The authors thank the reviewers for their encouraging and helpful comments. In the following, the comments are printed in **bold face**, our replie in normal face, and the actions taken to improve the manuscript in *italic face*.

Report 1

Comment: **This paper is clearly the result of a major and impressive undertaking with substantial investment by the authors – chapeau! The paper is centered around the numerical treatment of the inverse problem of deriving transport characteristics from tracer measurements, which by itself is novel and has the potential to make a fundamental and important contribution so should be published. My main concerns are with the physical interpretation of the inferred transport characteristics and the approximations used to derive the tracer continuity equations. As outlined in my major comment below, I think the authors need to include more discussion of these potential issues. I also have a few minor and editorial comments that should be taken into account before publication.**

Reply: The authors thank the reviewer for this encouraging comment.

Action: *N/A.*

Comment: **Major Comments:**
The purpose of the approach is to apply it to zonal-mean atmospheric tracer data. The corresponding continuity and tracer continuity Eqs are supposedly those arising from the zonally averaged 3-d Eqs, but in fact they are not, according to the derivations in the appendix. On line 7, page 25 it is claimed that density-weighting is performed (as is in section 3.1, line 6 on page 5): this would require redefining the zonal average and, more importantly, redefining the eddy part of the Reynolds decomposition. Also in the appendix, on line 15 of page 25 it is stated that the velocity field is assumed to be non-divergent. This essentially corresponds to applying a Boussinesq approximation, which is what the referenced Ko et al. (1985) use (discussed in their appendix). But applying a Boussinesq approximation means that the (relevant) density perturbations are neglected and therefore no density-weighting is used. I am skeptical that a Boussinesq approximation is suitable for this problem, although its possible that this is less of an issue in the height coordinates used here (it most certainly is an issue for the isentropic coordinates used by Ko et al). I recommend consulting Tungs 1986 paper (J. Atmos. Sci., 43, pages 2600-2618) that lays out a zonal-mean framework for isentropic coordinates and includes a detailed description of the mathematical treatment of density-weighting (his section 2). Similar frameworks apply to other coordinate systems; an exception

are pressure coordinates (or log-pressure coordinates) because they implicitly include mass/density-weighting. I think these issues should be discussed in the appendix. I recommend focusing on the conceptual points (approximations needed to make eddy tensor independent of tracer; interpretation of eddy tensor and symmetric / antisymmetric components as additional mixing / mean advection); the detailed mathematical treatment is not necessary I think or at least can be significantly condensed.

Reply: We agree that this part has been the week point in the initial submission. Since for us the correctness of the paper is more important than rapid publication, we would appreciate if the reviewer would have another critical look at the revised version.
Our inverse model will produce "effective velocities" and "effective mixing coefficients" instead of zonally averaged velocities and mixing coefficients, because of the eddy terms. The point is that the interpretation of these "effective kinematic quantities" depends on this choice and on the assumed approximations. The interpretation tells us which effects beyond the true zonal mean velocities and the true subscale effects are included in the "effective kinematic quantities". We have tackled the problem as follows:
First, we have studied the literature about 2D models involving mass-weighted zonal averages (Gallimore and Johnson, 1981, JAS **38**, 583-599 and 1870-1890; Tung, 1986, JAS **43**, 2600-2618). We found that the theory of mass-weighted zonal averaging is applied to θ coordinates in order to distinguish different dynamical processes. We are currently interested in the kinematic understanding and do not yet aim at a dynamical understanding (i.e. we want to know what the effective kinematic quantities are, and what they represent, but not by which forces they are caused). Further, our model is formulated in geometrical vertical coordinates. Thus, we have not pursued this path any further. We now work again with unweighted zonal averaged VMRs.
Second, we have studied the Tung 1986 paper, which left us somewhat clueless but contains a reference to Tung (1982, JAS 39 2330-2355) which appears helpful. In particular his formalism does not use the Boussinesq approximation about which we agree that it is not adequate for our purpose. We do not use it any more. In Tung (1982), a theory of a 2D model (non-weighted zonal means) is outlined which seems well applicable also to our case. We have taken the following steps:

1. we have checked that the formalism is also applicable to geometric vertical coordinates and have verified that the same approximations are valid.
2. we list the approximations made in this paper.
3. we have developed an interpretation of our kinematic quantities, which is based on these approximations.
4. we explicitly state that the interpretation of the inferred kinematic quantities depends on the assumptions and approximations made.

Action: *The entire appendix has been rewritten, based on Tung (1982). The new appendix is much shorter because we do not report the entire formalism, starting* *from the 3D continuity equation, but we start from Eq. D14 in Tung (1982). We now present only those equations which are indispensable to understand our interpretation of the kinematic variables.*

Comment: **Related to the above, I think the main description in section 3.1 needs more physical interpretation. The meanings of "effective transport velocity" and "eddy diffusion" are not unique, but depend on the kinds of approximations and treatment of perturbations outlined in the appendix. A different zonal mean / eddy treatment and different approximations would result in different effective transport velocities and eddy diffusivities. In fact, I would argue that the only unique definition of "effective transport velocity" would be one that sets the eddy diffusivities to zero (ad hoc), thereby absorbing all resolved transport into zonal-mean advection. After all, "eddy diffusion" is a parametrization of eddy advection and almost all transport is in reality advective (although, of course, some small-scale / subgrid scale diffusion is still necessary to close the system). I am sure the authors have thought about the physical interpretation of their deduced transport contributions and I would like to see some of those thoughts discussed somewhere in the paper. The way its currently written makes it sound too much like the inferred transport characteristics are unique.**

Reply: We agree.

Action: The assumptions, approximations, and interpretation of the involved quantities are now better specified. A clear statement is made that the interpretation of our effective kinematic quantities is not unique but depends on the approximations and assumptions made. As an alternative, it is now also proposed to use our method to validate 2D models, including their inherent assumptions and approximations to solve the eddy transport problem.

Comment: **Minor Comments, incl. editorial:**
page 1, line 15: "are under debate" feels out-of-date: as suggested by the introduction and by the other reviewer comments, a lot of what seemed to be a debate at first has become a more detailed and nuanced description

Reply: agreed

Action:Changed as follows:
OLD: are under debate
NEW: have become an important research topic

Comment: **page 1, line 21: "is not that one-dimensional": somewhat confusing what this refers to (I suppose it refers to the spatial vari-**

ability in age trends - see following sentence in text)

Reply: agreed

Action:Changed as follows:
OLD: ... not as one-dimensional. Instead, stratospheric age trends ...
NEW: ... more complex. Stratospheric age trends ...

Comment: **page 1, line 24: I think you either need to spell out the "ad hoc assumptions" or remove this remark**

Reply: agreed

Action:added: These include the adequacy of the Wald (inverse Gaussian) function for the representation of the age spectrum and the choice of its width parameter.

Comment: **page 2, line 6/7: I think most models in practice use the Earth's surface as a reference (see e.g. CCMVal reports)**

Reply: But not all. E.g. In Stiller et al. (2012, ACP **12** *3311-3331) a reviewer tried to force us to use the stratospheric entry point as reference, claiming that this was the usual reference.*

Action:Reworded as follows:
OLD: ... the modelling community has established the upper edge of the tropical tropopause layer as a reference (Hall and Waugh, 1994), which makes a difference ...
NEW: ... parts of the community have established the stratospheric entry point as a reference (Hall and Plumb, 1994), which makes a difference ...

Comment: **page 3, line 11: suggest "source and sink terms" for clarity**

Reply: agreed

Action: "and sink" inserted.

Comment: **page 4, line 2: "shallowness approximation": it will help some readers to spell out the approximation (i.e. z much smaller than r_E)**

Reply: agreed

Action:inserted "[which] simplifies the equations using the assumption that z is much smaller than r_E and which [is...]

Comment: **page 7, line 28: differentiate**

Reply: agreed, thanks for spotting

Action:*corrected.*

Comment: **page 10, line 4: I'd prefer "identity matrix"**

Reply: agreed.

Action: *"matrix" inserted after "identity".*

Comment: **page 12, line 4: equal to an integer multiple ...**

Reply: agreed

Action:*"with" replaced by "to".*

Comment: **page 12, line 12: I suggest "real atmosphere"**

Reply: agreed.

Action:*"true" replaced by "real".*

Comment: **page 19, line 22/23: velocities should be v and w, subscript for K_ϕ (also Fig. caption of Fig. 3)**

Reply: Agreed. Thanks for spotting.

Action:*corrected.*

Comment: **page 20, line 13: "roughly consistent" should be specified/quantified more - in what way are they consistent (list a few examples, such as overall latitudinal and vertical structures etc)?**

Reply: Agreed. Specified as reported below.

Action:*We have inserted "...in a sense that the quotient of the typical circulation velocity and the distance between equator to pole gives an age estimate of the correct order of magnitude".*

Comment: **page 20, line 14/15: "shows many detailed features demanding scientific investigation in their own right": which detailed features and why to they demand investigation? I think this should be discussed in the text.**

Reply: We think that it is too early to scientifically discuss the results when the proof of concept is still an issue. However, in order to give a flavour of what can be investigated with this method, we now mention a couple of features visible in our example shown, which may be interesting research issues.

Action: *We have inserted: "(e.g., the latitude offset between the intertropical convergence zone and the stratospheric tropical pipe, or the interfacing between the stratospheric two-cell circulation and the overturning mesospheric circulation and the transition altitude between them)"*

Comment: **page 20, line 17: should be "mixing coefficients"**

Reply: agreed, thanks for spotting.

Action: *Corrected.*

Comment: **page 20, line 25: weekly**

Reply: agreed, thanks for spotting.

Action: *Corrected.*

Comment: **page 20, line 33: view _of_ the same problem**

Reply: agreed, thanks for spotting.

Action: *corrected.*

Comment: **Fig. 1: right-hand column misses a color bar; labels are hard to read - please increase font size**

Reply: agreed.

Action: *The figure has been replaced*

Comment: **Fig. 2 caption: "a detail of this" - I suggest "a zoomed-in view"**

Reply: agreed

Action: *reworded as suggested.*

Report 2

Comment: **General comment:**
The paper presents a method for deducing atmospheric circulation

(wind field) and mixing parameters from trace gas measurements by inversion of the continuity equation. In a first step, the mathematical framework is defined and explained. Second, the method is applied to idealized tracer fields and to MIPAS satellite measurements (in the "proof of concept" section 5), to show that the inversion indeed results in reliable velocities and diffusivities. Deducing information about the circulation from measurements, without involving information from models, is a great challenge in atmospheric sciences. This paper seems to contain an important contribution to reach that goal, what renders it definitely publishable and of great interest to a large readership of ACP. However, I have two major points which the authors need to assess before publication. First, the paper is not easy-to-read and the presentation quality needs improvement - otherwise I feel that the paper will fail in addressing a large readership.

Reply: see related specific replies below.

Action:*see below.*

Comment: **Second, I have some concerns about the so-called "proof of concept".**

Reply: A 'proof of concept', we think, is something much more modest than, say, a 'validation'. A proof of concept shall provide evidence that the chosen pathway is likely to lead to interesting results when pursued further. It is not meant as a full test of the functionality. The validation in terms of comparison with models etc. is a full project in its own right.

Action:*see below.*

Comment: **Major comments:**
1) Presentation and Notation:
Overall, the paper is overloaded with detailed formulae, but lacking motivating and explanatory paragraphs. In their own words (P2, L33), the authors aim to avoid "that the reader does not see the forest for the trees" but, in my opinion, there are still too many trees around. For instance in section 3, there should be a clear motivation at the beginning, why the derivatives (which are calculated in the following) are needed and what the matrix notation means.

Reply: Well, removing the trees also would mean removing part of the forest. Instead of removing equations, we prefer to better guide the reader through the equations by adding some explanatory text.

Action:
text before Eq. 27 modified to "With these expressions, the prediction of air

density and volume mixing ratio can be rewritten in matrix notation for a single micro time increment. This notation simplifies the estimation of the uncertainties of the predicted atmospheric state and the inversion of the prediction equation. In matrix notation, the prediction reads"
added before Eq. 77: "i.e. we linearly predict the new atmospheric state for a given initial state as a function of wind and mixing ratios"
added after Eq. 77: "This formulation gives access to the winds and mixing ratios via inversion of $\mathbf{F}$*."*

Comment: **After that the equations (15-26) could be nicely combined into one single equation-array (similarly in section 4, starting with equation (37)).**

Reply: With this we would lose all the explaining comments between the equations.

Action:*none*

Comment: **Concerning all formulae, writing X; χ; μ; ... for mixing ratio instead of vmr would help to increase readability.**

Reply: With all the other variable names we have detected ambiguities. X (however lower case) is used as generalized atmospheric state variable (from Eq. 31 on), and χ would be confusing because χ^2 is our cost function in Eq. 78. Admittedly, μ is not used in our paper but its use to designate mixing ratios is not that much established while it is often used to designate other quantities. Following Wikipedia, it designates a measure in measure theory, minimalization in computability theory and recursion theory, the mean in the normal distribution, the integrating factor in ordinary differential equations, the learning rate in artificial neural networks, the Möbius function in number theory, the population mean or expected value in probability and statistics, the coefficient of friction, the reduced mass in the two-body problem, linear density, or mass per unit length, in strings and other one-dimensional objects, permeability in electromagnetism, the magnetic dipole moment of a current-carrying coil, dynamic viscosity in fluid mechanics, the amplification factor or voltage gain of a triode vacuum tube, the electrical mobility of a charged particle, the muon, the chemical potential of a system, and many others but not mixing ratio. Thus we prefer to use the self-explaining variable name vmr_g

Action:*none*

Comment: **Moreover, while many steps in the calculation are written in detail (like taking derivatives), at some points I was not able to understand the derivations in detail. One such example is the matrix notation in equation (27). First, a clear motivation should be given why this matrix notation is advantageous and what it means (this is**

the heart of the paper).

Reply: well, as usual, matrix notation saves a lot of writing (both in the paper and in the computer code). To better motivate our matrix formalism, a paragraph on the logical flow of the operations has been inserted, where also particular advantages of the matrix notation are highlighted.

Action:*In the Section "General Concept", the logical flow of the operations is outlined. For this purpose, the following paragraph has been added: "Our concept involves the following operations. First, a general solution of the forward problem is formulated (Section 3). The forward problem is the solution of the prediction equation as a function of the initial atmospheric state for given winds and mixing coefficients. For our chosen solver, which involves the MacCormack (1969) integration scheme for the solution of the transport problem (Eqs. 5–10), the relevant dependencies of the final state on the initial state are reported in Section 3.2 (Eqs. 15-26). The formulation in matrix notation (Eqs 27–28) allows the easy handling of multiple successive timesteps (Eq. 29) and an easy estimation of the prediction error via generalized Gaussian error estimation (Eq. 30). As a next step, the dependence of the predicted state on winds and mixing coefficients is estimated for a given initial state . This is achieved by differentiation of the solution of the prediction equation with respect to winds and mixing coefficients (Eqs 34–76). These partial derivatives form the Jacobian matrix of the problem, with which the estimation of winds and mixing coefficients can be reduced to a constrained least square optimization problem where the inversely variance-weighted residual between the predicted atmospheric state and the respective measured atmospheric state is minimized. The latter step involves the generalized inverse of the Jacobian matrix (Eqs. 78–90)."*

Comment: **Second, I did not succeed in understanding the dimensionalities of the quantities involved. As the authors state, the D-matrix is build from three submatrices of dimensions $K_0 \times K_0$ (I_K), $K_0 \times 2K_0$ (W_i), and $J_0 \times L_0$ $(D_\rho; nom)$. Therefore, the D-matrix has dimensions $(2K_0 + J_0) \times (3K_0 + L_0)$, which is, as far as I can see, not consistent with the vector it is acting on. Please check the dimensionalities again...**

Reply: Contrary to what we have state in our preliminary reply, the reviewer is right. Sorry for the premature author comment!

Action:*Equations have been corrected. Corrected dimensionalities are now reported.*

Comment: **... and explain clearly what equation (27) means. Equation (35) caused me similar problems with understanding. Please explain clearly where it comes from.**

Reply: All entries and the rationale behind them are explained in the list after this equation (admittedly hard to understand with the dimensionality error in the discussion version, which has now been corrected, see above. Instead of adding more words, we have decided to visualize which matrix blocks operate on which elements of the atmospheric state vector.

Action:*A figure has been added which illustrates which parts of the D-matrix operate on which part of the input vector.*

Comment: **The appendix is, in my opinion, not necessary. It just presents a recalculation of the existing literature. I would recommend to reduce such recalculations, but to add explanations at the critical and new steps of this paper (e.g., around Eqns. 27/35). If the authors want to keep that part, it could be moved to the supplementary material.**

Reply: agreed.

Action:*As stated above, we have removed the old appendix because it was incorrect. In the new appendix we do not recalculate anything. We only report Eq. D14 of Tung (1982) which we have rewritten in our system and notation, list the definitions of involved terms, and present the interpretations of the 2D velocities and mixing coefficients inferred from this.*

Comment: **2) "Proof of concept": In my opinion, section 5.3 does not really present a "proof of concept", as promised by the title of section 5. The method is used to deduce velocities and diffusivities from tracer measurements, but the true underlying circulation is not known. Therefore, this case is no proof that the inversion method yields the correct result. I think, for a true "proof of concept" the circulation and diffusivities must be known before and need to be reproduced by the method. Section 5.2 points into this direction, but is exclusively descriptive. The optimal "proof of concept" would be to have a 2D-model based on equations (3-4) with idealized velocities and diffusivities, and to invert the resulting trace gas distributions. At least, the cases described in section 5.2 should be explained in more detail and related results should be shown in the paper.**

Reply: As said above, a 'proof of concept', we think, is something much more modest than, say, a 'validation'. A proof of concept shall provide evidence that the chosen pathway is likely to lead to interesting results when pursued further and that the scheme is in principle functional. It is not meant as a full test of the functionality. The validation in terms of comparison with models etc. is a full project in its own right and is planned as future activity. Reporting and discussing all of the functionality tests performed during the development of the program would fill a book (but a boring one!). Such tests are often so trivial that the reader does not learn a lot. To give the reader a flavor of what the tests were *like, we have decided to report one representative example.*

Action:We have selected a testcase as an example and discuss it.

*Comment:***Specific comments:**
P2, L1ff: Another source of uncertainty when deducing mean age from SF6 is related to the fact that the tropospheric increase is not strictly linear (see Garcia et al., 2011, JAS).

Reply: We agree in part. The non-linearity of tropospheric SF_6 *is not an issue in itself. E.g., Stiller et al (2012) or Haenel et al. (2015) use o nonlinear tropospheric reference curve to infer the age of air from the tracer measurements. Up to that point, the inference of the age of air is unambiguous as long as the tropospheric growth is monotonical. Nonlineary becomes only a problem by mixing processes. If tropospheric growth was strictly linear, all effects implied by the asymmetric age spectrum would cancel out and no convolution with the age spectrum would be needed to obtain the correct mean age of air. In cases of nonlinear tropospheric increase, the consideration of the age spectrum becomes relevant. But, e.g. Stiller et al. (2012) or Haenel et al (2015) consider this. These authors use an iterative scheme to infer the age. As first guess, they use the age directly inferred from the SF6 mixing ratio. Then, they calculate the age spectrum for this initially guessed age, convolve the (non-linear) tropospheric time series with this age-dependent age spectrum, infer a correction, and iterate until convergence. Thus, the only remaining uncertainty in this context is the uncertainty of the age spectrum. However, we agree that not every reader might be aware of this method, and we will add a note on the non-linearity issue.*

Action:We have changed the text as follows:
OLD: ...by an age spectrum, on which some ad hoc assumptions have to be made...
NEW: ...by an age spectrum, which has to be considered since the tropospheric growth of SF_6 *mixing ratios is not strictly linear, and on which some ad hoc assumptions have to be made....*

Comment: **P2, L7: To my knowledge, in models usually the surface layer is used as a reference, not the upper edge of the TTL. Please clarify.**

Reply: Some people insist that the upper edge of the TTL is the only adequate reference. We admit that these people are not necessarily modellers.

Action:We have changed the text as follows
OLD: ... the modelling community has established the upper edge of the tropical tropopause layer as a reference (Hall and Waugh, 1994), which makes a difference ...
NEW: ... parts of the community have established the stratospheric entry point

as a reference (Hall and Plumb, 1994), which makes a difference ...

Comment: **P18, L11: How robust are the deduced velocities and diffusivities with respect to the choice of initial value for the iteration. Please give some quantitative estimate.**

Reply: In a constrained retrieval, where the initial guess is set equal to the a priori field, there are generally two possible mechanisms for a dependence of the result on the initial guess.

1. *The effect of the constraint: The solution of a constrained inversion always has a tendency to be pulled towards the a priori field, or, in our case with 1st order Tikhonov constraint, the field gradients are constrained to those of the a priori field. Our a priori constraint is chosen as zero velocity throughout. Since the resulting field gradients deviate largely from this initial guess and resemble those of the expected velocity fields, it is evident that the inversion is able to find a solution which is far away from the initial guess and is not overly constrained by the a priori assumption.*

2. *Non-linearity and possible secondary minima of the cost function: The observed convergence rate indicates an almost linear inverse problem; thus, no such related problems are seen. We have experienced that problematic cases regularily end up in non-converging iterations rather than converging to different results and are thus easily sorted out.*

Our results provide the smoothest field of velocities and the smallest mixing coefficients which are still consistent with the measurements. Any change of the a priori assumptions must have an impact on the results. How large this intentional dependence is, is fully user-controllable by adjustment of the regularisation strength.

Action:none.

Comment: **P20, L7: How can the residual be small for SF_6 if no chemical sink is included in the calculation? Is the sink effect absorbed in the transport terms, or is a significant sink only existing above the upper boundary for the calculation?**

Reply: The problem with the sink in Stiller et al. (2012) is, that age-calculations are sensitive to the accumulated decomposition since the air has left the troposphere. In our case, only the decomposition during the finite time-step of the calculation can contribute,because the atmospheric state at the beginning of the time step already is depleted in the trace gases. For SF6, the major sink is indeed above the altitude range under consideration.

Action:none.

Comment: **P20, L13: "Velocities are roughly consistent with mean ages...". Some misinterpretations in the past arose from relating mean age simply to the stratospheric circulation. However, mean age is known to be controlled by circulation and mixing (e.g., Neu and Plumb, 1999; Garny et al., 2014; Ploeger et al., 2015). So please discuss carefully what you mean here with "consistent".**

Reply: We emphasize the attribute "roughly". This statement is not meant as a quantitative assessment but is meant to say that there is no obvious major contradiction between our results and our current knowledge on stratospheric mean age of air. We will change the text towards a more careful wording.

Action:We have inserted "...in a sense that the quotient of the typical circulation velocity and the distance between equator to pole gives an age estimate of the correct order of magnitude".

Comment: **Technical corrections:**

Equation (11): I think there should be no minus here (in the supplement there is also no minus: Eq. (34)).

Reply: Thanks for spotting!

Action:*corrected.*

Comment: **Equation (13): Brackets missing around the argument of volume mixing ratio.**

Reply: Thanks for spotting!

Action:*corrected.*

Comment: **Equation (18): Missing point behind equation.**

Reply: Thanks for spotting!

Action:*corrected.*

Comment: **Equation (27): Are the dimensions correct - see my major comment 1.**

Reply: No! Contrary to what we have stated in the preliminary reply, dimensions are indeed incorrect.

Action:*corrected.*

Comment: **P12, L15: Point behind "numerical artefacts."**

Reply: agreed.

Action:*corrected.*

Comment: **P13, L6: K's should have an index j, like v and w.**

Reply: yes, indeed.

Action:*corrected.*

Comment: **P13, L15: vmr in italics - or better: use some symbol instead (e.g., X, see also my major comment 1).**

Reply: agreed for the italics.

Action:*corrected*

Comment: **P18, L27: "set to zero."**

Reply: agreed.

Action:*corrected.*

Comment: **P19, L13: Replace "In order to fight..." by "Due to..."**

Reply: We find the original wording more specific.

Action:

Comment: **P19, L27: "...macro timestep"**

Reply: agreed.

Action:*corrected.*

Comment: **P20, L21: I guess you mean Figure 2.**

Reply: yes indeed, thanks for spotting!

Action:*corrected.*

Comment: **P20, L23: Figure 3 has no middle column.**

Reply: I cannot find a middle column either.

Action:*corrected!*

Comment: **Figure 1: The figure, and particularly the descriptions need to be enlarged.**

Reply: Agreed

Action:*Figure as a whole, and particularly the labels have been enlarged, and the figure is now represented in the way that the rightmost colourbar is visible.*

Comment: **Figure 2: Give a color bar for the velocities in the upper/left panel. Caption: "...(upper right panel). Bottom panels show...". And write K_ϕ instead of K_{phi}.**

Reply: This is not easily possible because the length of the arrows and their colours represent a quadratic norm involving vertical and horizontal speeds of different units. No simple unit can be assigned to the length of the arrows. The colours just represent the length of the arrow and are meant simply to guide the eye. Instead we report (in the upper part of the plot) the values of horizontal and vertical windspeed that were used for the calculation of the norm. Agreed for K_{phi}.

Action:K_{phi} *changed to* K_ϕ*.*

Comment: **Figure 3: The lower/right panel is the same as the upper/right - it should show K_z.**

Reply: Thanks for spotting!

Action:*corrected.*

Comment: **Caption: Use $v; w$ instead of $v_\phi; \ldots$ and K_ϕ, to be consistent with the rest of the paper.**

Reply: agreed.

Action:*corrected.*

In the following we include a version of the manuscript with the differences marked. Since, however, the LaTeXdiff program available did not produce a source file which LaTeX would compile, a lot of hand editing was necessary, which naturally is prone to errors. In case of doubt the manuscript file without the marked differences should be used as reference.

Manuscript prepared for Atmos. Chem. Phys.
with version 2015/04/24 7.83 Copernicus papers of the LaTeX class copernicus.cls.
Date: 7 October 2016

Direct Inversion of Circulation and Mixing from Tracer Measurements: I. Method

Thomas von Clarmann[1] **and Udo Grabowski**[1]

[1]Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Karlsruhe, Germany

Correspondence to: T. von Clarmann (thomas.clarmann@kit.edu)

Abstract. From a series of zonal mean global stratospheric tracer measurements sampled in altitude versus latitude, circulation and mixing patterns are inferred by the inverse solution of the continuity equation. As a first step, the continuity equation is written as a tendency equation, which is numerically integrated over time to predict a later atmospheric state, i.e. mixing ratio and air density. The integration is formally performed by multiplication of the initially measured atmospheric state vector by a linear prediction operator. Further, the derivative of the predicted atmospheric state with respect to the wind vector components and mixing coefficients is used to find the most likely wind vector components and mixing coefficients which minimize the residual between the predicted atmospheric state and the later measurement of the atmospheric state. Unless multiple tracers are used, this inversion problem is under-determined, and dispersive behaviour of the prediction further destabilizes the inversion. Both these problems are fought by regularization. For this purpose, a first order smoothness constraint has been chosen. The usefulness of this method is demonstrated by application to various tracer measurements recorded with the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS). This method aims at a diagnosis of the Brewer-Dobson circulation without involving the concept of the mean age of stratospheric air, and related problems like the stratospheric tape recorder, or intrusions of mesospheric air into the stratosphere.

1 Introduction

In the context of climate change, possible changes of the intensity of the Brewer-Dobson circulation ~~are under debate~~have become an important research topic. Climate models predict an intensification of the Brewer-Dobson circulation (Butchart et al., 2006). Engel et al. (2009), however, found a weakly significant slow increase of the mean age of stratospheric air. The latter is defined as the mean time lag between the date of the transition of tropospheric air into the stratosphere and the date when the mixing ratio of a monotonically growing tracer was measured in the air volume under investigation, and its increase hints at a deceleration of the Brewer-Dobson circulation. These measurements have been challenged as not representative (Garcia et al., 2011), and global mean age of air measurements by Stiller et al. (2012) suggest that the true picture is not that one-dimensional. Instead, stratospheric age trends vary with altitude and with latitude. Determination of the age of air and its use as diagnostic of the intensity of the Brewer-Dobson circulation, however, has its own limitations: First, due to mixing processes, the age of a stratospheric air volume is not unique but characterized by an age spectrum, which has to be considered since the tropospheric growth of SF_6 mixing ratios is not strictly linear, and on which some *ad hoc* assumptions have to be made (Waugh and Hall, 2002). These include the adequacy of the Wald (inverse Gaussian) function for the representation of the age spectrum and the choice of its width parameter. Second, the most suited age tracer, SF_6, which has significant and monotonic growth rates in the troposphere, is not fully inert: It has a mesospheric sink (Hall and Waugh, 1998; Reddmann et al., 2001) and introduces some age uncertainty when mesospheric air subsides into the stratosphere in the polar winter vortex (Stiller et al., 2008). Third, the determination of the mean age relies upon a reference airmass where the age, by definition, is zero. When the age of air concept was introduced, the reference was simply the troposphere, which is well mixed and thus avoids any related complication (Solomon, 1990; Schmidt and Khedim, 1991). Since the age of air has become a model diagnostic, ~~the modelling community has established~~

~~the upper edge of the tropical tropopause layer~~ parts of the community have established the stratospheric entry point as a reference (Hall and Plumb, 1994), which makes a difference due to the slow ascent of air through the tropical tropopause layer (Fueglistaler et al., 2009). For model validation, however, this redefined age of air is of limited use, because no long measured time series of tracer mixing ratios are available there.

Facing these difficulties, it is desirable to infer the atmospheric circulation directly from tracer measurements, without going back to the age of air concept. Multiple approaches have been developed to infer windfields from measured atmospheric state variables. Sequential data assimilation, and in its optimal form, the extended Kalman filter approach (e.g. Ghil and Malanotte-Rizzoli, 1991; Ghil, 1997), calculates the optimal average of the forecasted meteorological variables for the time of the observation and the observed meteorological variables themselves and uses this average to initialize the next forecast step. The wind field is calculated by a dynamical model. This method involves the generalized inversion of the observation operator where the forecast is used as a constraint. In contrast, so-called[1] variational data assimilation minimizes the residual between the forecasted and the measured atmospheric state variables by optimally adjusting the initialization of the forecast via inversion of an adjoint forecast model, constrained by some background state (Thompson, 1969). Both approaches rely on dynamical models[2] and are suited to infer the most probable atmospheric state variables rather than the windfield, which is a by-product of the assimilation. The windfield or atmospheric circulation can also be inferred directly by kinematic methods from tracer measurements. Such methods rely solely on the continuity equation, do not involve a dynamic model, and thus do not depend on any *ad hoc* parametrisation of effects which are either not resolved by the discrete model, computationally too expensive for explicit modeling, or simply not well understood. While this work is targeted primarily at an assessment of the Brewer-Dobson circulation, its applicability is much wider and includes stratospheric-tropospheric exchange, the mesospheric overturning circulation and others. Early approaches to infer the circulation from tracer measurements include Holton and Choi (1988) as well as Salby and Juckes (1994) who used approaches which share several ideas with ours. Direct inversion of wind speeds from tracer measurements in a volcano plume has, e.g., been suggested by Krueger et al. (2013), however without consideration of mixing. The continuity equation including diffusion terms has been exploited by Wofsy et al. (1994) for assessment of diffusion of stratospheric aircraft exhaust.

[1]The term ‘so-called’ is used here, because it is challenged that this method is really variational in the context of discrete variables (Wunsch, 1996)[p368].

[2]This statement refers to meteorological data assimilation. Chemical data assimilation uses chemistry transport models.

In this paper we present a method to infer two-dimensional (latitude/altitude) circulation and mixing coefficients from subsequent measurements of inert tracers. The application of this method, i.e., the inference of the Brewer-Dobson circulation from global SF_6 distributions (Stiller et al., 2008, 2012) measured with the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), is presented in a companion paper. In order to avoid that the reader does not see the forest for the trees, we give a short overview of our method in Section 2. The prediction of pressure and tracer mixing ratio fields on the basis of the continuity equation and related error estimation is described in Section 3. The estimation of circulation and mixing coefficients by inversion of the continuity equation is presented in Section 4. In Section 5, the applicability of the method and the need of further refinements is critically discussed. The benefits of the method are discussed in Section 6. The paper concludes with recommendations how these results should be used and with an outlook on future work (Section 7). Changes of the Brewer-Dobson circulation during 2002-2012, i.e. the MIPAS mission, are currently investigated by means of this method and will be published in a subsequent paper.

2 General concept

Knowing the initial state of the atmosphere in terms of mixing ratio and air density distributions, wind speed and mixing coefficients at each gridpoint, a future atmospheric state can be predicted with respect to the distribution of any inert tracer. This procedure we call the forward problem. If no ideal tracers are available, source and sink terms of related species have to be included in the forward model. The goal of this work is to invert the forward model in order to infer the circulation and mixing coefficients from tracer measurements by minimization of the residual between the predicted and measured atmospheric state. This approach is complementary to free running climate models because it makes no assumptions about atmospheric dynamics except the validity of the continuity equation. It is further considered more robust than age-of-air analysis (Stiller et al., 2012) because it does not depend on a reference point where the age is assumed zero, nor does it require knowledge on the history of an air parcel.

Our concept involves the following operations. First, a general solution of the forward problem is formulated (Section 3). The forward problem is the solution of the prediction equation as a function of the initial atmospheric state for given winds and mixing coefficients. For our chosen solver, which involves the MacCormack (1969) integration scheme for the solution of the transport problem (Eqs. 5 –10), the relevant dependencies of the final state on the initial state are reported in Section 3.2 (Eqs. 15–26). The formulation in matrix notation (Eqs. 27–28) allows the easy handling of multiple successive timesteps (Eq.29) and an easy estimation of the prediction error via generalized Gaussian error estimation (Eq. 30). As a next step, the dependence of the predicted state on winds and mixing coefficients is estimated for a given initial state . This is achieved by differentiation of the solution of the prediction equation with respect to winds and mixing coefficients (Eqs 34–eq.76). These partial derivatives form the Jacobian matrix of the problem, with which the estimation of winds and mixing coefficients can be reduced to a constrained least square optimization problem where the inversely variance-weighted residual between the predicted atmospheric state and the respective measured atmospheric state is minimized. The latter step involves the generalized inverse of the Jacobian matrix (Eqs. 78–90).

3 The forward problem

The forward model reads the measured atmospheric state at time t and predicts the atmospheric state (number density of air, c, and volume mixing ratios, vmr) at time $t+\Delta t$ for given wind vectors and mixing coefficients representing the time interval $[t; t+\Delta t]$ by solving the continuity equation. The continuity equation allows to calculate the local tendencies of the number densities and volume mixing ratios. These local tendencies $\frac{\partial \rho}{\partial t}$ and $\frac{\partial vmr}{\partial t}$ are then integrated over time to give the new number densities and mixing ratios.

3.1 The continuity equation

The local change of number density ρ of air is in spherical coordinates (for all auxiliary calculations, see supplement):

$$\frac{\partial \rho}{\partial t} = -\frac{1}{r}\frac{\partial \rho v}{\partial \phi} + \frac{\rho v}{r}\tan(\phi) - \frac{\partial \rho w}{\partial z} - \frac{2\rho w}{r} - \frac{1}{r\cos(\phi)}\frac{\partial \rho u}{\partial \lambda} \quad (1)$$

where

t	=	time
λ	=	longitude
ϕ	=	latitude
z	=	altitude above surface
r	=	$r_E + z$
r_E	=	radius of Earth
u	=	$(r_E + z)\cos\phi \, d\lambda/dt$
v	=	$(r_E + z) d\phi/dt$
w	=	dz/dt

Here the shallowness approximation (Hinkelmann (1951); Phillips (1966), quoted after Kasahara (1977)), which ~~is~~simplifies the equations using the assumption that z is much smaller than r_E and which is, often implicitly, used in the usual textbooks on atmospheric sciences (e.g. Brasseur and Solomon, 2005, their Eq. 3.46a), is intentionally not used for reasons which will become clear in Section 3.2.

The local change of the volume mixing ratio of gas g can be calculated from known velocities and mixing coefficients as well as source/sink terms as

$$\begin{aligned}\frac{\partial vmr_g}{\partial t} = & \frac{S_g}{\rho} - \frac{u}{r\cos\phi}\frac{\partial vmr_g}{\partial \lambda} - \frac{v}{r}\frac{\partial vmr_g}{\partial \phi} \\ & -w\frac{\partial vmr_g}{\partial z} + \frac{1}{r^2}\frac{\partial}{\partial \lambda}\left[\frac{K_\lambda}{\cos^2\phi}\frac{\partial vmr_g}{\partial \lambda}\right] \\ & + \frac{1}{r^2\cos\phi}\frac{\partial}{\partial \phi}\left[K_\phi \cos\phi \frac{\partial vmr_g}{\partial \phi}\right] \\ & + \frac{1}{r^2}\frac{\partial}{\partial z}\left[r^2 K_z \frac{\partial vmr_g}{\partial z}\right]\end{aligned} \quad (2)$$

where

vmr_g	=	volume mixing ratio of species g
K_λ	=	zonal diffusion coefficient
K_ϕ	=	meridional diffusion coefficient
K_z	=	vertical diffusion coefficient
S_g	=	the production/loss rate of species g in terms of number density over time

(Brasseur and Solomon (e.g. 2005, Eq. 3.46b) and Jones et al. (2007)).

Since we are only interested in a two-dimensional representation of the atmosphere in altitude and latitude coordinates, zonal advection and mixing terms are ignored in Eqs. (1–2). In this two-dimensional representation, all atmospheric state variables represent zonal mean values~~, and the diffusion coefficient K_ϕ does not only describe physical diffusion but also eddy diffusion components arising from the symmetric part of the two-dimensional eddy tensor. Accordingly, the velocities are no longer windspeed only but include also an eddy component arising from the antisymmetric part of the eddy tensor (Ko et al., 1985, and references therein) and thus are effective transport velocities. Therefore, the transition from the 3D to the 2D system and involved absorption of eddy flux terms in mixing coefficients~~. The kinematic variables, viz. the velocities and ~~velocities implies a re-interpretation of the relevant quantities. Details of the transition from the three-dimesnional to the two-dimensional problem are discussed in Appendix A~~mixing coefficients, have to be re-interpreted because they do not represent merely the zonally averaged velocities and mixing coefficients. Instead, they include also eddy transport and diffusion terms (see Appendix A for details of the interpretation of the kinematic quantities). The local change of number density ρ of air in a two-dimensional atmosphere thus is

$$\frac{\partial \rho}{\partial t} = -\frac{1}{r}\frac{\partial \rho v}{\partial \phi} + \frac{\rho v}{r}\tan(\phi) - \frac{\partial \rho w}{\partial z} - \frac{2\rho w}{r} \quad (3)$$

and the local change of vmr_g is calculated as

$$\frac{\partial vmr_g}{\partial t} = \frac{S_g}{\rho} - \frac{v}{r}\frac{\partial vmr_g}{\partial \phi} - w\frac{\partial vmr_g}{\partial z} + \frac{1}{r^2\cos\phi}\frac{\partial}{\partial \phi}\left[K_\phi \cos\phi \frac{\partial vmr_g}{\partial \phi}\right] + \frac{1}{r^2}\frac{\partial}{\partial z}\left[r^2 K_z \frac{\partial vmr_g}{\partial z}\right] \quad (4)$$

~~In order to comply with the continuity equation, zonal averages of vmr_g have to be calculated number-density-weighted.~~

3.2 Integration of tendencies

Integration of Eqs (3–4) is performed numerically for timesteps of Δt_p. For practical reasons, processes (advection, diffusion, sinks) are splitted, i.e. the tendencies due to these three classes of processes are integrated independently. The timesteps Δt_p used for the integration are chosen smaller than the time difference Δt between two measurements, in order not to clash with the Courant limit (Courant et al., 1952). In the following we call Δt_p 'micro time increment' and the latter 'macro time increment'. The atmospheric state after a macro time increment is predicted by successive prediction over the micro time increment. In the following, index i designates time t, $i+1$ designates the time $t+\Delta t_p$, etc, and I designates the time after the final micro time increment, i.e. the next macro time increment.

For the discrete integration of the advection part of the tendencies the MacCormack (1969) method is used in a generalized multidimensional version similar to the one described by (Perrin and Hu, 2006). This is a predictor-corrector method. For a general state variable $c(t,x,y) = c_i(x,y)$ at location (x,y), and time t with $e(c)$ and $f(c)$ being functions of c, an equation of the form

$$\frac{\partial c}{\partial t} + \frac{\partial e(c)}{\partial x} + \frac{\partial f(c)}{\partial y} = 0 \quad (5)$$

is solved by preliminary predictions of the state variable as a first step: x is

$$c^*_{i+1}(x,y) = c_i(x,y) - \frac{\Delta t_p}{\Delta x}\left(e_i(x+\Delta x,y) - e_i(x,y)\right) - \frac{\Delta t_p}{\Delta y}\left(f_i(x,y+\Delta y) - f_i(x,y)\right). \quad (6)$$

These are then used in a subsequent correction step which gives the final prediction:

$$c_{i+1}(x,y) = \frac{1}{2}\left[\, c_i(x,y) + c^*_{i+1}(x,y) - \frac{\Delta t_p}{\Delta x}\left(e(c^*_{i+1},x,y) - e(c^*_{i+1},x-\Delta x,y)\right) - \frac{\Delta t_p}{\Delta y}\left(f(c^*_{i+1},x,y) - f(c^*_{i+1},x,y-\Delta y)\right)\right] \quad (7)$$

Application to the continuity equation in spherical coordinates requires reformulation of Eq. (3) (c.f., e.g., Chang and St.-Maurice, 1991):

$$\frac{\partial r^2\rho\cos(\phi)}{\partial t} = -\frac{\partial r\rho\, v\, \cos(\phi)}{\partial \phi} - \frac{\partial r^2 \rho\, w\, \cos(\phi)}{\partial z} \quad (8)$$

The predictor of $r^2\rho\cos(\phi)$ is then calculated as

$$[r^2\rho_{i+1}(\phi,z))\cos(\phi)]^* = r^2\rho_i(\phi,z)\cos(\phi) - \frac{\Delta t_p}{\Delta \phi}\left(r\rho_i(\phi+\Delta\phi,z)\cos(\phi+\Delta\phi)v(\phi+\Delta\phi,z) - r\rho_i(\phi,z)\cos(\phi)v(\phi,z)\right) - \frac{\Delta t_p}{\Delta z}\left((r+\Delta z)^2\rho_i(\phi,z+\Delta z)\, w_{\phi,z+\Delta z}\, \cos(\phi) - r^2\rho_i(\phi,z)\, w_{\phi,z}\, \cos(\phi)\right) \quad (9)$$

and the corrected prediction for ρ then gives

$$\rho_{i+1}(\phi,z) = \frac{1}{2r^2\cos(\phi)} \times \Bigg[\rho_i(\phi,z)r^2\cos(\phi) + \left[\rho_{i+1}(\phi,z)r^2\cos(\phi)\right]^* - \frac{\Delta t_p}{\Delta\phi}\left[\left[\rho_{i+1}(\phi,z)rv(\phi,z)\cos(\phi)\right]^* - \left[\rho_{i+1}(\phi-\Delta\phi,z)rv(\phi-\Delta\phi,z)\cos(\phi-\Delta\phi)\right]^*\right] - \frac{\Delta t_p}{\Delta z}\left[\left[\rho_{i+1}(\phi,z)r^2 w(\phi,z)\cos(\phi)\right]^* - \left[\rho_{i+1}(\phi,z-\Delta z)(r-\Delta z)^2 w(\phi,z-\Delta z)\cos(\phi)\right]^*\right]\Bigg] \quad (10)$$

For the local change of mixing ratio, operator splitting is performed. The horizontal and vertical advective parts of the continuity equation for mixing ratios in two dimensions are transformed into the following Mac-Cormack-integrable forms:

$$\left[\frac{\partial \frac{r\, vmr}{v}}{\partial t}\right]_{adv.horiz} = -\frac{\partial vmr_g}{\partial \phi} \quad (11)$$

and

$$\left[\frac{\partial \frac{vmr_g}{w}}{\partial t}\right]_{adv.vert} = \frac{\partial vmr_g}{\partial z}, \quad (12)$$

respectively.

For the diffusive component we use simple Eulerian integration:

$$\left[vmr_{g;i+1}(\phi,z) - vmr_{g;i\phi,z_{g;i}(\phi,z)}\right]_{\text{diff}} = \frac{\Delta t_p}{2r^2(\Delta\phi)^2\cos(\phi)} \cdot \Bigg[(K_\phi(\phi,z)+K_\phi(\phi+\Delta\phi))\cos(\phi+\frac{\Delta\phi}{2}) (vmr_{g;i}(\phi+\Delta\phi,z)-vmr_{g;i}(\phi,z)) - (K_\phi(\phi,z)+K_\phi(\phi-\Delta\phi))\cos(\phi-\frac{\Delta\phi}{2})\cdot (vmr_{g;i}(\phi,z)-vmr_{g;i}(\phi-\Delta\phi,z))\Bigg] + \frac{\Delta t_p}{2r^2(\Delta z)^2}\Bigg[(r+\frac{\Delta z}{2})^2\cdot (K_z(\phi,z)+K_z(\phi,z+\Delta z))\cdot (vmr_{g;i}(\phi,z+\Delta z)-vmr_{g;i}(\phi,z)) - (r-\frac{\Delta z}{2})^2\cdot(K_z(\phi,z)+K_z(\phi,z-\Delta z))\cdot (vmr_{g;i}(\phi,z)-vmr_{g;i}(\phi,z-\Delta z))\Bigg] \quad (13)$$

Sinks of the species considered here are treated as unimolecular processes (c.f., e.g. Brasseur and Solomon, 2005, their Eq. 2.27d) and integrated as

$$\rho_{g;i+1} = \rho_{g;i}e^{-k_g\Delta t_p} \quad (14)$$

where k_g is the sink strength of the gas g.

The abundance of gas g after time-step Δt_p is then simply the sum of the increments due to horizontal and vertical advection, diffusion, and chemical losses.

Admittedly, there exist more elaborate advection schemes than the one used here. However, the need to provide the Jacobians needed in Sections 3.3–4 justifies a reasonable amount of simplicity. Further, numerical errors cannot easily accumulate, because after each timestep Δt, the system is re-initialized with measured data.

Since we do not have a closed system but have mass exchange and mixing with the atmosphere below the lowermost model altitude and above the uppermost altitude, the atmospheric state is not predicted for the lowermost and uppermost altitudes. Prediction is only possible from the second to one below the uppermost altitude. This restricted altitude range we henceforth call 'nominal altitude range'. Instead, the atmospheric state of the uppermost and lowermost altitude is estimated by linear interpolation of measured values at times t and $t+\Delta t$ and used as boundary condition for prediction within the nominal altitude range. Alternatively, derivatives at the border can be approximated by asymmetric difference quotients.

We use the following convention: Atmospheric state variables are sampled on a regular latitude-altitude grid. For some gridpoints, no valid measurements may be available but we assume that, for each state variable, we have a contiguous subset of this grid with valid measurements. For state variable g, we have a total of J_g valid measurements within the 'nominal altitude range', each denoted by index j. A state variable in this context is either air density $\boldsymbol{\rho}$ ($g=0$) or the mixing ratio of one species $\boldsymbol{vmr}_g$. The nominal altitude range at latitude ϕ is the altitude range where, for each gridpoint, a valid measurement is available at the gridpoint itself, and for its northern, southern, upper, lower and diagonal neighbours. The use of asymmetric difference quotients can be emulated by extrapolation,generating artificial border values byextrapolation , which guarantee that each gridpoint within the nominal altitude and latitude range has all required northern, southern, upper, lower and diagonal neighbours. The availability of neighbour-values is necessary to allow the calculation of numerical derivatives of the state variable with respect to latitude and altitude. Further, we We therefore have K_g border elements of each quantity g, each denoted by index k. And finallyThis implies, for each state variable g, we have a total of $L_g = J_g + K_g$ gridpoints ,with indices l.

3.3 Integration in operator notation

For further steps (error propagation and the solution of the inverse problem) it is convenient to rewrite the prediction of air density and mixing ratios in matrix notation. For this purpose, we differentiate the predicted air densities (Eq. 10) and mixing ratios (Eqs. 11–13) with respect to air density and mixing ratios of the gases under assessment at all relevant locations. The sensitivities of the densities of the first predictive step with respect to the initial densities at the same latitude and altitude are

$$\frac{\partial\rho_{i+1}(\phi,z)}{\partial\rho_i(\phi,z)} = \frac{1}{2}\Bigg[2 - \frac{\Delta t_p}{\Delta\phi}\Bigg[\frac{v(\phi,z)}{r}\left(\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi,z)}{r}+\frac{\Delta t_p}{\Delta z}w(\phi,z)\right) + \frac{\Delta t_p}{\Delta\phi}\frac{v(\phi-\Delta\phi,z)v(\phi,z)}{r^2}\Bigg] - \frac{\Delta t_p}{\Delta z}\Bigg[w(\phi,z)\left(\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi,z)}{r}+\frac{\Delta t_p}{\Delta z}w(\phi,z)\right) + \frac{\Delta t_p}{\Delta z}w(\phi,z-\Delta z)w(\phi,z)\Bigg]\Bigg] \quad (15)$$

We further differenciate differentiate predicted air densities with respect to air densities at the adjacent southern latitude but the same altitude.

$$\frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi-\Delta\phi,z)} = \frac{1}{2}\left[\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi-\Delta\phi,z)}{r}\cdot\frac{\cos(\phi-\Delta\phi)}{\cos(\phi)} \left(1+\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi-\Delta\phi,z)}{r}+\frac{\Delta t_p}{\Delta z}w(\phi-\Delta\phi,z)\right)\right] \quad (16)$$

The derivative of the predicted air densities with respect to air densities at the adjacent northern latitude but the same altitude is

$$\frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi+\Delta\phi,z)} = \frac{1}{2}\left[\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi+\Delta\phi,z)}{r}\cdot\frac{\cos(\phi+\Delta\phi)}{\cos(\phi)} \left(-1+\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi,z)}{r}+\frac{\Delta t_p}{\Delta z}w(\phi,z)\right)\right]. \quad (17)$$

As a next step we ~~differenceiate~~ differentiate predicted air densities with respect to the initial air densities at the next higher altitude but the same latitude.

$$\frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi,z+\Delta z)} = \frac{1}{2}\left[\frac{\Delta t_p}{\Delta z}\cdot\frac{(r+\Delta z)^2}{r^2}w(\phi,z+\Delta z)\cdot \left(-1+\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi,z)}{r}+\frac{\Delta t}{\Delta z}w(\phi,z)\right)\right]. \quad (18)$$

The derivative of the predicted air densities with respect to the initial air densities at the next lower altitude but the same latitude is

$$\frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi,z-\Delta z)} = \frac{1}{2}\left[\frac{\Delta t_p}{\Delta z}w(\phi,z-\Delta z)\frac{(r-\Delta z)^2}{r^2}\cdot \left(1+\frac{\Delta t_p}{\Delta\phi}\frac{v(\phi,z-\Delta z)}{r-\Delta z}+\frac{\Delta t_p}{\Delta z}w(\phi,z-\Delta z)\right)\right]. \quad (19)$$

Finally we differentiate the predicted air densities with respect to the initial air densities at the adjacent southern latitude and higher altitude

$$\frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi-\Delta\phi,z+\Delta z)} = -\frac{1}{2}\left[\frac{v(\phi-\Delta\phi,z)}{r}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\frac{\Delta t_p}{\Delta z}\cdot\frac{(r+\Delta z)^2}{r^2}\cdot \frac{\cos(\phi-\Delta\phi)}{\cos(\phi)}w(\phi-\Delta\phi,z+\Delta z)\right] \quad (20)$$

and vice versa

$$\frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi+\Delta\phi,z-\Delta z)} = -\frac{1}{2}\left[w(\phi,z-\Delta z)\frac{\Delta t_p}{\Delta z}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\frac{(r-\Delta z)^2}{r^2}\cdot \frac{\cos(\phi+\Delta\phi)}{\cos(\phi)}\cdot\frac{v(\phi+\Delta\phi,z-\Delta z)}{r-\Delta z}\right] \quad (21)$$

where i is the index of the time increment, and where $\phi\pm\Delta\phi$ and $z\pm\Delta z$ refer to the adjacent model gridpoints in latitude and altitude, respectively.

For mixing ratios, the respective derivatives are:

$$\begin{aligned}\frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z)} = 11- & \\ -\left(\frac{\Delta t_p}{\Delta\phi}\right)^2\cdot\frac{v(\phi,z)}{r^2}\cdot\frac{1}{2}\left[v(\phi,z)+v(\phi-\Delta\phi,z)\right] & \\ -\left(\frac{\Delta t_p}{\Delta z}\right)^2\cdot w(\phi,z)\cdot\frac{1}{2}\left[w(\phi,z)+w(\phi,z-\Delta z)\right] & \\ -\frac{\Delta t_p}{2r^2(\Delta\phi)^2\cos(\phi)}\left[\left(K_\phi(\phi,z)+K_\phi(\phi+\Delta\phi,z)\right)\cdot\cos\left(\phi+\frac{\Delta\phi}{2}\right)\right. & \\ \left.+\left(K_\phi(\phi,z)+K_\phi(\phi-\Delta\phi,z)\right)\cos\left(\phi-\frac{\Delta\phi}{2}\right)\right] & \\ -\frac{\Delta t_p}{2r^2(\Delta z)^2}\left[\left(r+\frac{\Delta z}{2}\right)^2\left(K_z(\phi,z)+K_z(\phi,z+\Delta z)\right)\right. & \\ \left.+\left(r-\frac{\Delta z}{2}\right)^2\left(K_z(\phi,z)+K_z(\phi,z-\Delta z)\right)\right] & \\ -Loss(month,\phi,z)\Delta t_p; &\end{aligned} \quad (22)$$

$$\begin{aligned}\frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi+\Delta\phi,z)} = & \\ -\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi,z)}{2r}\left(1-\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi,z)}{r}\right) & \\ +\frac{\Delta t_p}{2r^2(\Delta\phi)^2\cos(\phi)}\cdot & \\ \left(K_\phi(\phi,z)+K_\phi(\phi+\Delta\phi,z)\right)\cos(\phi+\frac{\Delta\phi}{2}); &\end{aligned} \quad (23)$$

$$\begin{aligned}\frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi-\Delta\phi,z)} = & \\ \frac{v(\phi,z)}{2r}\cdot\frac{\Delta t_p}{\Delta\phi}\left(1+\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi-\Delta\phi,z)}{r}\right) & \\ +\frac{\Delta t_p}{2r^2(\Delta\phi)^2\cos(\phi)}\cdot & \\ \left(K_\phi(\phi,z)+K_\phi(\phi-\Delta\phi,z)\right)\cos(\phi-\frac{\Delta\phi}{2}); &\end{aligned} \quad (24)$$

$$\begin{aligned}\frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z+\Delta z)} = & \\ -\frac{1}{2}\cdot w(\phi,z)\cdot\frac{\Delta t_p}{\Delta z}\left(1-\frac{\Delta t_p}{\Delta z}w(\phi,z)\right) & \\ +\frac{\Delta t_p}{2r^2(\Delta z)^2}(r+\frac{\Delta z}{2})^2\left(K_z(\phi,z)+K_z(\phi,z+\Delta z)\right); &\end{aligned} \quad (25)$$

$$\frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z-\Delta z)} = \frac{\Delta t_p}{\Delta z}\cdot\frac{1}{2}\cdot w(\phi,z)\left(1+w(\phi,z-\Delta z)\frac{\Delta t_p}{\Delta z}\right) + \frac{\Delta t_p}{2r^2(\Delta z)^2}(r-\frac{\Delta z}{2})^2\left(K_z(\phi,z)+K_z(\phi,z-\Delta z)\right), \tag{26}$$

where $Loss(month,\phi,z)$ is the relative loss rate in the respective month at latitude ϕ and altitude z. These derivatives are simplifications in a sense that they do not consider the full chemical Jacobian but assume instead that the source strength depends on no other concentration than the actual concentration of the same species. For the typical long-lived so-called tropospheric source gases considered here, like SF_6 or CFCs, this assumption is appropriate. Pretabulated loss rates are used which have been calculated by locally integrating loss rates over an entire month at a time resolution adequate to resolve the diurnal cycle. From the monthly losses, the $Loss(month,\phi,z)$ values, which are the contribution of losses to the partial derivatives of the local mixing ratios with respect to the initial local mixing ratios, are calculated as the secant of the local decay curve.

With these expressions, the prediction of air density and volume mixing ratio can be rewritten in matrix notation for a single micro time increment~~:~~. This notation simplifies the estimation of the uncertainties of the predicted atmospheric state and the inversion of the prediction equation. In matrix notation, the prediction reads

$$\boldsymbol{\rho}_{i+1} = \begin{pmatrix} \rho_{I;k=1,K_0} \\ \rho_{i+1;k=1,K_0} \\ \rho_{i+1;j=K_0+1,L_0} \end{pmatrix}_{i+1} = \mathbf{D}_{\rho;i}\boldsymbol{\rho}_i = \begin{pmatrix} \mathbf{I}_K & \mathbf{0} & \mathbf{0} \\ & \mathbf{W}_i & \mathbf{0} \\ \mathbf{0} & & \mathbf{D}_{\rho,\mathrm{nom}} \end{pmatrix} \begin{pmatrix} \rho_{I;k=1,K_0} \\ \rho_{i;k=1,K_0} \\ \rho_{i;j=K_0+1,L_0} \end{pmatrix} \tag{27}$$

where

$\mathbf{D}_{\rho;i}$	is the ~~$L_0\times L_0$~~ $(L_0+K_0)\times(L_0+K_0)$ Jacobian matrix of air density for time increment i, i.e. the sensitivities of the prediction with respect to the initial state, $\frac{\partial c_{i+1,m}}{\partial c_{i,n}}$, here m and n run over the model gridpoints
$\mathbf{I}_K$	is $K_0\times K_0$ identity matrix;
$\mathbf{0}$	are zero submatrices of the required dimensions;
$\mathbf{W}_i$	is a $K_0\times 2K_0$-dimensional interpolation matrix;
$\mathbf{D}_{\rho,\mathrm{nom}}$	is an $J_0\times L_0$ Jacobian containing the partial derivatives $\partial\rho_{i+1;j}/\partial\rho_{i;l}$, applied to the nominal altitude range;
$\boldsymbol{\rho}_{I;k=1,K_0}$	is the K_0-dimensional vector of air densities in the border region after the final timestep, i.e. for the time of the next measurement;
$\boldsymbol{\rho}_{i;k=1,K_0}$	is the K_0-dimensional vector of air densities in the border region at the current timestep as resulting from interpolation in time;
$\boldsymbol{\rho}_{i;j=K_0+1,L_0}$	is the ~~K_0~~ J_0-dimensional vector of air densities in the nominal region at the current timestep as resulting from integration according to the MacCormack scheme as described above.

The operation of these sub-matrices is illustrated in the uppermost three (violet, green, blue) blocks of Figure 1.

Since the source term depends on air density, the integration in matrix notation for vmr requires simultaneous treatment of vmr and air density, and we get, using notation accordant with air density:

$$\begin{pmatrix} \boldsymbol{\rho}_{i+1} \\ \boldsymbol{vmr}_{i+1} \end{pmatrix} = \begin{pmatrix} \rho_{I;k=1,K_0} \\ \rho_{i+1;k=1,K_0} \\ \rho_{i+1;j=K_0+1,L_0} \\ vmr_{g;I;k=1,K_g} \\ vmr_{g;i+1;k=1,K_g} \\ vmr_{g;i+1;j=K_g+1,L_g} \end{pmatrix} = \mathbf{D}_i \begin{pmatrix} \boldsymbol{\rho_i} \\ \boldsymbol{vmr}_{g;i} \end{pmatrix} = \begin{pmatrix} \mathbf{D}_{\rho;i} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & & \mathbf{W}_i & \mathbf{0} \\ & \mathbf{D}_{\mathrm{g,nom}} & & \end{pmatrix} \begin{pmatrix} \rho_{I,k=1,K_0} \\ \rho_{i;k=1,K_0} \\ \rho_{i;j=K_0+1,L_0} \\ vmr_{g;I,k=1,K_g} \\ vmr_{g;i,k=1,K_g} \\ vmr_{g;i,j=K_g+1,L_g} \end{pmatrix} \tag{28}$$

where $\mathbf{D}_i$ is the total Jacobian with respect to air densities and all involved gas mixing ratios, and where g runs over all gases. Note that

1. The Jacobian $\mathbf{D}_i$ is time-dependent because it includes submatrices controlling the interpolation between the

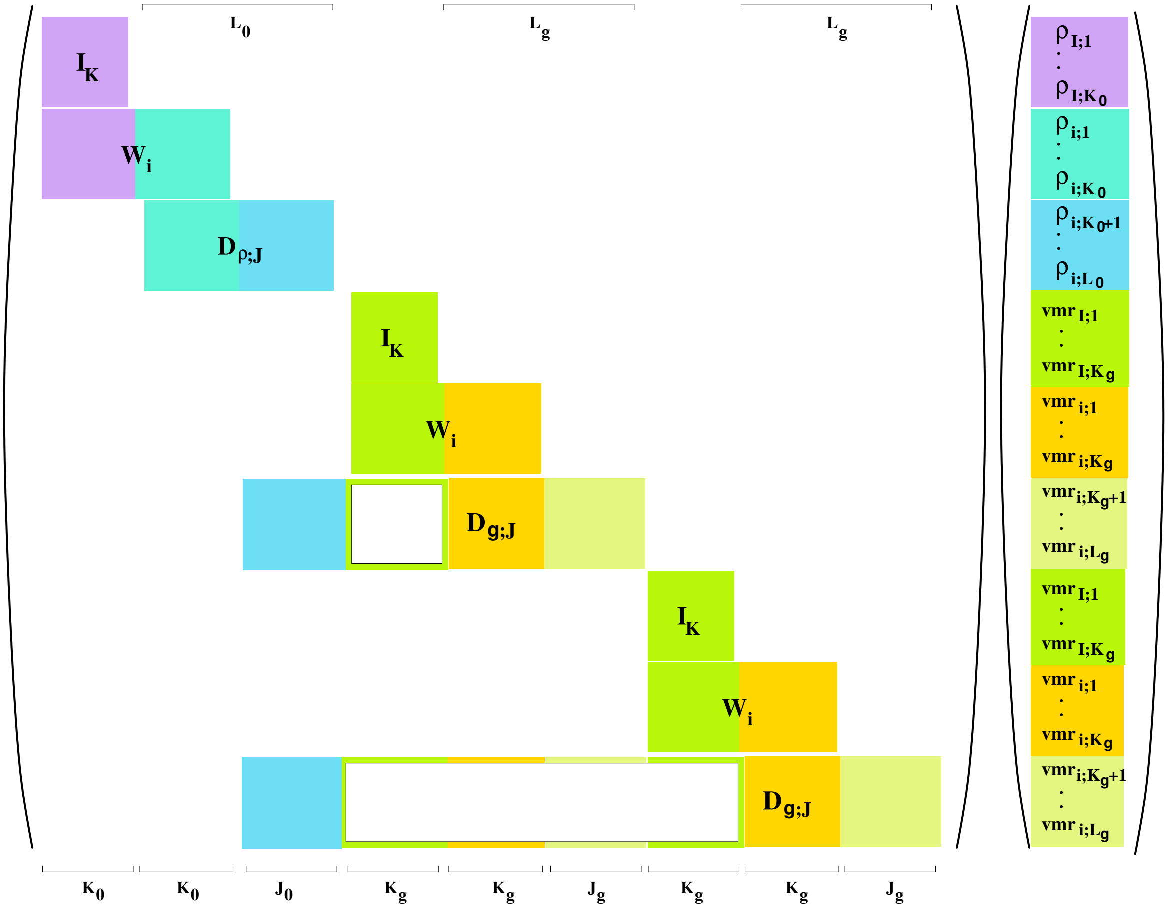

Figure 1. : Matrix structure of the right-hand part of Eq. 28. Colours indicate which matrix blocks operate on which part of the input vector. An example with two gases is shown.

initial time and the end time. In the case of vmr, a further time dependence is introduced by the time-dependent source function.

2. the first ‘row’ of the Jacobian matrix includes identity $\mathbf{I}_K$ because the prediction is not supposed to change the measured $\boldsymbol{\rho}_I$ and $\boldsymbol{vmr}_I$ at the end of the macro time increment. This value is used to construct the boundary condition. Row is here written in quotes because the elements of this ‘row’ are matrices in themselves. Introduction of unity Jacobian elements is necessary because Eqs. (27–28) are autonomized, originally non-autonomous systems of differential equations.

3. $\mathbf{W}$ is used to interpolate the boundary state between the initial time of the micro time interval, $t+(i-1)\Delta t_p$, and the time at the end of the time interval $t+\Delta t$ to give the atmospheric state at the border region at time $t+i\Delta t_p$.

4. The Jacobian submatrices $\mathbf{D}_{\rho,\mathrm{nom}}$ and $\mathbf{D}_{g,\mathrm{nom}}$ are used to predict the atmospheric state in the nominal range after one further micro time increment from the atmospheric state at the current time and the boundary condition. Its elements are described in Eqs. (15–21) and (22–26). The part of $\mathbf{D}_{g,\mathrm{nom}}$ which refers to the border mixing ratios ~~($\boldsymbol{vmr}_{g;I,k=1,K_g}$~~ ($vmr_{g;I,k=1,K_g}$) is zero. ~~In the case of multiple species, $\mathbf{D}_{\mathrm{g,nom}}$ has a block diagonal structure.~~ The dimension of $\mathbf{D}_{\rho,\mathrm{nom}}$ is $\sum_g J_g \times (L_0+K_0+\sum_g (L_g+K_g))$.

5. No simple mapping mechanism between the field of atmospheric state variables sampled at latitudes and altitudes and the vectors $\boldsymbol{\rho}$ and $\boldsymbol{vmr}_g$ is provided because the fields are irregular in a sense that the number of relevant altitudes is latitude-dependent. Pointer variables have to be used instead.

The matrix structure is exemplified in Fig. 1. For the macro time increment Δt we get

$$\begin{pmatrix} \boldsymbol{\rho}_I \\ \boldsymbol{vmr}_I \end{pmatrix} = \left(\prod_{i=I}^{1} \mathbf{D}_i\right) \begin{pmatrix} \boldsymbol{\rho}_0 \\ \boldsymbol{vmr}_0 \end{pmatrix} \qquad (29)$$

3.4 Prediction errors

Let $\mathbf{S}_0$ be the $L \times L$ covariance matrix describing the uncertainties of all involved measurements $\boldsymbol{\rho}_0$ and $\boldsymbol{vmr}_0$, with diagonal elements $s_{0;l,l} = \sigma^2_{0;l,l}$ and $L = \sum_0^{gases} L_g$. We assume that these measurement errors in the state variables used for the prediction are the only relevant error sources. With $\mathbf{S}_0$ and $\prod_{i=I}^{1} \mathbf{D}_i$ available, generalized Gaussian error propagation for $\boldsymbol{\rho}_I$ and $\boldsymbol{vmr}_I$ can be easily formulated as:

$$\mathbf{S}_I = \left(\prod_{i=I}^{1} \mathbf{D}_i \right) \mathbf{S}_0 \left(\prod_{i=I}^{1} \mathbf{D}_i \right)^T . \quad (30)$$

Even if $\mathbf{S}_0$ is diagonal, i.e. the initial errors are assumed to be uncorrelated, error propagation through the forward model will generate non-zero error covariances in $\mathbf{S}_I$ representing the atmospheric state at time $t+\Delta t$. $\mathbf{S}_I$ will be needed in the inversion of circulation and mixing coefficients described in Section 4.

3.5 A note on finite resolution measurements

The measurements used are not a perfect image of the true atmospheric state but contain some prior information. In the case of the IMK data, a priori profiles are usually set zero, and the constraint is built with a Tikhonov-type first order finite differences smoothing constraint (c.f. von Clarmann et al. (2009). That means that, besides the mapping of measurement and parameter errors, the only distortion of the truth via the retrieval is reduced altitude resolution; no other effect of the prior information is to be considered. Usually, any comparison between modelled and measured fields requires application of the averaging kernels of the retrieval to the model data in order to account for the smoothing by the constraint of the retrieval (assuming that the model grid is much finer than the resolution of the retrieval).

In our case, the situation is different: The model is initialized with measurements of reduced altitude resolution, and the fields predicted by the model are then compared to measurements of the same altitude resolution. It is fair to assume that the model does not dramatically change the altitude resolution of the profiles, and thus comparable quantities are compared when the residuals between predicted and measured atmospheric state are evaluated.

3.6 A note on numerical mixing

Let the initial mixing ratio field be homogeneous except one point with delta-type excess mixing ratio. Assume further a homogeneous velocity field and zero mixing coefficients. If the velocity is such that the position of the excess mixing ratio is displaced during Δt by a distance which is not equal ~~with~~ to an integer multiple of the gridwidth, then the resulting distribution will no longer be a delta-type distribution but will be smoothed. The widening of the delta peak we refer to as numerical mixing. The MacCormack transport scheme is designed to fight this diffusivity but some higher order effects may still survive. One might think that, during the inversion, the widening is misinterpreted as mixing, leading to too large mixing coefficients.

Again, in our case, the situation is different: The widening does not accumulate over the Δt_p timesteps, because we first calculate the operator $\prod_{i=I}^{1} \mathbf{D}_i$, which is applied only once to the initial field, which avoids accumulation of numerical mixing over timesteps. Still one widening process as described above can occur, when the forward model leads to a position of the new peak which cannot be represented in the grid chosen. However, since the gas distributions $vmr_{g;I}$ at the end of timestep Δt are sampled on the same grid, the maximum in the ~~true~~ real atmosphere would be widened in the same way, and there would be no residual the inversion would try to get rid of by increasing the mixing coefficients. And the next time step Δt is initialized again with measured data, which also excludes accumulation of numerical mixing effects.

These considerations aside, there are other numerical artefacts~~:~~. These are related to the numerical evaluation of partial derivatives of the state variables in our transport scheme chosen. Particularly in the case of delta functions in the state variable field, these cause side-wiggles behind and smearing in front of the transported structure. To keep these artefacts small, it is necessary to set the spatial grid fine enough that every structure is represented by multiple gridpoints.

4 The inverse problem

For convenience, we combine the variables of the initial atmospheric state and the predicted state at the end of the macro time interval, respectively, into the vectors

$$\tilde{\boldsymbol{x}}_0 = \begin{pmatrix} \boldsymbol{\rho}_0 \\ \boldsymbol{vmr}_0 \end{pmatrix} = (\tilde{x}_{0;1} \ldots \tilde{x}_{0;L})^T, \quad (31)$$

and

$$\tilde{\boldsymbol{x}}_I = \begin{pmatrix} \boldsymbol{\rho}_I \\ \boldsymbol{vmr}_I \end{pmatrix} = (\tilde{x}_{I;1} \ldots \tilde{x}_{I;L})^T, \quad (32)$$

The related subsets of $\tilde{\boldsymbol{x}}_0$ and $\tilde{\boldsymbol{x}}_I$ which contain only state variables in the nominal altitude range but not those in the border region are $\boldsymbol{x}_0 = (x_{0;1} \ldots x_{0;J})^T$ and $\boldsymbol{x}_I = (x_{I;1} \ldots x_{I;J})^T$, respectively. The reason why the distinction between $\tilde{\boldsymbol{x}}$ and $\boldsymbol{x}$ is made is that, contrary to the prediction step, for the inversion vector elements related to the interpolation of values in the border region are no longer needed. Further, we combine the fields of meridional and vertical wind components and mixing coefficients into the vector

$$\boldsymbol{q} = \begin{pmatrix} \boldsymbol{v} \\ \boldsymbol{w} \\ \boldsymbol{K}_\phi \\ \boldsymbol{K}_z \end{pmatrix}, \quad (33)$$

and assume constant velocities and mixing coefficients during the macro timestep. To infer circulation patterns and mixing coefficients from the measurements of air densities and mixing ratios, the Jacobian matrix $\mathbf{F}$,

$$\mathbf{F} = (\boldsymbol{f}_1,\ldots,\boldsymbol{f}_N) = (f_{j,n}) = \left(\frac{\partial x_{I,j}}{\partial q_n}\right) = \left(\frac{\partial \boldsymbol{x}_I}{\partial q_1},\ldots,\frac{\partial \boldsymbol{x}_I}{\partial q_N}\right), \tag{34}$$

is needed, where $N = 4J$ where $J = \sum_0^{gases} J_g$, because there are four unknown quantities, v_j, w_j, ~~K_ϕ, and K_z~~ $K_{\phi;j}$, and $K_{z;j}$ at each gridpoint of the nominal region where these variables shall be inferred. The elements of $\mathbf{F}$ are calculated from Eq. (28) by application of the product rule:

$$\tilde{\boldsymbol{f}}_n = \sum_{i=1}^{I}\left[\left(\prod_{k=I}^{i+1}\mathbf{D}_k\right)\left(\frac{\partial \mathbf{D}_i}{\partial q_n}\right)\left(\prod_{k=i-1}^{1}\mathbf{D}_k\right)\tilde{\boldsymbol{x}}_0\right], \tag{35}$$

where the tilde symbol in $\tilde{\boldsymbol{f}}_n$ indicates that the vectors resulting from Eq. (35) still include the border elements which ~~still~~ have to be discarded to obtain $\boldsymbol{f}_n$. The quantity $\tilde{\boldsymbol{f}}_n$is more efficiently computed using the following recursive scheme, where $\tilde{\boldsymbol{f}}_{l,i}$ is the respective column of the Jacobian after micro timestep i:

$$\tilde{\boldsymbol{f}}_{n,i} = \mathbf{D}_i \boldsymbol{f}_{n,i-1} + \frac{\partial \mathbf{D}_i}{\partial q_n}\left(\prod_{k=i-1}^{1}\mathbf{D}_k\right)\tilde{\boldsymbol{x}}_0 \tag{36}$$

With the argument of $\mathbf{D}$ specifying the column of the $\mathbf{D}$-matrix such that $D_{c,i}(\phi,z)$ relates $\rho_{i+1}(\phi,z)$ to $\rho_i(\phi,z)$, $D_{\rho,i}(\phi\pm\Delta\phi,z)$ relates $\rho_{i+1}(\phi,z)$ to $\rho_i(\phi\pm\Delta\phi,z)$, and $D_{\rho,i}(\phi,z\pm\Delta z)$ relates $\rho_{i+1}(\phi,z)$ to $\rho_i(\phi,z\pm\Delta z)$, and for ~~vmr~~ vmr accordingly, the entries of $\mathbf{D}_i$ relevant to $\boldsymbol{v}$ are:

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi,z)}{\partial\rho_i(\phi,z)}}{\partial v(\phi,z)} = -\frac{\Delta t_p}{2\Delta\phi}\cdot \cdot\left(\frac{\Delta t_p}{\Delta\phi}\cdot\frac{2v(\phi,z)+v(\phi-\Delta\phi,z)}{r^2}+2\frac{\Delta t_p}{\Delta z}\cdot\frac{w(\phi,z)}{r}\right) \tag{37}$$

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi+\Delta\phi,z)}{\partial\rho_i(\phi+\Delta\phi,z)}}{\partial v(\phi,z)} = -\frac{1}{2}\cdot\left(\frac{\Delta t_p}{\Delta\phi}\right)^2\cdot\frac{v(\phi+\Delta\phi,z)}{r^2} \tag{38}$$

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi+\Delta\phi,z)}{\partial\rho_i(\phi,z)}}{\partial v(\phi,z)} = \frac{1}{2r}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\frac{\cos(\phi)}{\cos(\phi+\Delta\phi)}\cdot \cdot\left(1+2\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi,z)}{r}+\frac{\Delta t_p}{\Delta z}w(\phi,z)\right) \tag{39}$$

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi,z)}{\partial\rho_i(\phi+\Delta\phi,z)}}{\partial v(\phi,z)} = \frac{1}{2}\cdot\left(\frac{\Delta t_p}{\Delta\phi}\right)^2\cdot\frac{v(\phi+\Delta\phi,z)}{r^2}\cdot\frac{\cos(\phi+\Delta\phi)}{\cos(\phi)} \tag{40}$$

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi-\Delta\phi,z)}{\partial\rho_i(\phi,z)}}{\partial v(\phi,z)} = \frac{1}{2r}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\frac{\cos(\phi)}{\cos(\phi-\Delta\phi)}\cdot \cdot\left(-1+\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi-\Delta\phi,z)}{r}+\frac{\Delta t_p}{\Delta z}w(\phi-\Delta\phi,z)\right) \tag{41}$$

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi,z+\Delta z)}{\partial\rho_i(\phi,z)}}{\partial v(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{\Delta z}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\frac{r^2}{(r+\Delta z)^2}\frac{w(\phi,z)}{r} \tag{42}$$

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi,z)}{\partial\rho_i(\phi,z+\Delta z)}}{\partial v(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{\Delta z}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\frac{(r+\Delta z)^2}{r^3}w(\phi,z+\Delta z) \tag{43}$$

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi+\Delta\phi,z)}{\partial\rho_i(\phi,z+\Delta z)}}{\partial v(\phi,z)} = -\frac{1}{2}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\frac{\Delta t_p}{\Delta z}\cdot \cdot\frac{(r+\Delta z)^2}{r^3}\cdot\frac{\cos(\phi)}{\cos(\phi+\Delta\phi)}w(\phi,z+\Delta z) \tag{44}$$

$$\frac{\partial\frac{\partial\rho_{i+1}(\phi-\Delta\phi,z+\Delta z)}{\partial\rho_i(\phi,z)}}{\partial v(\phi,z)} = -\frac{1}{2}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\frac{\Delta t_p}{\Delta z}\cdot \cdot\frac{r}{(r+\Delta z)^2}\cdot\frac{\cos(\phi)}{\cos(\phi-\Delta\phi)}\cdot w(\phi-\Delta\phi,z) \tag{45}$$

$$\frac{\partial\frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z)}}{\partial v(\phi,z)} = -\left(\frac{\Delta t_p}{\Delta\phi}\right)^2\cdot\frac{1}{r^2}\cdot\left(v(\phi,z)+\frac{v(\phi-\Delta\phi,z)}{2}\right) \tag{46}$$

$$\frac{\partial\frac{\partial vmr_{i+1}(\phi+\Delta\phi,z)}{\partial vmr_i(\phi+\Delta\phi,z)}}{\partial v(\phi,z)} = -\left(\frac{\Delta t_p}{\Delta\phi}\right)^2\cdot\frac{v(\phi+\Delta\phi,z)}{2r^2} \tag{47}$$

$$\frac{\partial\frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi-\Delta\phi,z)}}{\partial v(\phi,z)} = \frac{1}{2r}\cdot\frac{\Delta t_p}{\Delta\phi}\cdot\left(1+\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi-\Delta\phi,z)}{r}\right) \tag{48}$$

$$\frac{\partial\frac{\partial vmr_{i+1}(\phi+\Delta\phi,z)}{\partial vmr_i(\phi,z)}}{\partial v(\phi,z)} = \frac{1}{2r}\left(\frac{\Delta t_p}{\Delta\phi}\right)^2\frac{v(\phi+\Delta\phi,z)}{r} \tag{49}$$

$$\frac{\partial\frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi+\Delta\phi,z)}}{\partial v(\phi,z)} = -\frac{\Delta t_p}{\Delta\phi}\cdot\frac{1}{r}\left(\frac{1}{2}-\frac{\Delta t_p}{\Delta\phi}\cdot\frac{v(\phi,z)}{r}\right) \tag{50}$$

Entries not mentioned here are zero. Entries relevant to $\boldsymbol{w}$ are:

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi,z)}}{\partial w(\phi,z)} = -\frac{\Delta t_p}{\Delta \phi}\cdot\frac{\Delta t_p}{\Delta z}\cdot\frac{v(\phi,z)}{r} - \left(\frac{\Delta t_p}{\Delta z}\right)^2 w(\phi,z) - \frac{1}{2}\left(\frac{\Delta t_p}{\Delta z}\right)^2 w(\phi,z-\Delta z) \tag{51}$$

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi,z+\Delta z)}{\partial \rho_i(\phi,z+\Delta z)}}{\partial w(\phi,z)} = -\frac{1}{2}\left(\frac{\Delta t_p}{\Delta z}\right)^2 w(\phi,z+\Delta z) \tag{52}$$

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi+\Delta\phi,z)}}{\partial w(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{\Delta \phi}\cdot\frac{\Delta t_p}{\Delta z}\cdot\frac{v(\phi+\Delta\phi,z)}{r}\cdot\frac{\cos(\phi+\Delta\phi)}{\cos(\phi)} \tag{53}$$

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi+\Delta\phi,z)}{\partial \rho_i(\phi+\Delta\phi,z-\Delta z)}}{\partial w(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{\Delta \phi}\cdot\frac{\Delta t_p}{\Delta z}\cdot\frac{v(\phi,z)}{r}\cdot\frac{\cos(\phi)}{\cos(\phi+\Delta\phi)} \tag{54}$$

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi,z+\Delta z)}{\partial \rho_i(\phi,z)}}{\partial w(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{\Delta z}\cdot\frac{r^2}{(r+\Delta z)^2}\cdot \cdot\left(1+\frac{\Delta t_p}{\Delta \phi}\cdot\frac{v(\phi,z)}{r}+2\frac{\Delta t_p}{\Delta z}\cdot w(\phi,z)\right) \tag{55}$$

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi,z)}{\partial \rho_i(\phi,z+\Delta z)}}{\partial w(\phi,z)} = \frac{1}{2}\cdot\left(\frac{\Delta t_p}{\Delta z}\right)^2\cdot\frac{(r+\Delta z)^2}{r^2}\cdot w(\phi,z+\Delta z) \tag{56}$$

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi,z-\Delta z)}{\partial \rho_i(\phi,z)}}{\partial w(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{\Delta z}\cdot\frac{r^2}{(r-\Delta z)^2}\cdot \cdot\left(-1+\frac{\Delta t_p}{\Delta \phi}\cdot\frac{v(\phi,z-\Delta z)}{r-\Delta z}+\frac{\Delta t_p}{\Delta z}\cdot w(\phi,z-\Delta z)\right) \tag{57}$$

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi+\Delta\phi,z-\Delta z)}{\partial \rho_i(\phi,z)}}{\partial w(\phi,z)} = -\frac{1}{2}\cdot\frac{\Delta t_p}{\Delta \phi}\cdot\frac{\Delta t_p}{\Delta z}\cdot \cdot\frac{r^2}{(r-\Delta z)^2}\cdot\frac{\cos(\phi)}{\cos(\phi+\Delta\phi)}\frac{v(\phi,z-\Delta z)}{r-\Delta z} \tag{58}$$

$$\frac{\partial \frac{\partial \rho_{i+1}(\phi,z+\Delta z)}{\partial \rho_i(\phi+\Delta\phi,z)}}{\partial w(\phi)} = -\frac{1}{2}\cdot\frac{\Delta t_p}{\Delta \phi}\cdot\frac{\Delta t_p}{\Delta z}\cdot \cdot\frac{r^2}{(r+\Delta z)^2}\cdot\frac{\cos(\phi+\Delta\phi)}{\cos(\phi)}\cdot\frac{v(\phi+\Delta\phi,z)}{r} \tag{59}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z)}}{\partial w(\phi,z)} = -\left(\frac{\Delta t_p}{\Delta z}\right)^2\cdot\left(w(\phi,z)+\frac{w(\phi,z-\Delta z)}{2}\right) \tag{60}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z+\Delta z)}{\partial vmr_i(\phi,z+\Delta z)}}{\partial w(\phi,z)} = -\left(\frac{\Delta t_p}{\Delta z}\right)^2\cdot\frac{w(\phi,z+\Delta z)}{2} \tag{61}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z-\Delta z)}}{\partial w(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{\Delta z}\cdot\left(1+w(\phi,z-\Delta z)\cdot\frac{\Delta t_p}{\Delta z}\right) \tag{62}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z+\Delta z)}{\partial vmr_i(\phi,z)}}{\partial w(\phi,z)} = \frac{1}{2}\cdot\left(\frac{\Delta t_p}{\Delta z}\right)^2\cdot w(\phi,z+\Delta z) \tag{63}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z+\Delta z)}}{\partial w(\phi,z)} = -\frac{1}{2}\cdot\frac{\Delta t_p}{\Delta z}\cdot\left(1-2\frac{(\Delta t_p)}{(\Delta z)}\cdot w(\phi,z)\right) \tag{64}$$

Entries relevant to $\boldsymbol{K}_\phi$ are:

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z)}}{\partial K_\phi(\phi,z)} = -\frac{\Delta t_p}{(\Delta\phi)^2}\cdot\frac{1}{2r^2}\cdot \cdot\frac{\cos\left(\phi+\frac{\Delta\phi}{2}\right)+\cos\left(\phi-\frac{\Delta\phi}{2}\right)}{\cos(\phi)} \tag{65}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi\mp\Delta\phi,z)}{\partial vmr_i(\phi\mp\Delta\phi,z)}}{\partial K_\phi(\phi,z)} = -\frac{\Delta t_p}{(\Delta\phi)^2}\cdot\frac{1}{2r^2}\cdot\frac{\cos\left(\phi\mp\frac{\Delta\phi}{2}\right)}{\cos\left(\phi\mp\Delta\phi\right)} \tag{66}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi-\Delta\phi,z)}}{\partial K_\phi(\phi,z)} = \frac{1}{2r^2}\cdot\frac{\Delta t_p}{(\Delta\phi)^2}\cdot\frac{\cos(\phi-\frac{\Delta\phi}{2})}{\cos(\phi)} \tag{67}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi+\Delta\phi,z)}}{\partial K_\phi(\phi,z)} = \frac{1}{2r^2}\cdot\frac{\Delta t_p}{(\Delta\phi)^2}\cdot\frac{\cos(\phi+\frac{\Delta\phi}{2})}{\cos(\phi)} \tag{68}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi+\Delta\phi,z)}{\partial vmr_i(\phi,z)}}{\partial K_\phi(\phi,z)} = \frac{1}{2r^2}\cdot\frac{\Delta t_p}{(\Delta\phi)^2}\cdot\frac{\cos(\phi+\frac{\Delta\phi}{2})}{\cos(\phi+\Delta\phi)} \tag{69}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi-\Delta\phi,z)}{\partial vmr_i(\phi,z)}}{\partial K_\phi(\phi,z)} = \frac{1}{2r^2}\cdot\frac{\Delta t_p}{(\Delta\phi)^2}\cdot\frac{\cos(\phi-\frac{\Delta\phi}{2})}{\cos(\phi-\Delta\phi)} \tag{70}$$

And finally, entries relevant to $\boldsymbol{K}_z$ are:

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z)}}{\partial K_z(\phi,z)} = -\frac{\Delta t_p}{(\Delta z)^2}\cdot\frac{\left(r+\frac{\Delta z}{2}\right)^2+\left(r-\frac{\Delta z}{2}\right)^2}{2r^2} \tag{71}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z\mp\Delta z)}{\partial vmr_i(\phi,z\mp\Delta z)}}{\partial K_z(\phi,z)} = -\frac{\Delta t_p}{(\Delta z)^2}\cdot\frac{\left(r\mp\frac{\Delta z}{2}\right)^2}{2(r\mp\Delta z)^2} \tag{72}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z+\Delta z)}{\partial vmr_i(\phi,z)}}{\partial K_z(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{(\Delta z)^2}\cdot\frac{\left(r+\frac{\Delta z}{2}\right)^2}{(r+\Delta z)^2} \tag{73}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z-\Delta z)}}{\partial K_z(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{(\Delta z)^2}\cdot\frac{\left(r-\frac{\Delta z}{2}\right)^2}{r^2} \tag{74}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z)}{\partial vmr_i(\phi,z+\Delta z)}}{\partial K_z(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{(\Delta z)^2}\cdot\frac{\left(r+\frac{\Delta z}{2}\right)^2}{r^2} \tag{75}$$

$$\frac{\partial \frac{\partial vmr_{i+1}(\phi,z-\Delta z)}{\partial vmr_i(\phi,z)}}{\partial K_z(\phi,z)} = \frac{1}{2}\cdot\frac{\Delta t_p}{(\Delta z)^2}\cdot\frac{\left(r-\frac{\Delta z}{2}\right)^2}{(r-\Delta z)^2} \tag{76}$$

~~We~~ With these derivatives we linearize the prediction with respect to wind and mixing coefficients, i.e., we linearly predict the new atmospheric state for a given initial state $\boldsymbol{q}_0$ as a function of wind and mixing ratios.

$$\boldsymbol{x}_I = \boldsymbol{x}_0 + \mathbf{F}(\boldsymbol{q}-\boldsymbol{q}_0). \tag{77}$$

This formulation gives access to the winds and mixing ratios via inversion of $\mathbf{F}$.

Assuming linearity and Gaussian statistics, the most likely set $\boldsymbol{q}$ of winds and mixing ratios during the macro time interval minimizes the following cost function:

$$\begin{aligned}\chi_1^2 &= (\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_I)^T\mathbf{S}_\mathrm{r}^{-1}(\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_I) \\ &\approx (\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_0-\mathbf{F}(\boldsymbol{q}-\boldsymbol{q}_0))^T\mathbf{S}_\mathrm{r}^{-1} \\ &\quad (\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_0-\mathbf{F}(\boldsymbol{q}-\boldsymbol{q}_0))\end{aligned} \tag{78}$$

where $\boldsymbol{x}_{\mathrm{m};I}$ is the measured state at the end of the macro time step and $\mathbf{S}_\mathrm{r}$ is the error covariance matrix of the residual, which is, under the assumption that prediction error and measurement errors are uncorrelated, the sum of the prediction covariance matrix and the measurement covariance matrix, both after the macro time step:

$$\mathbf{S}_\mathrm{r} = \mathbf{S}_{\mathrm{m},I} + \mathbf{S}_\mathrm{p}, \tag{79}$$

where $\mathbf{S}_\mathrm{p}$ is an $J\times J$-matrix containing those elements of $\mathbf{S}_I$ which are relevant to $\boldsymbol{x}_I$. $\mathbf{S}_{\mathrm{m},I}$ is the measurement error covariance matrix of the atmospheric state after the macro time step. The minimization of the cost function gives the following estimate $\hat{\boldsymbol{q}}$ of winds and mixing coefficients:

$$\hat{\boldsymbol{q}} = \boldsymbol{q}_0 + \left(\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}(\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_I) \tag{80}$$

The matrix $\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}$ can be singular either because the related system of equations is under-determined or ill-posed due to nearly linearly dependent equations. Singularity is fought by adding the following constraint term to the cost function of Eq. (78):

$$\begin{aligned}\chi^2_\mathrm{con} &= (\boldsymbol{q}-\boldsymbol{q}_a)^T\mathbf{R}(\boldsymbol{q}-\boldsymbol{q}_a) & (81)\\ \chi^2 &= \chi_1^2+\chi^2_\mathrm{con}, & (82)\end{aligned}$$

where $\boldsymbol{q}_a$ is some prior assumption on velocities and mixing coefficients. $\mathbf{R}$ is a $J\times J$ regularization matrix of which the choice is discussed below. From this, the constrained estimate of velocities and mixing coefficients can be inferred:

$$\hat{\boldsymbol{q}} = \boldsymbol{q}_a + \left(\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}+\mathbf{R}\right)^{-1}\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}(\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_I) \tag{83}$$

An equivalent formulation, which is more efficient if the dimension of $\boldsymbol{q}$ is larger than that of $\boldsymbol{x}$ (underdetermined problem), but which requires a non-singular regularization matrix ~~, and which~~ and does not give easy access to diagnostics (see below), is (Rodgers, 2000):

$$\hat{\boldsymbol{q}} = \boldsymbol{q}_a + \mathbf{R}^{-1}\mathbf{F}^T\left(\mathbf{F}\mathbf{R}^{-1}\mathbf{F}^T+\mathbf{S}_{\mathrm{r},I}\right)^{-1}(\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_I). \tag{84}$$

The covariance matrix characterizing the uncertainty of estimated winds and mixing coefficients is

$$\mathbf{S}_q = \left(\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}+\mathbf{R}\right)^{-1}\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}\left(\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}+\mathbf{R}\right)^{-1}, \tag{85}$$

and the estimated winds and mixing coefficients are related to the true ones as

$$\mathbf{A} = \frac{\partial\hat{\boldsymbol{q}}}{\partial\boldsymbol{q}} = \left(\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}+\mathbf{R}\right)^{-1}\mathbf{F}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}, \tag{86}$$

which is unity in the case of unconstrained estimation of $\boldsymbol{q}$. In the case of Newtonian iteration, Eqs. (85-86) are evaluated using the Jacobian $\mathbf{F}$ valid at the solution.

Due to the concentration-dependence of the source function and the q-depecdence of $\mathbf{F}$, Eq. (29) is valid only in linear approximation. This is helped by putting the inversion in the context of a Newtonian iteration (see, e.g., Rodgers (2000, p. 85). Eq. (80) becomes

$$\hat{\boldsymbol{q}}_{it+1} = \boldsymbol{q}_{it} + \left(\mathbf{F}_{it}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}_{it}\right)^{-1}\mathbf{F}_{it}^T\mathbf{S}_\mathrm{r}^{-1}(\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_{I,it}), \tag{87}$$

where it is the iteration index. Equation (83) becomes

$$\begin{aligned}\hat{\boldsymbol{q}}_{it+1} &= \boldsymbol{q}_{it} + \left(\mathbf{F}_{it}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}_{it}+\mathbf{R}\right)^{-1} \\ &\quad \left(\mathbf{F}_{it}^T\mathbf{S}_\mathrm{r}^{-1}(\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_{I,it})-\mathbf{R}(\boldsymbol{q}_{it}-\boldsymbol{q}_a)\right)\end{aligned} \tag{88}$$

or alternatively

$$\begin{aligned}\hat{\boldsymbol{q}}_{it+1} &= \boldsymbol{q}_a + \left(\mathbf{F}_{it}^T\mathbf{S}_\mathrm{r}^{-1}\mathbf{F}_{it}+\mathbf{R}\right)^{-1} \\ &\quad \mathbf{F}_{it}^T\mathbf{S}_\mathrm{r}^{-1}\left(\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_{I,it}+\mathbf{F}_{it}(\boldsymbol{q}_{it}-\boldsymbol{q}_a)\right),\end{aligned} \tag{89}$$

and Eq. (84) becomes

$$\begin{aligned}\hat{\boldsymbol{q}}_{it+1} &= \boldsymbol{q}_a + \mathbf{R}^{-1}\mathbf{F}_{it}^T\left(\mathbf{F}_{it}\mathbf{R}^{-1}\mathbf{F}_{it}^T+\mathbf{S}_{\mathrm{r},I}\right)^{-1} \\ &\quad \left(\boldsymbol{x}_{\mathrm{m};I}-\boldsymbol{x}_{I,it}+\mathbf{F}_{it}(\boldsymbol{q}_{it}-\boldsymbol{q}_a)\right).\end{aligned} \tag{90}$$

With $\boldsymbol{q}_a = \mathbf{0}$ and diagonal $\mathbf{R} = \gamma \mathbf{I}$ we get the smallest possible velocities and mixing coefficients still consistent with the measurement, where tuning parameter γ will be set depending on how large fit residuals the user still considers to be 'consistent'. With $\mathbf{R}$ being diagonally blockwise composed of squared and scaled first order finite differences operators and $\boldsymbol{q}_a = \mathbf{0}$, smooth fields of wind vectors and mixing coefficients can be enforced. Setting $\boldsymbol{q}_a$ the result of the previous macro time step ~~and~~ corresponds to sequential data assimilation. In this application $\mathbf{R}$ ~~its reciprocal uncertainty~~ is set to the reciprocal uncertainty of $\boldsymbol{q}_a$ plus some margin for allowance of variability of velocity and mixing coefficients in time~~corresponds to sequential data assimilation~~. And finally, if prior knowledge is formed by independent measurements and their reciprocal uncertainties as constraint matrix, or within the debatable framework of Bayesian statistics, estimates $\hat{\boldsymbol{q}}$ would even be the most probable estimate of velocities and mixing ratios.

5 Proof of concept

5.1 Prediction of the atmospheric state

In a first step we test the predictive power of the formalism defined by Eqs. (3–29). Since the formalism itself is deductive and starts from a well established theoretical concept, the purpose of the test is solely to verify that the implementation of the formalism is correct and that involved numerical approximations are adequate. As a consequence of the Bonini paradox (c.f. Bonini (1963) and Starbuck (1975)), a model is the harder to understand the more complex it is. While the predictive power of a model usually increases with complexity, this does not necessarily hold for its explanatory power. Thus we have decided to test our model on the basis of very simple test cases, where major failure of the model is immediately obvious. Four test cases have been chosen, each dedicated to one kinematic variable (v, w, K_ϕ and K_z), while the other three were set to zero.

In the first case, v was set ~~close to~~ to a tenth of the Courant limit (Courant et al., 1928) (about ~~$0.17\cos(\phi)$~~ $0.17/\cos(\phi)$ ms^{-1}) everywhere. As one would expect from the continuity equation applied to a spherical atmosphere, no changes in air density except boundary effects at the poles were observed, and structures near the equator were transported by about 4° within a month, as expected from the equation of motion. A Gaussian-shaped perturbation of a halfwidth of one latitudinal gridwidth (4°) causes an upwind wiggle of less than 0.7% of the amplitude of the perturbation at a meridional velocity of one gridpoint per month. There is no discernable change in the width of the transported structure. Similarly, for the second case a constant field of w of 1.1×10^{-3} ms^{-1} lifts a structure upwards by about 3 km per month. Mixing coefficients were verified to smear out structures in the respective direction while leaving air density and structures in the orthogonal direction unchanged.

5.2 Inversion of simulated measurements

Case studies based on real measurement data are inadequate as the sole proof of concept because the truth is unknown and the result thus is unverifiable. Instead we first test our scheme on the basis of simulated atmospheric states and consider the scheme as verified if the velocities and mixing coefficients used to simulate the atmospheric states are sufficiently well reproduced. In the noise-free well conditioned case one might even expect, within the numerical precision of the system, the exact reproduction of the reference data; due to the – weak but non-zero – dispersivity of the numerical transport scheme the wiggles discussed in the previous subsection cause, at some gridpoints, $\mathbf{D}$-matrix entries of the wrong sign. In order to fight resulting convergence problems of the inversion, at least some small regularization is adequate, even if the system of equations to be solved is well or over-determined. Since the system in reality is, in tendency, ill-conditioned and the constraint applied to the inversion prevents reproduction of the reference data, we use a variety of idealized tracers instead. After this initial test of functionality, more and more realistic test cases are constructed in order to study the competing influence of constraint and measurement data on the solution.

~~The trace gas distributions used for this test were chosen such that~~During code development, a series of basic tests of increasing complexity were performed, including a variety of mixing ratio distributions transported with various velocity fields. The main lesson learned was that, even when the rows of Equation (80) ~~are independent. Four artificial gases were chosen whose mixing ratios had a linear, quadratic and two variants of exponential latitude and altitude dependences, respectively. Further, a hydrostatic air density distribution was assumed. After successful separate inversion of v_{phi}, v_z, K_{phi} and K_z, these kinematic variables were inverted in combination, whereby no further problems were encountered.~~

~~In a following step, the linear latitude-dependence was replaced by a stepwise linear function, i.e., a dependence on the absolute amount of latitude. Here it showed up that, besides~~ were independent and no formal ill-posedness ~~due to linear dependence of equations~~could be diagnosed, the unphysical upwind wiggles in the vicinity of the mixing ratio peak as discussed in the previous subsection ~~can~~ could trigger errors which are boosted during the iteration. This problem, which is associated with sharp structures and large velocities (of the order of one gridwidth per ~~macrotimestep~~macro timestep) can be solved by the use of a smoothing regularization matrix $\mathbf{R}$ as discussed in the last paragraph of Section 4, however at the cost of degraded spatial resolution of the result.

As an example we show the following test. An altitude-independent meridional velocity field (in degrees per month)

$$v = \frac{23.6562}{2} \cos\phi \qquad (91)$$

was chosen while the vertical velocity was set zero (Fig. 2a). This particular choice of the meridional velocity field keeps boundary problems at the poles due to divergence in a circulation which is not closed reasonably small. Initial mixing ratio distributions (in ppmv) of four idealized gases were constructed such that gas (a) was sugar-loaf shaped while those of the gases (b) to (d) were monotonical slopes.

$$vmr_a = \left(e^{-\left(\frac{z-36}{1.5}\right)^2/72} + 5.0*10^{-4}\right) e^{-\left(\frac{6(\phi+6)}{40}\right)^2/12} \qquad (92)$$

$$vmr_b = \frac{(30+0.02z^2-0.2\phi)^3}{1000} \qquad (93)$$

$$vmr_c = 20 * e^{0.02z+0.03\phi} \qquad (94)$$

and

$$vmr_d = 30 * e^{-0.03z-0.002\phi}, \qquad (95)$$

where z is altitude above surface in km and phi is latitude in degrees. The distribution of gas (a) is shown Fig. 2b. The transport problem was solved in the forward mode using the velocity field as defined in Eq. 91 for integrating the tendency equation (eq. 28) over one month. The resulting distribution of gas a after this macro-timestep is shown Fig. 2c. The "sugarloaf" is transported to the North by the expected distance. Due to the latitudinal gradients with lower velocities in the luff of the sugar-loaf and larger velocities in the lee, the shape of the "sugar-loaf" is slightly distorted, leading to a steeper slope in the luff and a flatter slope in the lee. No indication of problems with diffusity or dispersivity of the transport scheme is seen. At the poles boundary problems are visible which are inavoidable when such an unrealistic (but instructive) velocity field is used. The inversion nicely reproduces the initial velocity field (not shown because hardly discernable from Fig. 2b). Due to the smoothing constraint the peak velocity is decreased by 18%. This smoothing is the price to pay for the stabilization of the retrieval in the presence of the boundary problems discussed above. The residual of between the simulated measurement of the distribution of gas (a) at $t_0 + \Delta t$ and the distribution predicted with the resulting velocity field is shown in (Fig 2d).

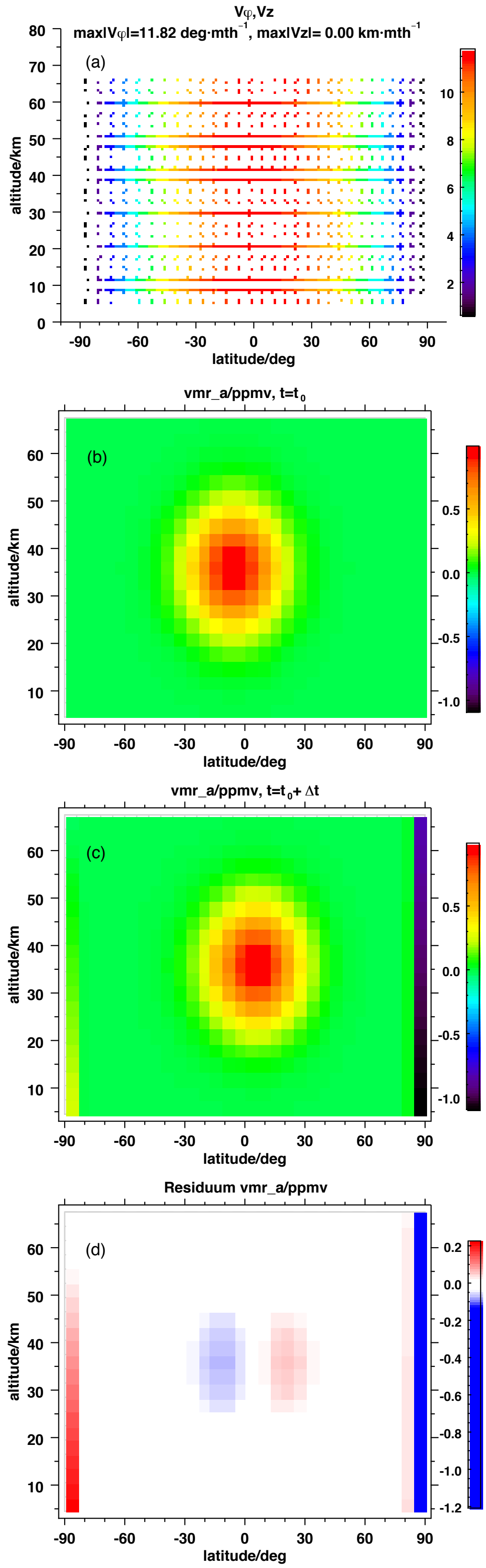

Figure 2. : Test case. (a) velocity field; (b) initial distribution of gas a; (c) distribution of gas a after one month; (d) residual between distribution of gas a after one month predicted with the resulting and the correct velocities.

5.3 Case Study with MIPAS measurements

The risk of case studies based on simulated data typically is that not all difficulties encountered with real data are foreseen during theoretical studies. In order to demonstrate applicability to real data, global monthly latitude/altitude distributions of CFC-12, CH_4, N_2O and SF_6 (Kellmann et al., 2012; Plieninger et al., 2015; Haenel et al., 2015) measured with the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) (Fischer et al., 2008) were used. The purpose of these tests is demonstration of the feasibility of the method presented. An investigation of the atmospheric circulation on the basis of this method applied to MIPAS data is left for a companion paper. For this proof of concept, sinks of these long-lived tracers have been ignored but these will certainly be considered in scientific applications.

For this case study, zonal monthly mean distributions of air densities and mixing ratios of these four species from September and October 2010 were used. Figure 3 shows the measured distributions of these quantities in September (left column) and October (middle column), and the residuals between the measured and predicted contributions for October (right column). The residuals are reasonably small and show, except for methane in the polar upper stratosphere, no patterns which would hint at peculiarities with the inferred kinematic quantities.

The resulting circulation vectors which best explain the change of the mixing ratio distributions from September to October 2010 are shown in the upper left panel of Figure 4. Winter polar subsidence, summer polar upwelling, the mesospheric overturning circulation, the upper and lower branches of the Brewer-Dobson circulation and the tropical pipe are clearly visible. Details of the tropical pipe are visible in the right upper panel. As expected, the Brewer-Dobson circulation is much more pronounced in the northern (early) winter hemisphere. Velocities are roughly consistent with mean ages of stratospheric air as determined by Stiller et al. (2008) and Haenel et al. (2015) in a sense that the quotient of the typical circulation velocity and the distance between equator to pole gives an age estimate of the correct order of magnitude. While the inferred field of circulation vectors shows many detail features demanding scientific investigation in their own right ~~, the~~ (e.g., the latitude offset between the intertropical convergence zone and the stratospheric tropical pipe, or the interfacing between the stratospheric two-cell circulation and the overturning mesospheric circulation and the transition altitude between them), the reproduction of the expected features justifies confidence in the method proposed. Resulting mixing coefficients K_ϕ and K_z are shown in the left and right lower panels, respectively. Negative mixing ~~ratios~~ coefficients indicate counter-gradient mixing, which seems to be most pronounced in the tropical upper stratosphere.

Jacobian elements with respect to v values and K values seem to form a null space. Thus the K-values were constrained to zero using diagonal components only in the $\mathbf{R}$ matrix diagonal blocks associated with the K values. The strength of this constraint was adjusted such that the K values were as small as possible as long as this did not boost the residual. Resulting K_ϕ and K_z distributions are shown in Figure ~~5~~4.

The errors in the estimated transport velocities and mixing coefficients have been estimated according to Equation (85) and are shown in Figure 5~~, middle column. The errors~~. The uncertainties in the transport velocities caused by the propagation of measurement errors are in the one percent range, indicating that the information contained in the measurements is adequate for the purpose of retrieving circulation parameters. It seems even possible to improve the time resolution of the circulation analysis and aim at ~~weakly~~ weekly instead of monthly temporal sampling.

Larger errors above 65 km altitude and at the bins closest to the pole are border effects, resulting from the fact that no symmetric derivatives can be calculated there. The uncertainties in K_ϕ show the same patterns as the K_ϕ values themselves.

6 Discussion

The analysis of the age of stratospheric air can be understood as an integrated view at the equations of motion of stratospheric air, because the total travel time of the air parcel through the stratosphere is represented. The refinement of this method which analyzes the mean age just considers a weighted mean of the above, but it is still an integral method. Contrary to these integral methods, our direct inversion scheme supports a – in approximation, due to discrete sampling in the time domain – differential view ~~at~~ of the same problem. The related advantages are: (a) independence of assumptions on the age spectrum, because during each time step mixing is explicitly considered; (b) insensitivity to SF_6 depletion in the mesosphere (c.f., e.g., Reddmann et al., 2001; Stiller et al., 2012), because the scheme uses the actual entry values of subsiding air as a reference; (c) applicability to non-ideal tracers in the stratosphere; since the atmospheric state is updated for each time step by measured value, depletion does not accumulate, even if no sink functions are considered; and (d) the logical circle that the lifetimes of non-ideal tracers depend on their trajectories (and thus atmospheric circulation), while the determination of the circulation requires knowledge of the lifetimes, can be solved. Our scheme requires knowledge only on the local, not the global, lifetimes; (e) the method is an empirical method which does not involve any dynamical model, i.e. the forces which cause the circulation are not required. The method only finds that kinematic state of the atmosphere which, according to the continuity equation, fits best to the measurements. These kinematic state values are provided as model diagnostics to assess the performance of dynamical models.

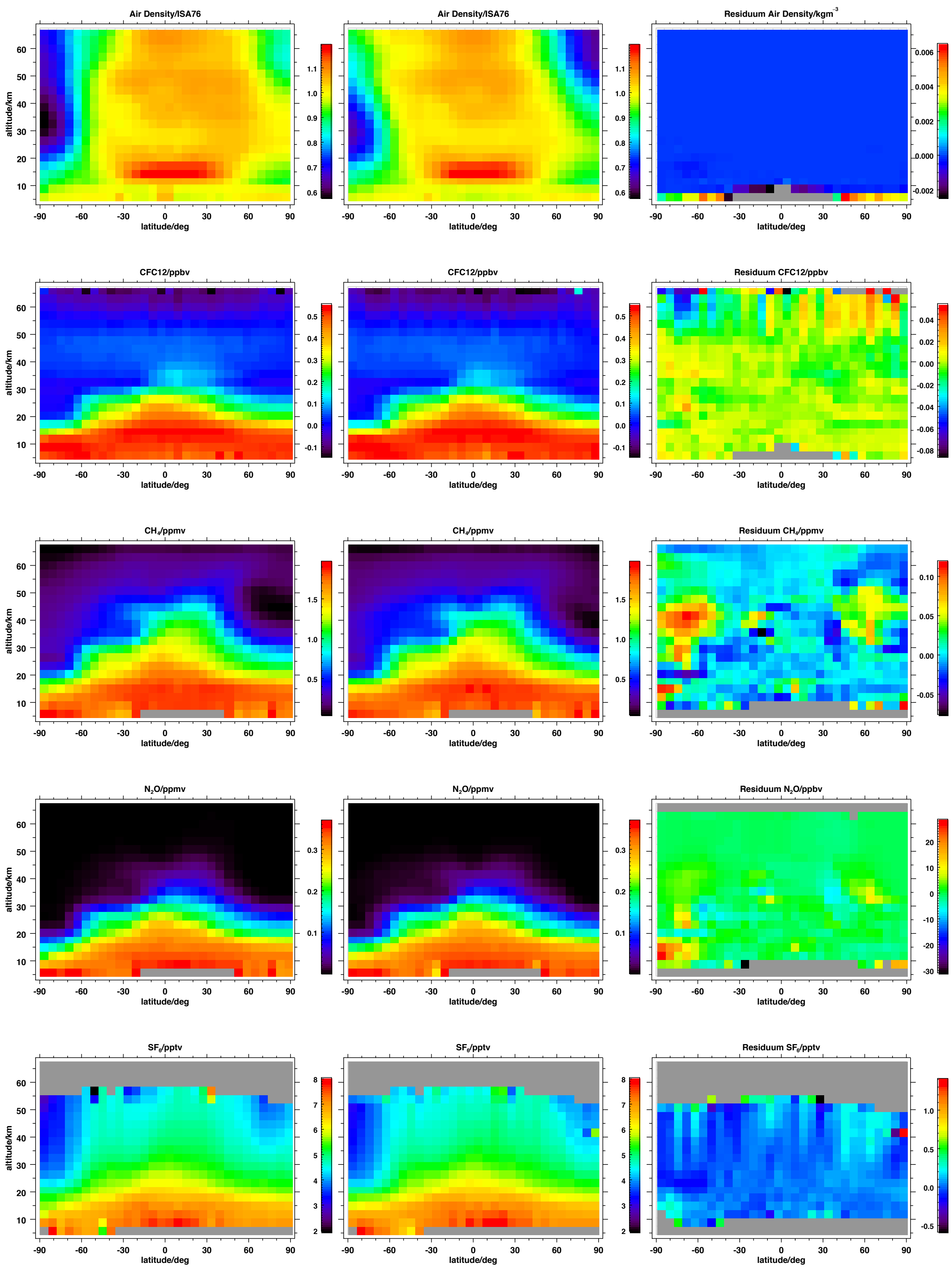

Figure 3. : Measured distributions in September (left column), October (middle column) and residual distributions between October measurements and predictions for October (right column) for air density and mixing ratios of CFC-12, CH_4, N_2O, and SF_6 (top to bottom). Grey gridboxes indicate non-availability of valid data.

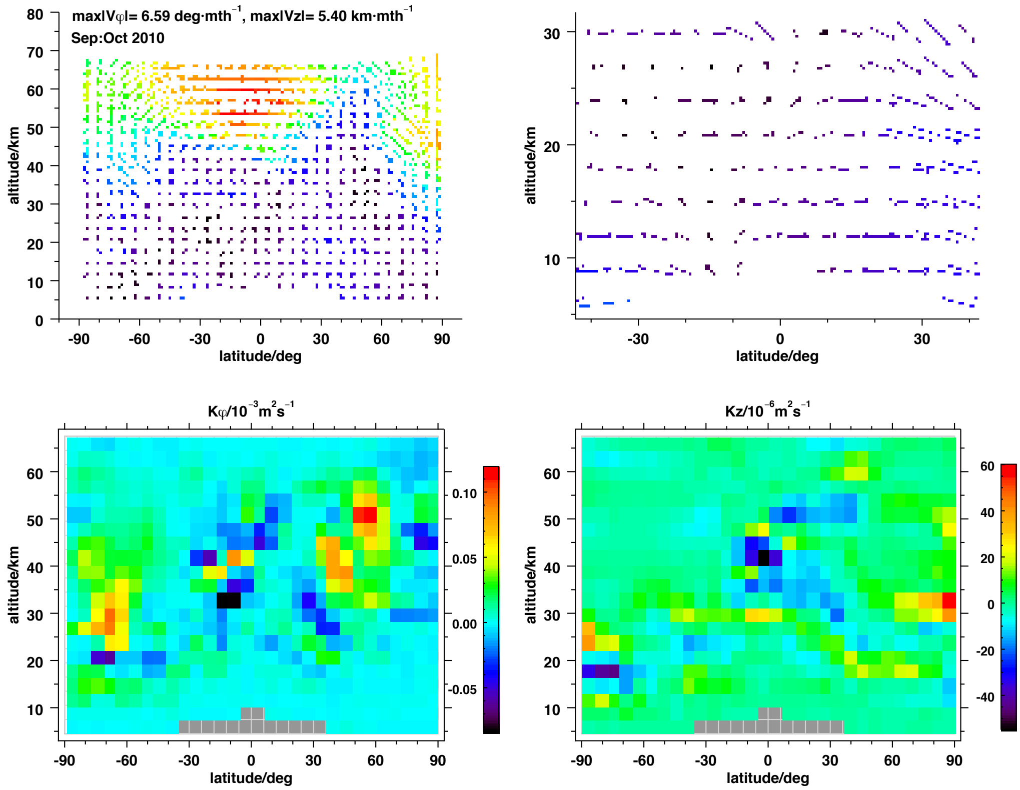

Figure 4. : Resulting circulation vectors (~~$v_\phi(z,\phi); v_z(z,\phi$~~$v(z,\phi); w(z,\phi)$) (upper left panel), where colours on the red side of the rainbow colour scale represent higher velocities; a ~~detail~~ zoomed-in view of this (upper right panel); mixing coefficients ~~K_{phi}~~ K_ϕ (lower left panel) and K_z (lower right panel).

Due to these advantages, the major problems in the empirical analysis of the Brewer-Dobson circulation as mentioned by Butchart (2014) are solved. Problems related to our method are (a) sensitivity of the inferred kinematic quantities to locally varying biases, (b) a tendency towards ill-posedness of the inversion if distributions of too few tracers with too similar morphology are used, and (c) the usual artefacts arising if the numerical discretization is chosen too coarse. Results of the case study presented in Section 5.3 suggest that these problems ~~have successfully been solved~~ are under control in the current application of the proposed scheme.

7 Conclusions and outlook

We have presented a method which infers mixing coefficients and effective velocities of a 2-D atmosphere by inversion of the continuity equation. The main steps of this procedure are (a) integration of the continuity equation over time to predict pressure and mixing ratios for given initial pressures and mixing ratios and initially guessed velocities and mixing coefficients; (b) propagation of errors of initial pressures and mixing ratios onto the predicted pressures and mixing ratios, by differentiation of the predicted state with respect to the initial state and generalized Gaussian error propagation; (c) estimation of the sensitivities of the predicted state with respect to the velocities and mixing coefficients; and (d) minimization of a quadratic cost function involving the residual between measured and predicted state at the end of the forecasting interval by inversion of the continuity euqation. The inferred velocities are suggested to be used as a model diagnostic in order to avoid problems encountered with other model diagnostics like mean age of stratospheric air. It is important to note that the diagnostics inferred here are effective transport velocities and effective mixing coefficients in a sense that they include eddy transport and diffusion terms. Thus, they cannot simply be compared to zonal mean velocities and mixing coefficients of a 3D model but the eddy terms have to be considered when these diagnostics quanities are calculated. The application of this method on SF_6 distributions measured by MIPAS (Stiller et al., 2012) to diagnose the Brewer-Dobson circulation are discussed in a compan-

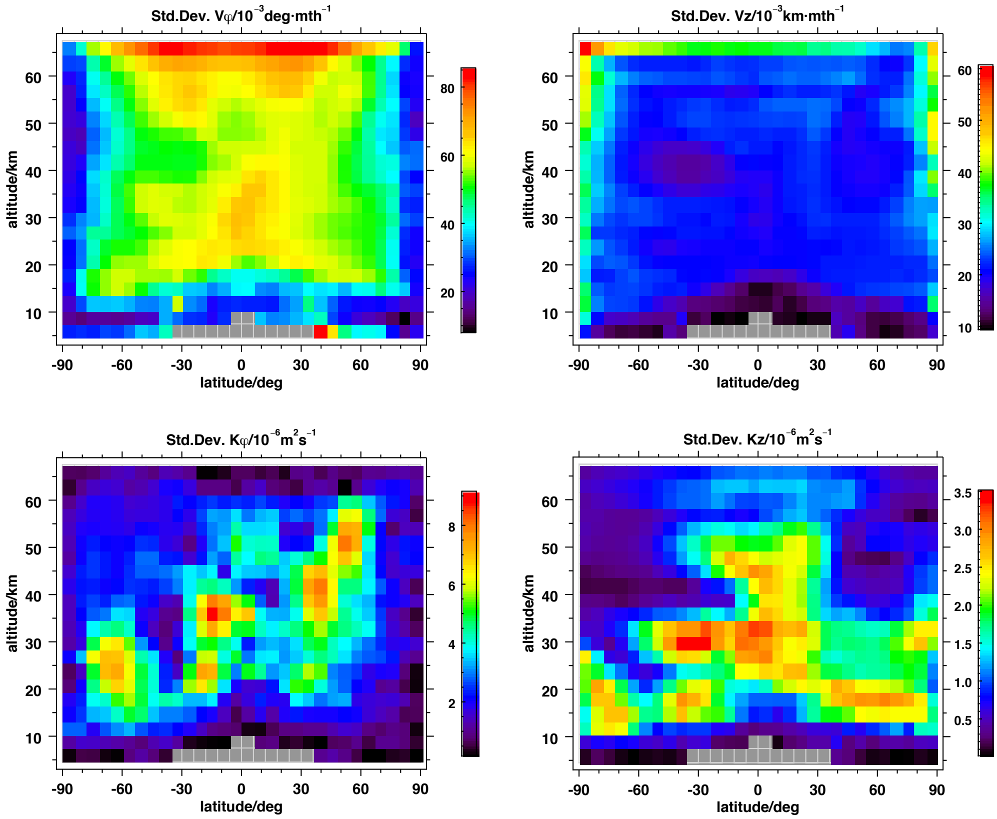

Figure 5. : Estimated uncertainties of ~~$v_\phi(z,\phi)$~~$v(z,\phi)$, (upper left panel), ~~$v_z(z,\phi$~~$w(z,\phi)$ (upper right panel), ~~K_{phi}~~ K_ϕ (lower left panel) and K_z (lower right panel).

ion paper. Obvious future activities are the extension of this method to three dimensions and inclusion of sink functions of non-inert species to explore a larger number of tracers in order to better constrain the related inverse problem.

Appendix A: From 3D to 2D

The inference of effective two-dimensional transport velocities and effective mixing coefficients from measurements discussed in the main paper relies on the fact that ~~under usual conditions (species chemical lifetimes large compared to transport lifetimes, slow change of mean state compared to the effect of eddies)~~ , within certain assumptions and approximations, all eddy effects ~~can be parametrized by transport and mixing terms . The additional~~ emerging from Reynolds decomposition of the three- dimensional continuity equation and subsequent zonal averaging can be understood as additional pseudo-advection and pseudo-mixing terms according to the advection equation and the Fickian law, with pseudo-velocities and pseudo-mixing coefficients which are gas-independent. The exact interpretation of the two-dimensional velocities and mixing coefficients inferred from the measurements depends on the approximations made. We apply our scheme to zonally averaged mixing ratios (no mass-weighted averaging!). Contrary to the main text, where the symbols v, w, K_ϕ, and K_z are used in the 2-dimensional system, here zonal averages are indicated by a bar.

UNFORTUNATELY LATEXDIFF DOES NOT WORK PROPERLY, THUS THE DIFFERENCES CANNOT BE SHOWN; THE ENTIRE REMAINING PART OF THE OLD APPENDIX HAS BEEN DELETED. NOT SHOWN HERE.

THE FOLLOWING NEW APPENDIX HAS BEEN INCLUDED:

Assuming

- that the deviations from the zonal mean are small compared to the zonal mean itself such that linearization is justifiable,
- that meridional advection is negligibly small compared to zonal advection,

– that the time variation of the zonal mean quantities is assumed to be much slower than the time variation of the deviations from the zonal mean, which corresponds to the assumption of a quasi-steady state,

Tung (1982) derives a two-dimensional approximation to the continuity equation which, adjusted to our notation, written for geometric altitudes instead of potential temperatures, and using the shallow water approximation, reads

$$\bar{\rho}\frac{\partial}{\partial t}\overline{vmr_g} + \left(\overline{\rho v} - \frac{\partial}{\partial t}\overline{\rho'\eta'}\right)\frac{1}{r}\frac{\partial}{\partial \phi}\overline{vmr_g} + \left(\overline{\rho w} - \frac{\partial}{\partial t}\overline{\rho'\Phi'}\right)\frac{\partial}{\partial z}\overline{vmr_g} - \frac{1}{r}\frac{\partial}{\partial \phi}\left(\bar{\rho}K_{\phi\phi}\frac{1}{r}\frac{\partial}{\partial \phi}\overline{vmr_g}\right) - \frac{1}{r}\frac{\partial}{\partial \phi}\left(\bar{\rho}K_{\phi z}\frac{\partial}{\partial z}\overline{vmr_g}\right) - \frac{\partial}{\partial z}\left(\bar{\rho}K_{z\phi}\frac{1}{r}\frac{\partial}{\partial \phi}\overline{vmr_g}\right) - \frac{\partial}{\partial z}\left(\bar{\rho}K_{zz}\frac{\partial}{\partial z}\overline{vmr_g}\right) = \bar{S} - \frac{1}{r}\frac{\partial}{\partial \phi}(\overline{(\rho v)'\sigma'}) - \frac{\partial}{\partial z}(\overline{(\rho z)'\sigma'}) - \frac{\partial}{\partial t}(\overline{\rho'\sigma'}). \quad \text{(A1)}$$

where η', Φ' and σ' are defined by

$$\left(\frac{\partial}{\partial t} + \frac{\bar{u}}{r\cos\phi}\frac{\partial}{\partial \lambda}\right)\eta' = v', \quad \text{(A2)}$$

$$\left(\frac{\partial}{\partial t} + \frac{\bar{u}}{r\cos\phi}\frac{\partial}{\partial \lambda}\right)\Phi' = w' \quad \text{(A3)}$$

and

$$\left(\frac{\partial}{\partial t} + \frac{\bar{u}}{r\cos\phi}\frac{\partial}{\partial \lambda}\right)\sigma' = S'. \quad \text{(A4)}$$

Further,

$$K_{\phi\phi} = \frac{1}{\bar{\rho}}\overline{(\rho v)'\eta'} \quad \text{(A5)}$$

$$K_{zz} = \frac{1}{\bar{\rho}}\overline{(\rho w)'\Phi'} \quad \text{(A6)}$$

$$K_{\phi z} = \frac{1}{\bar{\rho}}\overline{(\rho v)'\Phi'} \quad \text{(A7)}$$

and

$$K_{z\phi} = \frac{1}{\bar{\rho}}\overline{(\rho w)'\eta'} \quad \text{(A8)}$$

Assuming further that

– our scheme is applied only to long-lived species, such that chemical eddy terms can be ignored because chemical lifetimes are long compared to transport lifetimes (c.f. Pyle and Rogers (1980)),

– wave disturbances are dominated by steady or periodic terms, such that the terms with the mixed second derivative terms tend to disappear (Matsuno (1980); Clark and Rogers (1978); Pyle and Rogers (1980), quoted after Tung (1982)). Here an even weaker approximation is sufficient, namely

$$\frac{\partial}{\partial t}\overline{\eta'\Phi'} \approx 0. \quad \text{(A9)}$$

This is equivalent with the assumption that $K_{z\phi} \approx -K_{\phi z}$

Eq. A1 can be rewritten as a tendency equation of the type

$$\frac{\partial}{\partial t}\overline{vmr_g} = -\frac{v^*}{r}\frac{\partial}{\partial \phi}\overline{vmr_g} - w^*\frac{\partial}{\partial z}\overline{vmr_g} + \frac{1}{r^2}\frac{\partial}{\partial \phi}\left[K^*_{\phi}\frac{\partial}{\partial \phi}\overline{vmr_g}\right] + \frac{\partial}{\partial z}\left[K^*_{z}\frac{\partial}{\partial z}\overline{vmr_g}\right] \quad \text{(A10)}$$

where

$$v^* = \bar{v} - \frac{1}{\bar{\rho}}\frac{\partial}{\partial t}\overline{\rho'\eta'} - \frac{1}{\bar{\rho}r}\frac{\partial\bar{\rho}}{\partial \phi}K_{\phi\phi} - \frac{\partial}{\partial z}K_{z\phi} - \frac{1}{\bar{\rho}}\frac{\partial\bar{\rho}}{\partial z}K_{z\phi}, \quad \text{(A11)}$$

$$w^* = \bar{w} - \frac{1}{\bar{\rho}}\frac{\partial}{\partial t}\overline{\rho'\Phi'} - \frac{1}{\bar{\rho}}\frac{\partial\bar{\rho}}{\partial z}K_{zz} - \frac{1}{r}\frac{\partial}{\partial \phi}K_{\phi z} - \frac{1}{\bar{\rho}r}\frac{\partial\bar{\rho}}{\partial \phi}K_{\phi z}, \quad \text{(A12)}$$

$$K^*_{\phi} = K_{\phi\phi} \quad \text{(A13)}$$

and

$$K^*_{z} = K_{zz} \quad \text{(A14)}$$

can be understood as virtual velocities and virtual mixing coefficients. All terms in Eq. (A10) have either the mathematical structure of an advection equation or a Fickian law. Thus, Eqs.(A11 - A14) provide the interpretation of the velocities and mixing coefficients inferred in the main part of the paper. These can be identified with the virtual velocities and virtual mixing coefficients inferred above. Obviously, this interpretation depends on the approximations made, and different approximations would lead to a different interpretation. If mixing coefficients describing subscale effects in the original 3-D model (the last two terms in Eq. (3)) are to be considered, then the effective mixing coefficient is the sum of the mixing coefficients describing these subscale effects plus the respective K^* term accounting for the eddy mixing.

We like to emphasize that none of the approximations and assumptions discussed above are used in our proposed method to infer velocities from zonal mean mixing ratio measurements. The discussion in this Appendix only tries to relate the resulting velocities to the velocities in a 3-D world. The ambiguities in the interpretation of the inferred 'effective' 2-D velocities suggests that it might be promising to switch from a theoretical to an empirist view and to conceive no longer the zonal mean of the 3-D velocities as the 'true' 2-D velocities, but those which satisfy the 2-D continuity equation. These can be – admittedly indirectly – observed with our suggested method and can be used to validate 2-D models, including their underlying concept of solving the 2-D transport problem. With this, also the adequacy of the assumptions made to approximate away the headache terms which can be expressed neither as advection nor as Fickian law terms can be tested by means of comparison of the measured and 2-D modelled effective, i.e. transport relevant, velocities. This empiristic turn in argumentation might not fully solve all aspects of the problem of interpretation of the observed 2-D velocities from a 3-D perspective, but at least it moves the problem from the desk of the observation scientist onto the desk of the 2-D modeller.

Acknowledgements. The authors thank two anonymous reviewers for their thorough examination of the original manuscript and for helpful and important comments. Furthermore, TvC wishes to thank Hendrik Elbern, Richard Menard, Peter Braesicke, Björn-Martin Sinnhuber and Thomas Birner for drawing his attention to some important literature and for encouragement as well as Arne Babenhauserheide for helpful discussions.

References

Bonini, C. P.: Simulation of information and decision systems in the firm, Prentice-Hall, Englewood Cliffs, N. J., 1963.

Brasseur, G. and Solomon, S.: Aeronomy of the Middle Atmosphere–Chemistry and Physics of the Stratosphere and Mesosphere, Atmospheric and Oceanographic Sciences Library 32, Springer, P. O. Box 17, 3300 AA Dordrecht, The Netherlands, third edn., 2005.

Butchart, N.: The Brewer-Dobson Circulation, Rev. Geophys., 52, 157–184, doi:10.1002/2013RG000448, 2014.

Butchart, N., Scaife, A. A., Bourqui, M., de Grandpre, J., Hare, S. H. E., Kettleborough, J., Langematz, U., Manzini, E., Sassi, F., Shibata, K., Shindell, D., and Sigmond, M.: Simulations of anthropogenic change in the strength of the Brewer-Dobson circulation, Clim. Dyn., 27, 727–741, doi:10.1007/s00382-006-0162-4, 2006.

Chang, C. A. and St.-Maurice, J.-P.: Two-dimensional high-latitude thermospheric modeling: A comparison between moderate and extremely disturbed conditions, Can. J. Phys., 69, 1007–1031, doi:10.1139/p91-159, 1991.

Clark, J. H. E. and Rogers, T. G.: The transport of trace gases by planetary waves, J. Atmos. Sci., 35, 2232–2235, 1978.

Courant, R., Friedrichs, K., and Lewy, H.: Über die partiellen Differentialgleichungen der mathematischen Physik, Mathematische Annalen, 100, 32–74, doi:10.1007/BF01448839, 1928.

Courant, R., Isaacson, E., and Rees, M.: On the Solution of Nonlinear Hyperbolic Differential Equations by Finite Differences, Comm. Pure Appl. Math., 5, 243–255, doi:10.1002/cpa.3160050303, 1952.

Engel, A., Möbius, T., Bönisch, H., Schmidt, U., Heinz, R., Levin, I., Atlas, E., Aoki, S., Nakazawa, T., Sugawara, S., Moore, F., Hurst, D., Elkins, J., Schauffler, S., Andrews, A., and Boering, K.: Age of stratospheric air unchanged within uncertainties over the past 30 years, Nature Geosci., 2, 28–31, doi:10.1038/ngeo388, 2009.

Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopmann, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R.: MIPAS: an instrument for atmospheric and climate research, Atmos. Chem. Phys., 8, 2151–2188, 2008.

Fueglistaler, S., Dessler, A. E., Dunkerton, T. J., Folkins, I., Fu, Q., and Mote, P. W.: Tropical Tropopause Layer, Rev. Geophys., 47, RG1004, doi:10.1029/2008RG000267, 2009.

Garcia, R. R., Randel, W. J., and Kinnison, D. E.: On the determination of age of air trends from atmospheric trace species, J. Atmos. Sci., 68, 139–154, 2011.

Ghil, M.: Advances in sequential estimation for atmospheric and oceanic flows, J. Meteorol. Soc. Jpn., 75, 289–304, 1997.

Ghil, M. and Malanotte-Rizzoli, P.: Data assimilation in meteorology and oceanography, Adv. Geophys., 33, 141–266, 1991.

Haenel, F. J., Stiller, G. P., von Clarmann, T., Funke, B., Eckert, E., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., and Reddmann, T.: Reassessment of MIPAS age of air trends and variability, Atmos. Chem. Phys., 15, 13 161–13 176, doi:10.5194/acp-15-13161-2015, 2015.

Hall, T. M. and Plumb, R. A.: Age as a diagnostic of stratospheric transport,, J. Geophys. Res., 99, 1059–1070, 1994.

Hall, T. M. and Waugh, D. W.: Influence of nonlocal chemistry on tracer distributions: Inferring the mean age of air from SF_6, J. Geophys. Res., 103, 13,327–13,336, doi:10.1029/98JD00170, 1998.

Hinkelmann, K.: Primitive Equations, in: Lectures on Numerical Shortrange Weather Prediction, vol. 297 of *Regional Training Seminar*, pp. 306–375, WMO, 1951.

Holton, J. R. and Choi, W.-K.: Transport Circulation Deduced from SAMS Trace Species Data, J. Atmos. Sci., 45, 1929–1939, 1988.

Jones, A. R., Thomson, D. J., Hort, M., and Devenish, B.: The U.K. Met Office's next-generation atmospheric dispersion model, NAME III, in: Proceedings of the 27th NATO/CCMS International Technical Meeting on Air Pollution Modelling and its Application, edited by Borrego, C. and Norman, A.-L., vol. XVII of *Air Pollution Modeling and its Application*, pp. 580–589, Springer, 2007.

Kasahara, A.: Computational Aspects of Numerical Models for Weather Prediction and Climate Simulation, in: General Circulation Models of the Atmosphere, edited by Chang, J., pp. 2–108, Academic Press, New York, 1977.

Kellmann, S., von Clarmann, T., Stiller, G. P., Eckert, E., Glatthor, N., Höpfner, M., Kiefer, M., Orphal, J., Funke, B., Grabowski, U., Linden, A., Dutton, G. S., and Elkins, J. W.: Global CFC-11 (CCl_3F) and CFC-12 (CCl_2F_2) Measurements with the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS):

retrieval, climatologies and trends, Atmos. Chem. Phys., 12, 11 857–11 875, doi:10.5194/acp-12-11857-2012, 2012.

Ko, M. K. W., Tung, K. K., Weisenstein, D. K., and Sze, N. D.: A Zonal Mean Model of Stratospheric Tracer Transport in Isentropic Coordinates: Numerical Simulations for Nitrous Oxide and Nitric Acid, J. Geophys. Res., 90, 2313–2329, 1985.

Krueger, A., Stremme, W., Harig, R., and Grutter, M.: Volcanic SO_2 and SiF_4 visualization using 2-D thermal emission spectroscopy - Part 2: Wind propagation and emission rates, Atmos. Meas. Tech., 6, 47–61, doi:10.5194/amt-6-47-2013, 2013.

MacCormack, R. W.: The effect of viscosity in hypervelocity impact cratering, in: AIAA Hypervelocity Impact Conference, Cincinnati, Chio/April 30 -May 2, 1969, pp. 69–354, AIAA, American Institute of Aeronautics and Astrophysics, 1969.

Matsuno, T.: Lagrangian motion of air parcels in the stratosphere in the presence of planetary waves, Pure Appl. Geophys., 118, 189–216, 1980.

Perrin, A. and Hu, H. H.: An explicit finite-difference scheme for simulation of moving particles, J. Comput. Phys., 212, 166–187, doi:10.1016/j.jcp.2005.06.021, 2006.

Phillips, N. A.: The equations of motion for a shallow rotating atmosphere and the ‘traditional approximation’, J. Atmos. Sci., 23, 626–628, 1966.

Plieninger, J., von Clarmann, T., Stiller, G. P., Grabowski, U., Glatthor, N., Kellmann, S., Linden, A., Haenel, F., Kiefer, M., Höpfner, M., Laeng, A., and Lossow, S.: Methane and nitrous oxide retrievals from MIPAS-ENVISAT, Atmos. Meas. Tech., 8, 4657–4670, doi:10.5194/amt-8-4657-2015, 2015.

Pyle, J. A. and Rogers, C. F.: A modified diabatic circulation model for stratospheric tracer transport, Nature, 287, 711–714, doi:10.1038/287711a0, 1980.

Reddmann, T., Ruhnke, R., and Kouker, W.: Three–dimensional model simulations of SF_6 with mesospheric chemistry, J. Geophys. Res., 106, 14,525–14,537, doi:10.1029/2000JD900700, 2001.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, vol. 2 of *Series on Atmospheric, Oceanic and Planetary Physics, F. W. Taylor, ed.*, World Scientific, Singapore, New Jersey, London, Hong Kong, 2000.

Salby, M. L. and Juckes, M. N.: An Algorithm for Retrieving Atmospheric Motion From Satellite Measurements of Tracer Behavior, J. Geophys. Res., 99, 1403–1417, 1994.

Schmidt, U. and Khedim, A.: In situ measurements of carbon dioxide in the winter Arctic vortex and at midlatitudes: An indicator of the ”age” of stratospheric air, Geophys. Res. Lett., 18, 763–766, 1991.

Solomon, S.: Antarctic ozone: Progress towards a quantitative understanding, Nature, 347, 347–354, 1990.

Starbuck, W. H.: Organizations and their environments, in: Handbook of industrial and organizational psychology, edited by Dunnette, M. D., pp. 1069–1123, Rand, Chicago, 1975.

Stiller, G. P., v. Clarmann, T., Höpfner, M., Glatthor, N., Grabowski, U., Kellmann, S., Kleinert, A., Linden, A., Milz, M., Reddmann, T., Steck, T., Fischer, H., Funke, B., López-Puertas, M., and Engel, A.: Global distribution of mean age of stratospheric air from MIPAS SF_6 measurements, Atmos. Chem. Phys., 8, 677–695, 2008.

Stiller, G. P., von Clarmann, T., Haenel, F., Funke, B., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Lossow, S., and López-Puertas, M.: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, Atmos. Chem. Phys., 12, 3311–3331, doi:10.5194/acp-12-3311-2012, 2012.

Thompson, P.: Reduction of analysis error through constraints of dynamical consistency, J. Appl. Meteor., 8, 739–742, 1969.

Tung, K. K.: On the Two-Dimensional Transport of Stratspheric Trace Gases in Isentropic Coordinates, J. Atmos. Sci., 39, 2330–2355, 1982.

von Clarmann, T., Höpfner, M., Kellmann, S., Linden, A., Chauhan, S., Funke, B., Grabowski, U., Glatthor, N., Kiefer, M., Schieferdecker, T., Stiller, G. P., and Versick, S.: Retrieval of temperature, H_2O, O_3, HNO_3, CH_4, N_2O, $ClONO_2$ and ClO from MIPAS reduced resolution nominal mode limb emission measurements, Atmos. Meas. Techn., 2, 159–175, 2009.

Waugh, D. W. and Hall, T. M.: Age of stratospheric air: theory, observations, and models, Rev. Geophys., 40, 1010, doi:10.1029/2000RG000101, 2002.

Wofsy, S. C., Boering, K. A., Daube, Jr., B. C., McElroy, M. B., Loewenstein, M., Podolske, J. R., Elkins, J. W., Dutton, G. S., and Fahey, D. W.: Vertical transport rates in the stratosphere in 1993 from observations of CO_2, N_2O and CH_4., Geophys. Res. Lett., 21, 2571–2574, 1994.

Wunsch, C.: The ocean circulation inverse problem, Cambridge University Press, 1996.