure an accurate interpretation, prior to invoking the TAMC, the code had to be prepared and rearranged slightly. In particular constructs that complicate the order of execution of the statements had to be replaced. Unlike the conventional use of adjoint models, where the adjoint model evaluates the derivative of a scalar valued cost function, which is then iteratively minimized by an optimization algorithm, the Jacobian computed here is the derivative of a linear vector valued function.

As a linear function mapping fluxes on concentrations at observational sites, the Jacobian contains all information

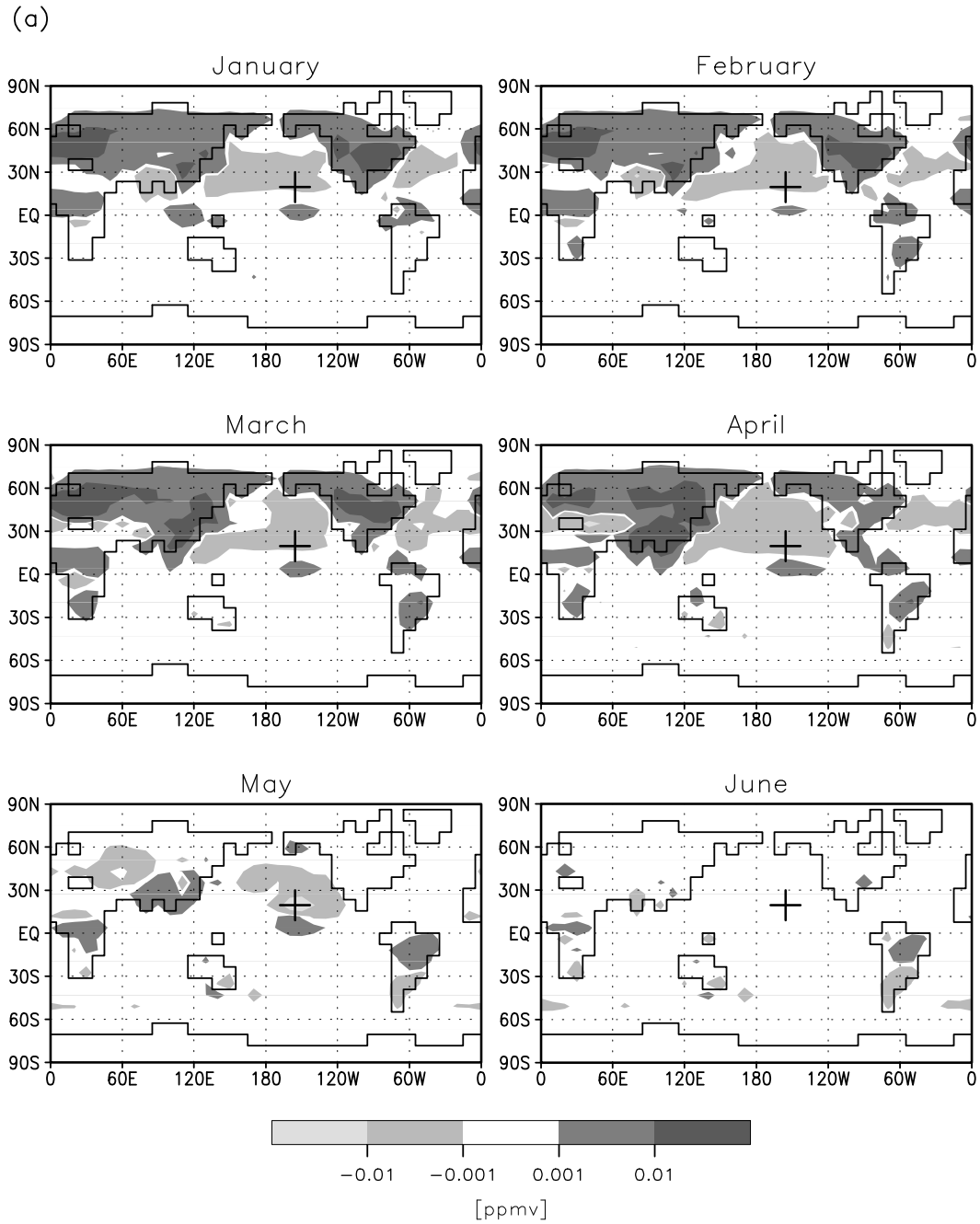
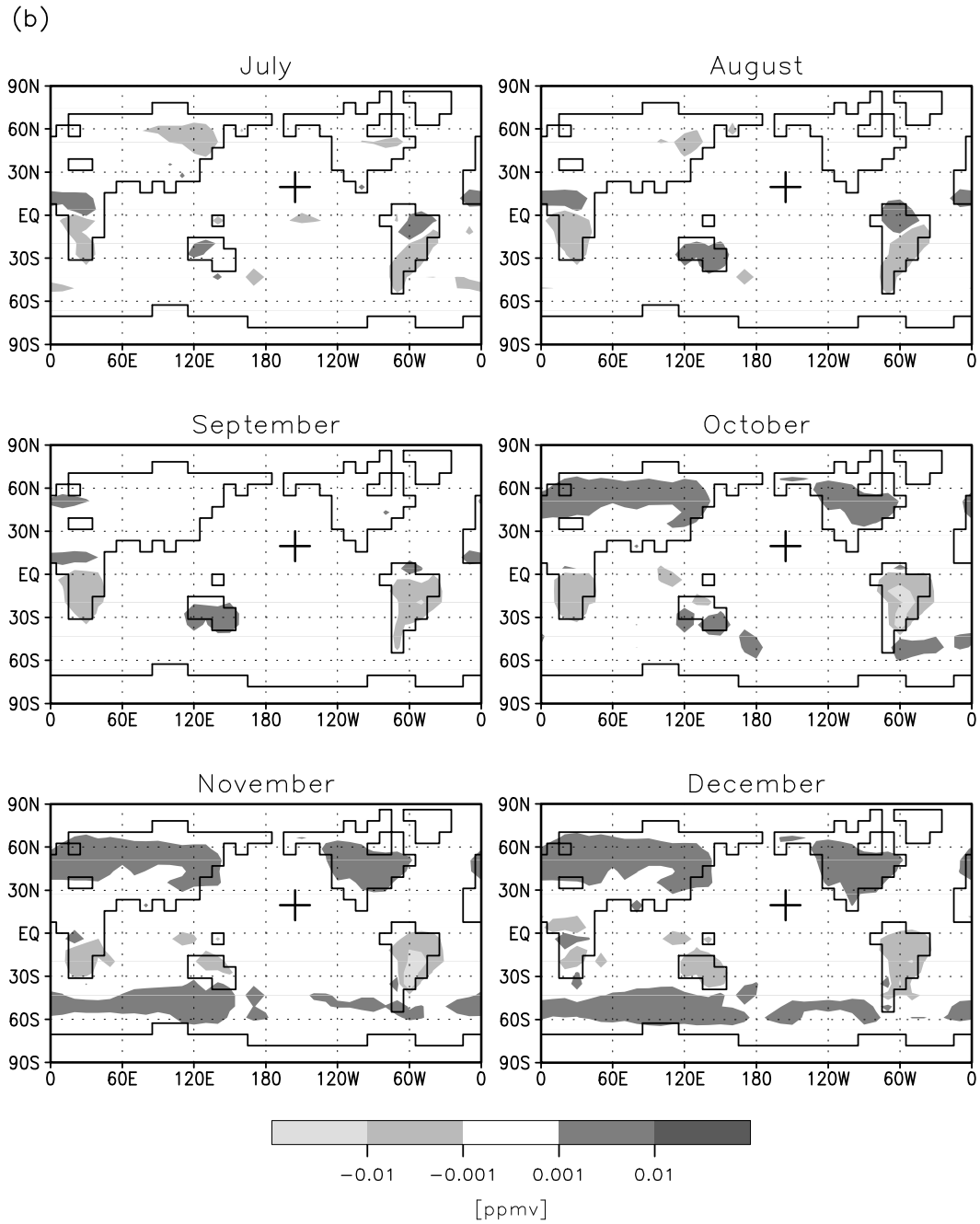


Figure 8. Decomposition of the May component in the quasi-stationary seasonal cycle at Mauna Loa based on the flux field inferred in the inversion of *Kaminski et al.* [this issue]. The respective maps quantify the contributions from the fluxes at all months and grid cells in per cent. (a) fluxes from January to June; (b) fluxes from July to December. Negative values mean that the fluxes in the respective months and grid cells have a negative contribution, i.e., increasing those fluxes would yield a reduced May component in the quasi-stationary seasonal cycle at Mauna Loa.



about the transport. Hence, once the Jacobian is available for a particular setup, it can replace the transport model: To simulate the concentrations at the station locations, instead of running the model for a given flux field, this flux field can be multiplied by the Jacobian, which is much more efficient in terms of both memory and CPU requirements.

Plots of the rows of the Jacobian provide information

about the potential impact of emissions at every location on the globe and in every month on the modeled concentration at a particular station and month. On the other hand, combining the Jacobian to a prescribed flux field, a simulated concentration value at a particular station and month can be analyzed: This value can be decomposed into the contributions of the fluxes in the respective grid cells and months.

Such maps of potential or simulated impact could provide valuable information about differences in the transport simulated by different models. In that respect the reverse approach could complement the maps of concentration fields simulated by running prescribed flux fields forward through different models. The reverse approach requires that adjoints of the respective transport models be available. Since transport models typically are implemented in Fortran, we suggest the use of automatic differentiation tools such as the TAMC.

The Jacobian contains all necessary transport information to infer the magnitude of cyclostationary CO_2 surface exchange fluxes together with their uncertainties from observed concentrations at the station locations and prior estimates of the fluxes. In a companion paper [*Kaminski et al.*, this issue], we present such an inversion study using atmospheric CO_2 observations of the period from January 1981 to January 1987 from the NOAA/CMDL program [*Conway et al.*, 1994; *Globalview–CO2*, 1996]. Our inversion contrasts the conventional use of adjoint models for optimization, where a (potentially expensive) computation of second derivatives is necessary to obtain estimates of the uncertainties in the unknown variables.

Acknowledgments. The authors thank two anonymous reviewers for helpful suggestions and comments as well as Michael Voßbeck for producing all GrADS plots. This work was supported in part by the Commission of the European Communities under contract EV5V-CT92-0120. Computing support was provided by the Deutsches Klimarechenzentrum (DKRZ) in Hamburg.

References

Bousquet, P., Optimisation des flux nets de CO_2: assimilation des mesures atmosphériques en CO_2 et en $\delta^{13}C$ dans un modele de transport tridimensionnel, Ph.D. thesis, Univ. Paris VI, 1997.

Brown, M., Deduction of emissions of source gases using an objective inversion algorithm and a chemical tranport model, *J. Geophys. Res.*, *98*, 12,639–12,660, 1993.

Brown, M., The singular value decomposition method applied to the deduction of the emissions and the isotopic composition of atmospheric methane, *J. Geophys. Res.*, *100*, 11,425–11,446, 1995.

Ciais, P., et al., Partitioning of ocean and land uptake of CO_2 as inferred by $\delta^{13}C$ measurements from the NOAA climate monitoring and diagnostics laboratory global air sampling network, *J. Geophys. Res.*, *100*, 5051–5070, 1995.

Conway, T. J., P. Tans, L. Waterman, K. Thoning, D. Buanerkitzis, K. Masarie, and N. Zhang, Evidence for interannual variability of the carbon cycle from the noaa–cmdl global air sampling network, *J. Geophys. Res.*, *99*, 831–855, 1994.

Corliss, G., and L. B. Rall, An introduction to automatic differentiation, in *Computational Differentiation: Techniques, Applications, and Tools*, edited by M. Berz, C. Bischof, G. Corliss, and A. Griewank, pp. 1–18, SIAM, Philadelphia, Pa., 1996.

Courtier, P., and O. Talagrand, Variational assimilation of meteorological observations with the adjoint equation, II, Numerical results, *Q. J. R. Meteorol. Soc.*, *113*, 1329–1347, 1987.

Denning, A. S., I. Y. Fung, and D. Randall, Latitudinal gradient of CO_2 due to seasonal exchange with biota, *Nature*, *376*, 240–243, 1995.

Enting, I. G., Green's function methods of tracer inversion, in *Geophysical Monograph Series*, edited by P. Kasibhatla, M. Heimann, D. Hartley, P. J. Rayner, N. Mahowald, and R. Prinn, AGU, Washington, D. C., in press, 1999.

Enting, I. G., and J. V. Mansbridge, Seasonal sources and sinks of atmospheric CO_2: Direct inversion of filtered data, *Tellus, Ser. B*, *41*, 111–126, 1989.

Enting, I. G., C. M. Trudinger, and R. J. Francey, A synthesis inversion of the concentration and $\delta^{13}C$ of atmospheric CO_2, *Tellus, Ser. B*, *47*, 35–52, 1995.

Giering, R., and T. Kaminski, Recipes for Adjoint Code Construction, *ACM Trans. Math. Software*, *24*, 437–474, 1998.

Globalview–CO2, *Cooperative Atmospheric Data Integration Project - Carbon Dioxide* [CD-ROM], NOAA/CMDL, Boulder, Colo., 1996.

Griewank, A., On automatic differentiation, in *Mathematical Programming: Recent Developments and Applications*, edited by M. Iri and K. Tanabe, pp. 83–108, Kluwer Acad., Boston, Mass., 1989.

Griewank, A., Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation, *Optimization Methods and Software*, *1*, 35–54, 1992.

Haas-Laursen, D., Regional estimates of carbon dioxide fluxes deduced with an inverse method, Ph.D. thesis, Georgia Inst. of Technol., Atlanta, 1997.

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

Heimann, M., The global atmospheric tracer model TM2, *Tech. Rep. 10*, Max-Planck-Institut f. Meteorol., Hamburg, Germany, 1995.

Heimann, M., and C. D. Keeling, A three-dimensional model of atmospheric CO_2 transport based on observed winds, 2, Model description and simulated tracer experiments, in *Aspects of Climate Variability in the Pacific and the Western Americas, Geophys. Monogr. Ser.*, edited by D. H. Peterson, vol. 55, pp. 237–275, AGU, Washington, D. C., 1989.

Heimann, M., C. D. Keeling, and I. Y. Fung, Simulating the atmospheric carbon dioxide distribution with a three-dimensional tracer model, in *The Changing Carbon Cycle; A Global Analysis*, edited by J. Trabalka and D. Reichle, pp. 16–49, Springer-Verlag, New York, 1986.

Hein, R., and M. Heimann, Determination of global scale emissions of atmospheric methane using an inverse modelling method, in *Non-CO_2 Greenhouse Gases*, edited by J. van Ham et al., pp. 271–281, Kluwer, Norwell, Mass., 1994.

Hein, R., P. Crutzen, and M. Heimann, An inverse modeling approach to investigate the global atmospheric methane cycle, *Global Biogeochem. Cycles*, *11*, 43–76, 1996.

Houghton, J. T., L. M. Filho, B. Callander, N. Harris, A. Dattenberg, and K. Maskell (Eds.), *Climate Change 1994 - Radiative Forcing of Climate Change*, Cambridge Univ., New York, 1995.

Iri, M., History of automatic differentiation and error estimation, in *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, edited by A. Griewank and G. F. Corliss, pp. 3–16, SIAM, Philadelphia, Pa., 1991.

Jacob, D. J., M. J. Prather, S. C. Wofsy, and M. B. McElroy, Atmospheric distribution of ^{85}Kr simulated with a general circulation model, *J. Geophys. Res.*, *92*, 6614–6626, 1987.

Juedes, D., A taxonomy of automatic differentiation tools, in *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, edited by A. Griewank and G. F. Corliss, pp. 315–329, SIAM, Philadelphia, Pa., 1991.

Kaminski, T., R. Giering, and M. Heimann, Sensitivity of the seasonal cycle of CO_2 at remote monitoring stations with respect to seasonal surface exchange fluxes determined with the adjoint of an atmospheric transport model, *Phys. Chem. Earth*, *21*, 457–462, 1996.

Kaminski, T., M. Heimann, and R. Giering, A coarse grid three-dimensional global inverse model of the atmospheric transport, 2, Inversion of the transport of CO_2 in the 1980s, *J. Geophys. Res.*, this issue.

Keeling, C. D., R. B. Bacastow, A. F. Carter, S. C. Piper, T. P. Whorf, M. Heimann, W. G. Mook, and H. Roeloffzen, A three-dimensional model of atmospheric CO_2 transport based on observed winds, 1, Analysis of observational data, in *Aspects of Climate Variability in the Pacific and the Western Americas, Geophys. Monogr. Ser.*, edited by D. H. Peterson, vol. 55, pp. 165–236, AGU, Washington, D. C., 1989.

Knorr, W., Satellitengestützte Fernerkundung und Modellierung des Globalen CO_2 -Austauschs der Landvegetation: Eine Synthese, Ph.D. thesis, Max-Planck-Institut f. Meteorol., Hamburg, Germany, 1997.

Knorr, W., and M. Heimann, Impact of drought stress and other factors on seasonal land biosphere CO_2 exchange studied through an atmospheric tracer transport model, *Tellus, Ser. B*, *47*, 471–489, 1995.

Law, R., and I. Simmonds, The sensitivity of deduced CO_2 sources and sinks to variations in transport and imposed surface concentrations, *Tellus, Ser. B*, *48*, 613–625, 1996.

Law, R. M., CO_2 sources from a mass balance inversion: sensitivity to the surface constraint, *Tellus, Ser. B*, *51*, 254–265, 1999.

Law, R. M., et al., Variations in modelled atmospheric transport of carbon dioxide and the consequences for CO_2 inversions, *Global Biogeochem. Cycles*, *10*, 783–796, 1996.

Louis, J. F., A parametric model of vertical eddy fluxes in the atmosphere, *Boundary Layer Meteorol.*, *17*, 187–202, 1979.

Marchuk, G. I., *Adjoint Equations and Analysis of Complex Systems*, Kluwer Acad., Boston, Mass., 1995.

Pearman, G., and P. Hyson, Activities of the global biosphere as reflected in atmospheric CO_2 records, *J. Geophys. Res.*, *85*, 4468–4474, 1980.

Rayner, P. J., and R. M. Law, A comparison of modelled responses to prescribed CO_2 sources, *Tech. Pap. 36*, CSIRO Div. of Atmos. Res., Aspendale, Victoria, Australia, 1995.

Rayner, P. J., I. G. Enting, R. J. Francey, and R. L. Langenfelds, Reconstructing the recent carbon cycle from atmospheric CO_2, δ^{13}C and O_2/N_2 observations, *Tellus, Ser. B*, *51*, 213–232, 1999.

Rostaing, N., S. Dalmas, and A. Galligo, Automatic differentiation in Odyssée, *Tellus, Ser. A*, *45*, 558–568, 1993.

Russel, G. L., and J. A. Lerner, A new finite-differencing scheme for the tracer transport equation, *J. Appl. Meteorol.*, *20*, 1483–1498, 1981.

Shah, P., Application of adjoint equations to estimation of parameters in distributed dynamic systems, in *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, edited by A. Griewank and G. F. Corliss, pp. 181–190, SIAM, Philadelphia, Pa., 1991.

Talagrand, O., The use of adjoint equations in numerical modelling of the atmospheric circulation, in *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, edited by A. Griewank and G. F. Corliss, pp. 169–180, SIAM, Philadelphia, Pa., 1991.

Talagrand, O., and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation, I, Theory, *Q. J. R. Meteorol. Soc.*, *113*, 1311–1328, 1987.

Thacker, W. C., Automatic differentiation from an oceanographer's perspective, in *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, edited by A. Griewank and G. F. Corliss, pp. 191–201, SIAM, Philadelphia, Pa., 1991.

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

Tiedtke, M., A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Mon. Weather Rev.*, *117*, 1779–1800, 1989.

Trampert, J., and R. Snieder, Model estimations biased by truncated expansions: Possible artifacts in seismic tomography, *Science*, *271*, 1257–1260, 1996.

Watson, R., M. Zinyowera, and R. Moss (Eds.), *Climate Change 1995 - Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses: Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press, New York, 1995.

R. Giering, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena CA 91109. (ralf@pacific.jpl.nasa.gov)

M. Heimann and T. Kaminski, Max-Planck-Institut für Meteorologie, Bundesstr. 55, D-20146 Hamburg, Germany. (heimann@dkrz.de; kaminski@dkrz.de)

September 2, 1998; revised February 26, 1999; accepted March 3, 1999.

This preprint was prepared with AGU's LaTeX macros v4, with the extension package 'AGU^{++}' by P. W. Daly, version 1.5d from 1997/04/28.