**Atmospheric
Chemistry
and Physics**

Evaluation of MIPAS ozone fields assimilated using a new algorithm constrained by isentropic tracer advection

M. N. Juckes

British Atmospheric Data Centre, SSTD, Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire, OX11 0QX, UK

Abstract. A new data assimilation algorithm, using the isentropic advection equation, is applied to MIPAS and SBUV measurements of stratospheric ozone. The system is solved separately on each isentropic level, with neither vertical advection nor chemical reactions represented. The results are validated against HALOE, POAM III, SAGE II & III, OSIRIS and ozone sonde data. The new assimilation algorithm has the accuracy of the Kalman smoother but is, for the systems studied here with up to 200 000 variables per time step and 61 million control variables in total, many orders of magnitude less computationally expensive. The analysis produced minimises a single penalty function evaluated over an analysis window of over one month. The cost of the analysis is found to increase nearly linearly with the number of control variables. Compared with over 800 profiles from Electrochemical Concentration Cell sondes at 29 sites the analysis is found to be merely 0.1% high at 420 K, rising to 0.4% at 650 K. Comparison against the other satellites imply that the bias remains small up to 1250 K (38 km) and then increases to around –10% at 1650 K (44 km). Between 20 and 35 km the root-mean-square difference relative to HALOE, SAGE II & III, and POAM is in the 5 to 10% range, with larger discrepancies relative to other instruments. Outside this height range rms differences are generally larger, though agreement with HALOE remains good up to 50 km. The assimilation has closer agreement to independent observations than found in direct near-neighbour comparisons between profiles, demonstrating that the assimilation can add value to the observations.

1 Introduction

The Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) instrument on the ENVISAT satellite

Correspondence to: M. Juckes
(m.n.juckes@rl.ac.uk)

measured ozone profiles with near global coverage from July 2002 to March 2004. Compared with the operational SBUV instruments, MIPAS has improvements both in vertical resolution and in its ability to make both night and day measurements. This paper assesses the quality of the MIPAS ozone observations, and some global gridded fields derived from them, by comparison with independent observations from ozone sondes, SAGE II & III, POAM III, HALOE, OSIRIS and SBUV.

The global gridded fields are generated by a new algorithm solving a variational formulation of the data assimilation problem (that is, the solution is defined to be that field which minimises a penalty function). The new algorithm exploits all the information available in a retrospective analysis. Many analyses are constructed using data assimilation systems which have been developed in an operational context (e.g. Bloom et al., 1996; Lorenc, 1996; Courtier, 1997; Talagrand, 1997; Courtier et al., 1998; Rabier et al., 1998; Jeuken et al., 1999; Khattatov et al., 2000; Chipperfield et al., 2002; Struthers et al., 2002; El Serafy and Kelder, 2003; Eskes et al., 2003) designed to produce real-time analyses. In that context the information content in the analysis comes predominantly from observation prior to the analysis time. The system used here, in contrast, makes full use of observations after the analysis time, providing a significantly increased information input. Cohn et al. (1994) have discussed the advantages of making use of the additional, post analysis, data, but found marginal benefit in their sub-optimal implementation. Some of the advantages of making full use of post-analysis time observations are discussed in Juckes and Lawrence (2006).

The methods described in the works cited above rely heavily on a background field which carries information from earlier analyses forward to the current analysis time or time window. The Generalised Inversion Method (GIM, e.g. Bennett et al., 1998) is closer to the approach described below, since it fits a series of observations which is sufficiently long that the influence of the background field is small in most of the

analysis window. For linear systems the Generalised Inversion Method gives the same solution as the Kalman Smoother (e.g. Rodgers, 2000), but it uses substantially less computational resources than the latter when dealing with systems with large numbers of points in the spatial domain. A useful measure of the size of a variational problem is the number of control variables (the control variables are a set of variables which completely define the solution and which are used to adjust the solution towards the minimum of the penalty function). The cost of GIM scales as the product of the number of observations times the number of control variables. This is much more efficient than the Kalman Smoother for systems with large numbers of variables, and makes it usable for interesting physical applications (Bennett et al., 1998), but it is still (in its current formulation) substantially more expensive than currently implemented operational systems.

Lyster et al. (1997) implemented a full Kalman filter (e.g. Rodgers, 2000) with the same modelling constraint, isentropic advection, as used here. The cost of the Kalman filter method scales with the square of the number of spatial mesh points. As with GIM, this makes the Kalman filter impractical for very large problems. Compared with this study, Lyster et al. (1997) used lower resolution and their algorithm could not propagate information back in time.

The large cost of optimal algorithms has led many authors to investigate sub-optimal systems, e.g. reverse domain filling trajectories (Dragani et al., 2002) or low order modelling of the background error covariance (e.g. Riishøjgaard, 1998, 2000, 2001).

The method used in this paper is "direct" in the sense that it uses neither a background field to convey information between analysis segments, as used in the above cited works on 3 and 4D-VAR, nor representers to derive the impact of individual observations, as in GIM. The new method preserves the quasi-elliptic nature of the problem which follows naturally from its specification in terms of a minimisation: the elliptical nature of the problem motivates the choice of a relaxation algorithm to derive the numerical solution.

2 Methodology and data

2.1 Background

Figure 1 shows results of an assimilation run at approximately 1/2 degree latitude/longitude resolution (768 points around the equator). The positions of MIPAS measurements occurring within 2 h of the time shown are marked as coloured crosses, and those occurring within 2 to 6 h are shown as black crosses. There is more information in the field than could be obtained from simple interpolation of the observations (this is discussed in more detail in relation to Fig. 9 below). It will be shown below that observations from several days either side of the displayed field contribute information. This is made possible through use of physical

knowledge in the form of the advection equation which describes the evolution of ozone on these timescales, at this height, to good approximation.

This section describes the mathematical and computational formulation that created this analysis, and then the following section will describe its validation against independent observations.

2.2 The weak physical constraint

Variational methods of data assimilation can be characterised as "strong" or "weak" depending on whether the model being used in the assimilation enters into the variational principle as a "strong constraint", which must be satisfied exactly, or as a "weak constraint", which must only be approximately satisfied (Kalnay, 2002). Here a weak constraint is used. The analysis is defined to satisfy, within error bars, an isentropic advection equation (e.g. Gill, 1982) (that is, advection on surfaces of constant potential temperature):

$$M\left[\chi\right] \equiv \chi_t + \boldsymbol{u} \cdot \nabla \chi = \epsilon, \tag{1}$$

where χ is the ozone mixing ratio, ϵ is a random process with spatial and temporal correlation scales smaller than the numerical discretization, and $\boldsymbol{u}$ is the horizontal wind taken from the operational analyses of the European Centre for Medium-range Weather Forecasts (see Table 3 in Appendix A for full reference). The discretized version of the variational principle stated below is simplified considerably if the ϵ is close to white noise. This can be quantified through the structure function, defined as follows:

$$E\left[\epsilon(\lambda, \phi, t)\epsilon(\lambda_0, \phi_0, t_0)\right] = \sigma_{\mathrm{ap}}^2 f_{\mathrm{ap}}(\lambda, \phi, t; \lambda_0, \phi_0, t_0) \tag{2}$$

where E is the expectation operator and the structure function f_{ap} is normalised so that

$$\int \int \int f_{\mathrm{ap}}(\lambda, \phi, t; \lambda_0, \phi_0, t_0) \cos(\phi) d\phi d\lambda dt = 1. \tag{3}$$

The structure function is assumed to be localised such that

$$f_{\mathrm{ap}}(\lambda, \phi, t; \lambda_0, \phi_0, t_0) \ll 1 \tag{4}$$

if $\ |t - t_0| > \tau_{\mathrm{ap}}$

or $\ (\phi - \phi_0)^2 + \cos(\phi)^2(\lambda - \lambda_0)^2 > l_{\mathrm{ap}},$

for constants τ_{ap} and l_{ap}. This will be exploited in deriving Eq. (5) below.

Note that the definition of M in Eq. (1) includes the time derivative and acts on the full field of $\chi(\lambda, \phi, t)$ values for the whole time window of the analysis. M will be referred to as the "process model", to distinguish it from error and observing models.

Equation (1) should contain both vertical advection and chemical sources and sinks, but here these are considered as unknowns and modelled with the random term, ϵ. In most treatments in the meteorological literature some conceptual simplification is gained by discretizing the problem at this
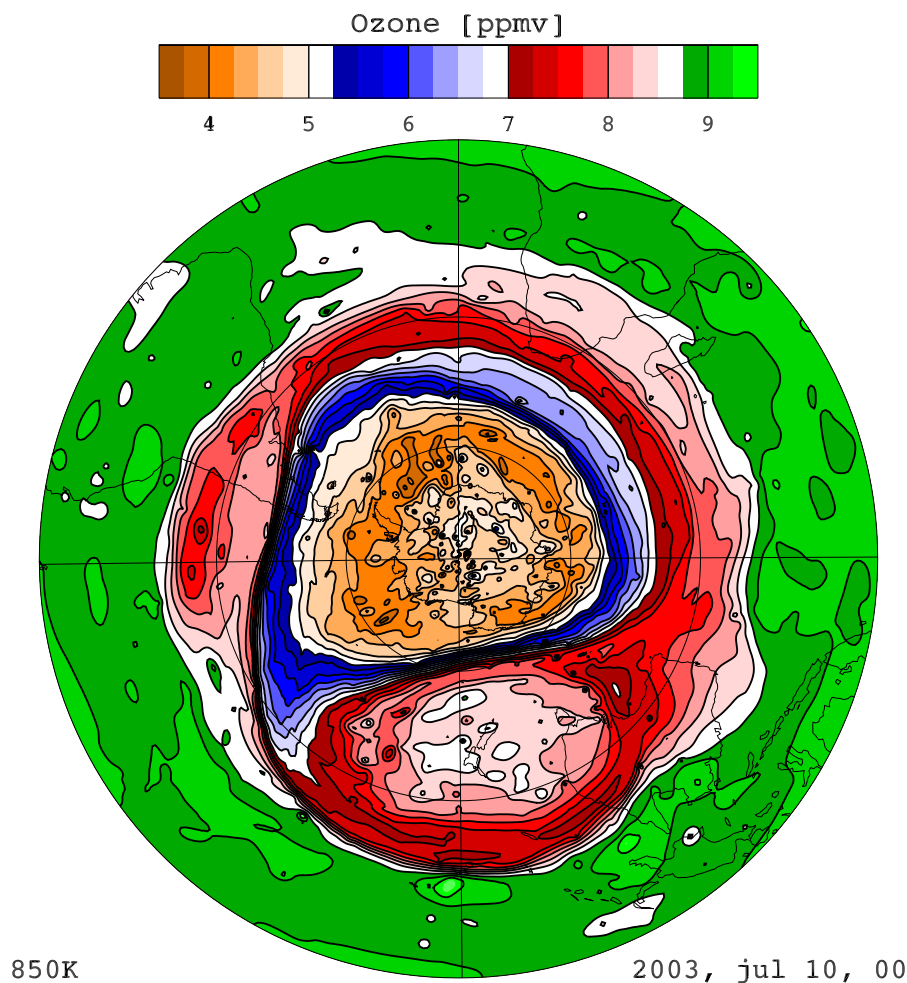
Fig. 1. Southern hemisphere, isentropic fields of ozone, on the 850 K surface, for 10 July 2003, derived from MIPAS observations, 0.5 degree resolution. The Greenwich meridian is at the top of the plot.

stage and representing M by a matrix. However, keeping the analytic form allows the structure which emerges below to be exploited in the choice of discretization.

If ϵ is a random process and if the errors in the observations can also be represented by random terms, the Bayesian formulation can be used to construct a joint likelihood density function (LDF) for the structure of the fields given the observations. The LDF can be expressed as the exponential of minus a penalty function so that the minimum of the penalty function corresponds to the maximum of the likelihood density. Here we restrict attention to Gaussian noise. Appendix B shows how the inverse of the noise autocorrelation function can be constructed for a representative class of such functions, and shows that the inverse can be approximated by a constant if the length scale of the analysis field is large compared to the auto-correlation length scale of ϵ. This will be assumed to be the case here.

Here the penalty function implied by the observations and Eq. (1) is augmented by smoothing terms:

$$\mathcal{J} = \sum_i \sigma_{\text{obs}:i}^{-2} \left[\chi_{\text{obs}:i} - \chi(\lambda_{\text{obs}:i}, \phi_{\text{obs}:i}, t_{\text{obs}:i}) \right]^2$$

$$+ \int\int \left\{ w_{\text{ap}} \left[\chi_t + \boldsymbol{u} \cdot \nabla\chi \right]^2 \right.$$

$$+ w_2 \left| \left(c_1 \frac{\partial}{\partial t}, \nabla \right) \right|^2 \tag{5}$$

$$\left. + w_{\text{num}} \left| \left(c_1 \frac{\partial}{\partial t}, \nabla \right) \left(c_1^2 \frac{\partial^2\chi}{\partial t^2} + \nabla^2\chi \right) \right|^2 \right\} dA dt, \tag{6}$$

where "dA" is an area element on an isentropic surface. The smoothing term, prefixed by w_{num}, imposes regularity near observations. w_2 will be set to zero for most of the results: it will be used to illustrate the effect of removing the advective constraint. c_1 is a constant determining the ratio between spatial and temporal smoothing. Here c_1=0.5 days per radian

is used. The scalar weighting coefficient w_{ap} would ideally be derived from the covariance of a random process model error. As the latter is generally not random, the value of w_{ap} will be determined empirically.

Applying the calculus of variations to the penalty function in Eq. (6) shows that it is minimised with respect to $\chi(\theta, \phi, t)$ when that field satisfies the following analysis equation:

$$\mathcal{A}[\chi] \equiv \mathcal{A}_{obs} + \mathcal{A}_{ap} + \mathcal{A}_{num} = \mathcal{S} \tag{7}$$

where

$$\mathcal{A}_{obs} = \sum_i \frac{\chi(\lambda_{obs:i}, \phi_{obs:i}, t_{obs:i})}{\sigma_{obs:i}^2} \times$$
$$\delta(t - t_{obs:i}) \delta(\lambda - \lambda_{obs:i}) \delta(\phi - \phi_{obs:i}), \tag{8a}$$

$$\mathcal{A}_{ap} = w_{ap} \left(-\frac{D^2 \chi}{Dt^2} \right), \tag{8b}$$

$$\mathcal{A}_{num} = -w_2 \left(c_1^2 \frac{\partial^2}{\partial t^2} + \nabla^2 \right) \chi \tag{8c}$$

$$-w_{num} \left(c_1^2 \frac{\partial^2}{\partial t^2} + \nabla^2 \right)^3 \chi, \tag{8d}$$

$$\mathcal{S} = \sum_i \frac{\chi_i}{\sigma_{obs:i}^2} \delta(t - t_i) \delta(\lambda - \lambda_i) \delta(\phi - \phi_i). \tag{8e}$$

The boundary conditions at the start and end times (t_{start} and t_{end} respectively) are:

$$\frac{D\chi}{Dt}, \quad \frac{\partial}{\partial t} \left(c_1^2 \frac{\partial^2}{\partial t^2} + \nabla^2 \right)^n \chi \equiv 0, \tag{9}$$
$$\text{for} \quad t = t_{start}, t_{end}, \quad \text{and} \quad n = 1, 2.$$

Equations (6) and (7) can also be derived by forming the Euler-Lagrange equations from (1) (e.g. Bennett, 1992) and then eliminating the Lagrange multiplier. The relaxation algorithm described below cannot be applied to the Euler-Lagrange equations themselves because of the intrinsic non-localness in those equations. A small forcing applied to either equation will generally produce a global response because the equations are hyperbolic in nature.

From Eqs. (6) and (8) it follows, by integrating over the globe and over the analysis time window, that

$$\sum_i \frac{\chi_i - \chi(\lambda_{obs:i}, \phi_{obs:i}, t_{obs:i})}{\sigma_{obs:i}^2} = 0, \tag{10}$$

that is, the analysis is unbiased with respect to the observations. This also means that any bias in the observations is present, without modification, in the analysis. Equation (10) gives a simple relation between the analysis averaged over the observation points and the observations. Unfortunately, there is no simple expression relating the observations to the global mean of the analysis: the latter will also depend on the distribution of observations in a non-trivial way.

In this formulation the satellite observations are treated as point measurements. If the spatial averaging kernel of the instrument were taken into account the observation term would contain a convolution with this averaging kernel rather than the delta functions in Eq. (7). If the averaging kernel is smaller than the grid size, the two approaches become identical after discretization. At the highest resolution described below the analysis should be capable of partially resolving the line of sight averaging of the limb viewing instruments. This issue is not addressed here as the focus will be on dealing with the constraint provided by the process model Eq. (1).

Since Eqs. (6) and (7) describes an optimal solution for given error statistics, it follows that the solution is equal to the solution of the Kalman smoother (e.g. Rodgers, 2000), apart from differences in the discretization which may be appropriate for the predictive equations used in the standard Kalman Smoother algorithm as opposed to the quasi-elliptic equation solved here.

2.3 Discretisation

The spatial mesh used is a latitude-longitude grid which thins towards the pole so that the longitudinal spacing does not decrease too drastically. It is described in more detail in Appendix C. This grid preserves some of the simplicity of the latitude-longitude grid but avoids the convergence of points at the poles which can create numerical conditioning problems (e.g. Thuburn and Li, 2000, and references therein). In the calculations described below the longitudinal spacing at the equator is equal to the latitudinal spacing. Five different resolutions will be employed having $12 \times 2^{N_{grid}}$ points around the equator for $N_{grid} = 3$ to 7. More details are listed in Table 1 below.

The solution method is a a multi-grid relaxation algorithm described in appendices D and E. This is an iterative method, and the cost of each iteration is proportional to the number of control variables. The number of iterations appears to be independent of the resolution, so that the net cost of the solution algorithm increases only linearly with the problem size.

The second order time derivative is discretized using the standard 3-point formula. The algorithm does not require the integration of any first order time derivatives, so the usual problems of numerical stability found with forecast and adjoint models do not arise here.

The Lagrangian derivative is evaluated using a semi-Lagrangian scheme with an implicit definition of the parcel displacements:

$$\frac{D^2\chi}{Dt^2} \stackrel{\text{def}}{=} \frac{\chi(\boldsymbol{x}_m, t_m) - 2\chi(\boldsymbol{x}, t) + \chi(\boldsymbol{x}_p, t_p)}{\Delta t^2} \tag{11}$$

$$\boldsymbol{x} - \boldsymbol{x}_m = \frac{\Delta t}{2} [\boldsymbol{u}(\boldsymbol{x}, t) + \mathbf{u}(\boldsymbol{x}_m, t_m)] \tag{12}$$

$$\boldsymbol{x}_p - \boldsymbol{x} = \frac{\Delta t}{2} [\boldsymbol{u}(\boldsymbol{x}_p, t_p) + \mathbf{u}(\boldsymbol{x}, t)], \tag{13}$$

where $t_p = t + \Delta t$, $t_m = t - \Delta t$, and x_p and x_m are the estimated positions at times t_p and t_m respectively, of a parcel which is at position x at time t. The wind fields are taken from the ECMWF operational analyses. The assimilation time step used here is 4 h, using winds interpolated linearly between the 6 hourly ECMWF analyses. Reducing the assimilation time step was not found to have any significant effect, but the effect of wind variability not resolved in the operational analyses could not be tested. The spatial interpolation required in Eqs. (11) and (12) is done with cubic splines for χ and linear interpolation for u.

2.4 The smoothing term

The smoothing term ensures that the problem has a unique solution. The dynamical penalty will be zero for any field satisfying $D\chi/Dt = 0$. Since there are an uncountable infinity of such fields and only a finite number of observations, it is clear that we will in general have an infinite number of solutions which fit all the observations and have zero dynamical penalty. These solutions are such that the value on any trajectory intersecting an observations is given by that observation. In the absence of any additional information the value on trajectories which do not intersect observations is undetermined.

The form of the smoothing term used is constrained by consideration of the structure of the solution near the observations. This structure is determined by a balance between the δ-function forcing (S in Eq. 6) and the highest derivatives in the equation (because the δ-function forcing generates small scales, and with small scales the higher derivatives have the largest magnitude). The highest derivatives here are, by construction, homogeneous in space and time when time is scaled by c_1. Let $r^2 = \Delta\phi^2 + \cos\phi^2\delta\lambda^2 + c_1^{-2}\Delta t^2$ be the space-time distance from an observation. The solution of Eq. (6) near the observation is then

$$\frac{r^3}{4\pi}\left[\chi(\lambda_i, \phi_i, t_i) - \chi_i\right] + a_1 r^2 + a_2$$

where a_1 and a_2 are constants determined by the larger scale solution. If the smoothing term in the analysis equation were 4th order in r, then the leading order term near the observations would be proportional to r. This would imply a singularity in the gradient at each observation. Having a 6th order term guarantees a sufficient degree of smoothness of the solution in the vicinity of the observations so that all the derivatives in the physically motivated a priori constraint can be accurately evaluated.

The solution method described in appendices D and E requires increasing numbers of ancillary variables as the order of the equation is increased. In order to keep the computational cost down it is desirable to use the smallest suitable value, which is 6th order.

2.5 Experimental setup

The main results will be presented at a resolution of $1.875°$. Constants and variables will be non-dimensionalised with length, time and ozone scales of 1 radian, 12 h and 0.2 ppmv respectively. The results discussed below use $w_{ap} = 8$ and $w_{num} = 0.1 \times 2^{-6N_{grid}}$ unless otherwise stated, where N_{grid} is the number of levels in the multi-grid hierarchy ($N_{grid} = 5$ for the resolution of $1.875°$).

Convergence is expressed in terms of a tolerance, $T_{conv} = 0.005$ unless otherwise stated. The iteration is stopped when: (i) the mean square residual in the assimilation equation is less then T_{conv} times the mean square solution, (ii) the fractional changes, over 3 iterations, in the a priori, numerical and observational components of the cost function are less than 2, 5 and 2 times T_{conv} respectively. It is found that these conditions ensure that, over the wide range of parameter values and horizontal resolutions reported here, the iteration reaches a point where further changes are substantially smaller than observational errors.

2.6 Analysis time windows

The 6 month period discussed here is broken up into 6 overlapping segments, and a separate minimisation carried out for each segment. The six time periods are: (1) 21 January to 10 March, (2) 21 February to 10 April, (3) 21 March to 10 May, (4) 21 April to 10 June, (5) 21 May to 10 July and (6) 21 June to 10 August, all in 2003. At points well away from the ends of the time windows the analysis benefits from information of both future and past observations. Near the ends of the time windows this is no longer the case and we can expect some loss of accuracy as a result. The differences between the analyses in the overlap period will be used below to assess the amplitude of the random error in the analysis.

The loss of accuracy at the ends of the time segment could be avoided by using a more sophisticated boundary condition at the end point, but this would require handling a large error covariance matrix. Here the period near the end of the assimilation is used only for quality control. In the forecast situation neither of these latter two options is available: we must simply accept that the error in a real-time assimilation, with no observations after the analysis, will be larger than in a retrospective one.

3 Results

3.1 Hemispheric fields

The MIPAS and SBUV datasets provide near global coverage (the MIPAS dataset has a number of gaps in it and SBUV only measures in the daylight hemisphere), so both can be used to generate global analyses.

Figure 2 shows fields for 00:00 GMT, July 10, 2003, for analyses of MIPAS and SBUV data. This is a period when
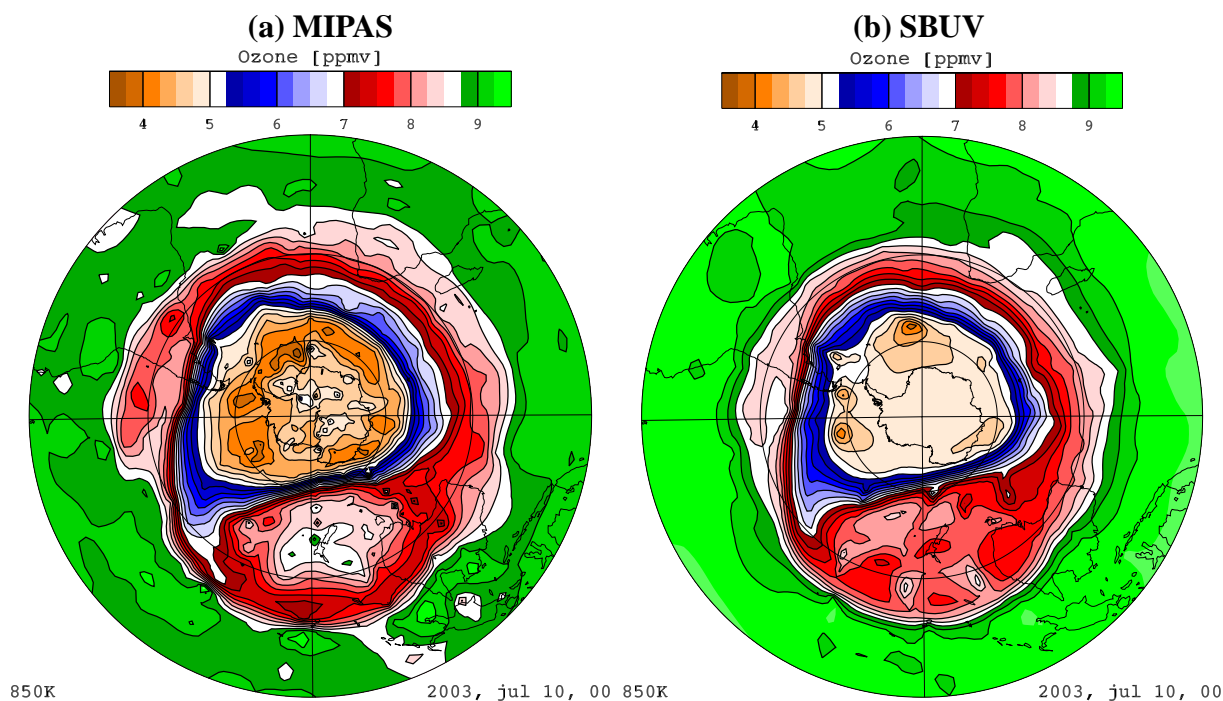
Fig. 2. Southern hemisphere, isentropic fields of ozone, on the 850 K surface, for 10 July 2003. **(a)** MIPAS, **(b)** SBUV. 2 degree resolution.
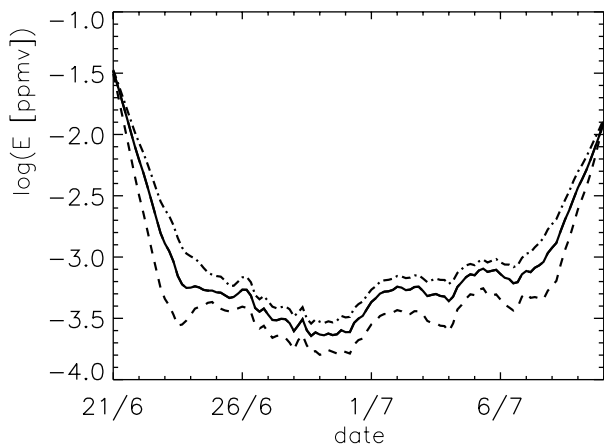

Fig. 3. Global root-mean-square difference between analyses from time windows 5 and 6, on the 850 K surface. For $w_{ap} = 8$ (solid), 1 (dashed) and 32 (dot-dashed).

a wave-breaking event is pulling a streamer of polar vortex air into mid-latitudes. These plots are taken from assimilations at 1.875 degree resolution. The MIPAS analysis shows greater detail: the streamer stretching from the Pacific over Australia at around 25°S, for instance, is not a coherent structure in the SBUV analysis.

Another interesting feature seen in the MIPAS analysis is the ring of low ozone values immediately inside the vortex edge. The physical interpretation of this feature will be discussed elsewhere using a longer study period to cover at least a full annual cycle.

Comparing Fig. 1 with Fig. 2a, it appears that the four-fold increase in spatial resolution used in Fig. 1 has not had a major impact, though there are some small scale features in Fig. 1 which are not resolved in Fig. 2. This point is reinforced below with a quantitative evaluation against independent measurements.

3.2 Random error

As discussed in Sect. 2.6, we expect there to be a loss of accuracy near the end of the analysis time windows. This can be assessed by looking at the difference between two overlapping assimilations. Figure 3 shows the logarithm of the RMS difference between the June and July segments:

$$E(t) = \sqrt{\frac{1}{4\pi} \int \left(\chi^{(5)} - \chi^{(6)} \right)^2 dA}, \qquad (14)$$

where the superscripts (5) and (6) refer to the analysis time segments listed in Sect. 2.6. In the central region the difference levels out at around 0.05 ppmv. This level appears to be related to the convergence criteria used to determine when to stop the iteration towards a solution of Eqs. (6) and (7).

The dashed lines in Fig. 3 show results from experiments using different values of w_{ap}. A larger weighting causes information to persist in the assimilation for longer, and, similarly, causes the detrimental effect of missing information to persist for longer. Hence the loss of accuracy at the ends of the assimilation window persists for longer when w_{ap} is larger.
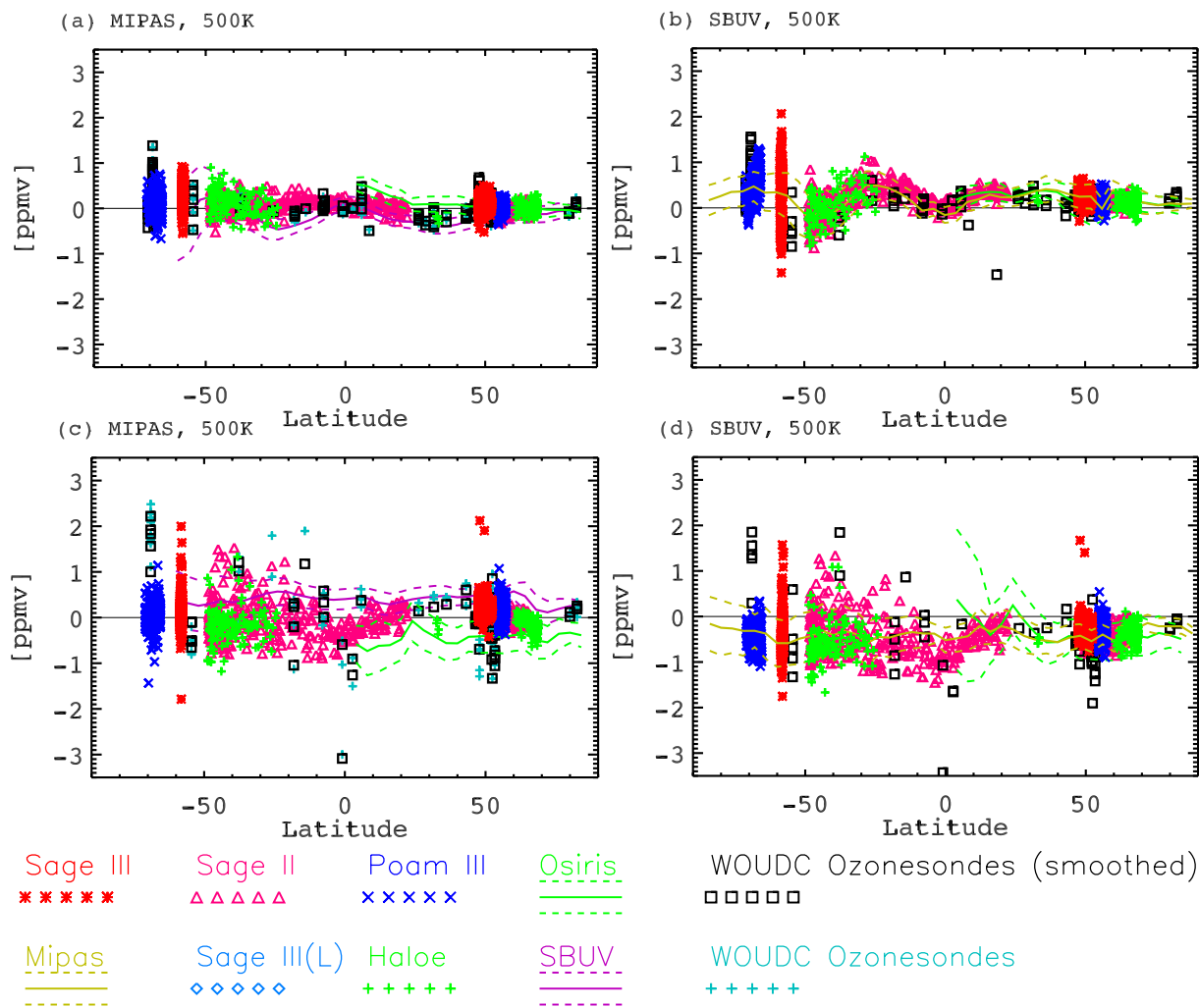
Fig. 4. Validation data, showing differences between independent observations and the gridded fields, for the month of July 2003.

At the ends of the overlap period shown in Fig. 3 we see a near exponential increase in E. The differences at the ends of the analysis windows (around 0.22 ppmv and 0.14 ppmv on 25 June and 12 July respectively) reflect the difference in accuracy between the Kalman Filter and the Kalman Smoother. The e-folding timescale for information persistence is shown by Fig. 3 to be around 2 to 3 days, depending on the magnitude of w_{ap}. This means that the information input to an analysis at any time, t_0 say, comes from a period of 4 to 6 days, and the impact of observations within the same time step as t_0 is relatively small. Consequently, the information content of the Kalman Smoother will be close to twice that of the Kalman Filter. In other words, the error variance at the end of the assimilation window will be twice that in the centre. This allows us to make a heuristic estimate of the random error in the assimilation: Let KF, KS and BF be the solutions of the Kalman Filter, Kalman Smoother and Backward (or reverse) Kalman Filter respectively. If the impact of observations shared by the forward and backward filters is

neglected:

$$\overline{(KF - KS)^2} = \overline{\left[\frac{1}{2}(BF - KF)\right]^2} \tag{15}$$

$$= \frac{1}{2}\sigma_{KF}^2 = \sigma_{KS}^2, \tag{16}$$

where σ^2 is the random error variance and the over bar denotes a global mean. Recall that the solutions of Eqs. (6) and (7) derived here are equivalent to solutions of the Kalman Smoother, and that the Kalman Smoother is equivalent to the Kalman Filter at the end of the time window and the Backward Kalman Filter at the start of the time window. Figure 3 then suggests that the random error of the assimilation, away from the ends of the assimilation window, is around $\sqrt{\exp(-3)} \approx 0.22$ ppmv. This value is consistent with the differences between the analyses and independent observations evaluated below. Figure 3 does not give any indication of possible systematic errors: a bias in the MIPAS measurements, for instance, would lead to a common error in the two

Table 1. Variation of the cost of the numerical algorithm with problem size, for a 50 day analysis, on one isentropic level. The columns show: the number of grid refinements used in the multi-grid method (N_{grid}); the CPU time required (T) [seconds]; the number of spatial mesh points (N_{mesh}); the number of control variables – the number of mesh points times the number of time levels (N_{cv}); the number of CPU cycles per control variable (cycles/cv); and the number of iterations on the finest grid (N_{it}).

N_{grid}	T	N_{mesh}	N_{cv}	cycles/cv	N_{it}
3	36	890	267 890	2.7×10^5	31
4	215	3482	1 048 082	4.1×10^5	47
5	824	13 658	4 111 058	4.0×10^5	38
6	4286	53 594	16 131 794	5.3×10^5	53
7	24 642	213 338	64 214 738	7.7×10^5	72

assimilation segments which would have no impact on the difference shown in Fig. 3. Independent observations will be used below to assess the systematic errors and give a second estimate of the random errors.

The estimate derived above depends on the time symmetry of the variational problem and the solution algorithm. There is no obvious generalisation to other algorithms which lack this symmetry.

3.3 Numerical cost of the method

Solving the elliptical equations with multi-grid relaxation algorithms is known to have a numerical cost that scales with N, where N is the number of variables in the solution. Table 1 shows how the numerical cost of solving the analysis equation (Eq. 8e) varies with resolution using the current algorithm. The computations have been carried out on a single 2 GHz processor. To provide a means of comparing with results on different platforms, the cost is also presented in terms of the number of CPU cycles per control variable. The software is still at an early stage of development with many optimisation issues yet to be explored, so there is potential for improvement on the absolute values of these costs. The most important result at this stage is the ability of the algorithm to deliver near linear dependence of cost on problem size, hence making it a promising candidate for significantly larger data assimilation problems.

3.4 Validation against ozone-sondes

Table 2 lists the biases and standard error of the differences between the different types of sondes and the MIPAS assimilation:

$$\mathcal{B}_{inst} = N_{inst}^{-1} \sum (\chi_{inst} - \chi), \tag{17a}$$

$$\mathcal{SE}_{inst} = \left\{ N_{inst}^{-1} \sum (\chi_{inst} - \chi - \mathcal{B}_{inst})^2 \right\}^{1/2}, \tag{17b}$$

respectively, where the subscript "inst" refers to the instrument used, the sum is taken over all observations at each level, N_{inst} is the number of observations in the sum. All the sonde data is taken from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) web-site (see Appendix A for details), the complete list of the 37 stations used is in Appendix B.

Smit and Kley (1998) describe an extensive intercomparison between different types of sondes. They conclude that the ECC sonde is the most precise. The table shows systematic differences between the analysis and this sonde well below 1% between 420 K and 650 K, averaged over a total of over 800 sonde profiles.

The Indian sondes give anomalously large differences, reading less than half the mixing ratio of the MIPAS analysis at 420 K. In the following diagrams, measurements from these sondes have been omitted.

The result is in line with the results of Migliorini et al. (2004), who found negligible biases in a sample of 30 near coincident MIPAS and ozone sonde measurements, and a difference standard error of around 0.35 ppmv. Here, the rms difference from the ECC sondes, averaged over the 3 lowest levels, is 0.3 ppmv. The differences are greater at 850 K, but the comparison with other satellite instruments (below) suggests that this is due to loss of accuracy in the sondes rather than in the analysis.

Figure 5 below also shows comparisons against ozone sonde profiles which have been vertically smoothed over 2 km. This does not affect the bias, but it can be seen that it makes a significant reduction in the standard error, implying that some of the disagreement between sondes and analysis is due to the lower vertical resolution in the latter. A more accurate method of dealing with this problem is described in Migliorini et al. (2004).

3.5 Validation against other satellite measurements

The scope of stratospheric observations is steadily increasing. This paper analyses ozone measurements from 7 satellite instruments: MIPAS (Fischer and Oelhaf, 1996; Mengistu Tsidu et al., 2003; Glatthor et al., 2005), SBUV (Planet et al., 2001) SAGE II and III (Thomason and Taha, 2003), POAM III (Lumpe et al., 2003; Pierce et al., 2003), OSIRIS (von Savigny et al., 2003), and HALOE (Brühl et al., 1996). The use of a wide range of instruments provides insight, in some instances, into the source of discrepancies when a single instrument departs from the majority. Brief details of the instruments are given in Appendix A.

Figure 4 shows a sample comparison for July 2003, at 500 K and 850 K. For SBUV and OSIRIS the increments, observation minus MIPAS analysis, evaluated at each profile location are averaged in 5 degree latitudinal bins and plotted as a solid purple and green lines respectively, with dashed lines showing plus and minus one standard deviation. For the other measurements a symbol is plotted, as detailed in

Table 2. Comparison of ozone sondes with the MIPAS assimilation, by level and sonde type. For details of the Electrochemical Concentration Cell (ECC) and other sondes, see Smit and Kley (1998). Each entry shows the mean value of the ozone sonde measurements minus the Mipas assimilation, the standard error in brackets and the number of profiles used in the comparison in square brackets. The mean and standard error are expressed as percentages of the MIPAS assimilation values averaged over the measurement locations.

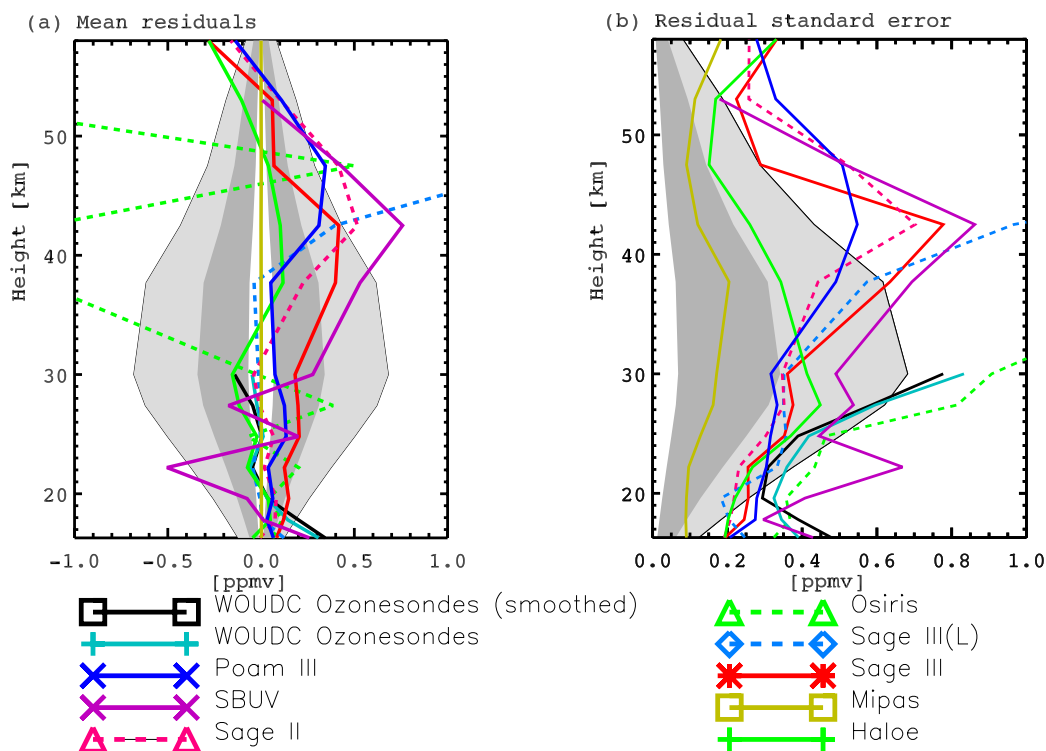| Level	ECC	Indian-sonde	Carbon-Iodine	Brewer-Mast
420 K	−0.1% (23.8%) [850]	−57.2% (29.1%) [32]	−12.7% (32%) [170]	2.8% (21.1%) [82]
500 K	−0.2% (10.1%) [843]	−40.0% (33.1%) [33]	−4.5% (20%) [170]	−0.7% (8.8%) [84]
650 K	−0.4% (8.3%) [813]	−28.5% (20.5%) [30]	2.5% (8.2%) [162]	−2.4% (5.4%) [81]
850 K	−1.0% (10.6%) [697]	−21.7% (22.5%) [9]	7.4% (8.9%) [117]	−9.6% (6.8%) [61]

Fig. 5. The mean value of the verifying observation minus the MIPAS assimilation. The background shading indicates the percentage of the mean profile (mean taken over all instruments, giving each instrument equal weight), dark shading from 1% to 5% and lighter shading from 5% to 10%.

the figure caption, for each measurement. The spread (in this month) grows towards the South pole. This may be related to greater variability in the ozone field itself. The SBUV measurements show a low bias at 850 K and a high bias at 500 K.

Figure 5 shows $\mathcal{B}_{\text{inst}}$ and $\mathcal{SE}_{\text{inst}}$ evaluated over the 6 month assimilation period, February–July 2003. Below 35 km the assimilation has little bias relative to the ozone-sondes (up to 30 km), HALOE, POAM III, or SAGE II. SAGE III measures slightly higher than the assimilation, SBUV is substantially higher, especially in the mid to upper stratosphere. The OSIRIS measurements show a large height dependent bias.

The bias in the SBUV measurements may be due to low vertical resolution: as the ozone number density has a con-cave profile in the mid and upper stratosphere a positive bias would be an expected consequence of vertical averaging associated with broad weighting functions.

Between 35 km and 50 km it appears that MIPAS is measuring low relative to all the other instruments. In this region the difference between HALOE and MIPAS is substantially smaller than the difference with respect to other instruments. This means that either these two instruments are substantially more accurate than the others in this region or they have a common error such that the difference between them is not representative of the actual errors. As HALOE and MIPAS both rely on infra-red measurements, while SAGE, POAM and SBUV exploit ultra-violet radiation, this may point to a
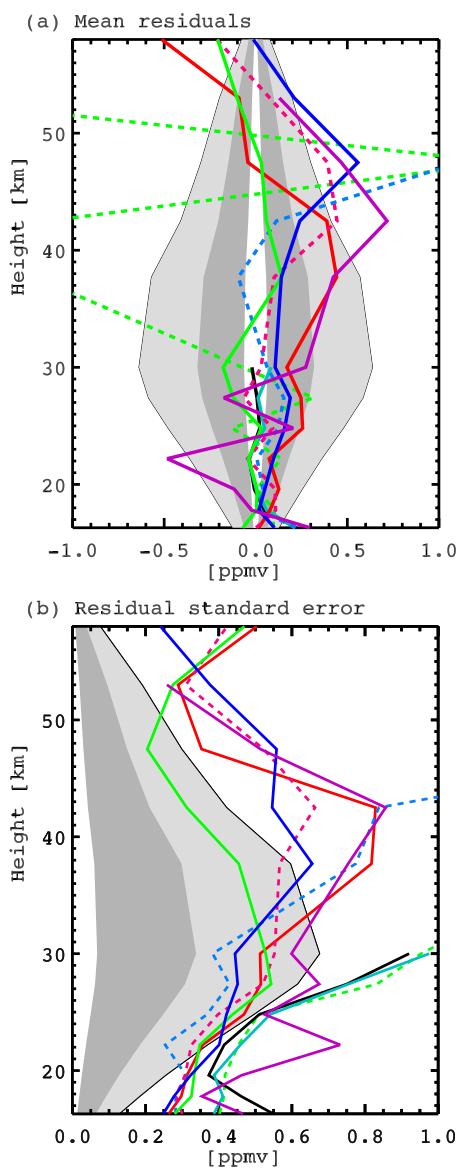
(a) Mean residuals

(b) Residual standard error

Fig. 6. As Fig. 5, except using nearly coincident profiles, with co-location criteria of spacing less than 2.5 degrees and 6 h.

(a) Mean residuals

(b) Residual standard error

Fig. 7. As Fig. 6, except with co-location criteria of spacing less than 5 degrees and 12 h.

problem with the underlying spectroscopy. These results do not, however, give any indication as to whether the UV or IR instruments are more reliable.

Figures 6 and 7 shows some analogous plots created without any data assimilation. Instead, differences between nearby measurements have been evaluated. Pairs of measurements are included in the comparison if they fall within 6 h and 2.5 great circle degrees of each other for Fig. 6, 12 h and 5° for Fig. 7. This constraint means that there are far fewer data points, but still enough for a meaningful comparison. The same pattern of differences emerges, showing that the systematic differences seen in Fig. 5 are not generated by the assimilation system. The differences are generally, though not exclusively, larger in the "nearest neighbour"
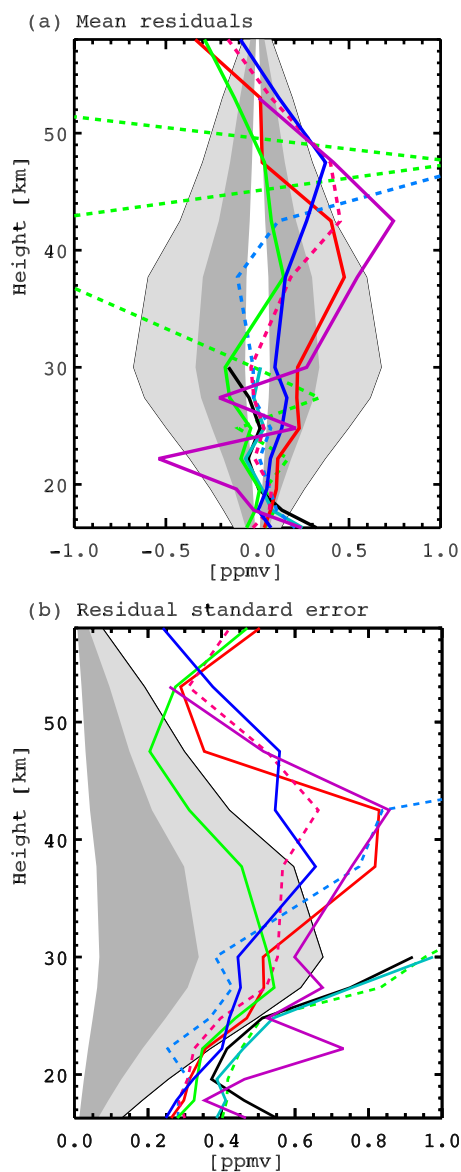
plots, showing that the gridded fields bring significant added value.

Two factors could account for this: firstly, the assimilation accounts for the spatial and temporal variation of the ozone field. This should improve the inter-comparison, provided that the variations are represented with sufficient accuracy. Secondly, the value of the assimilation at any point is a weighted average of many MIPAS observations. This could lead to a reduction in the random error.

Figure 4a also gives an indication of the biases between the other instruments. These results are in line with Danilin et al. (2002), who show HALOE ozone measurements to be around 4% systematically lower than SAGE II.

At 30 km there is a cluster of instruments with $\mathcal{SE}_{\text{inst}} \approx$ 0.35 ppmv (Fig. 5b). If this difference is equally partitioned between the assimilation and the verifying instruments, and if these differences are assumed to be independent, this implies a random error in the analysis, at this height, of 0.25 ppmv, which is very close to the estimate (0.22 ppmv) derived in section 2 from overlapping analyses.

3.6 Dependence on resolution

Comparison of Figs. 1 and 2a suggest that there is little change in the analysis when the resolution is changed from 0.47 degree resolution to 1.875 degree. The higher resolution does capture some extra small scale features, such as small filaments coming off the vortex at around 10 E and 50 E and a thin strip of intermediate valued ozone between the vortex edge and the extended filament which is "reconnecting" with the vortex at 130 E.

A quantitative evaluation is shown in Fig. 8, with number of mesh points around the equator varying from $N_{\text{eq}}=48$ (7.5°) up to $N_{\text{eq}}=768$ (0.47°). The root-mean-square departure from observations is plotted, using the 850 K analyses and the entire 6 month study period. In Fig. 8 the mean is taken as an average of monthly means, whereas Fig. 6 shows an average with each observation equally weighted. This means that months with more observations get greater weight in Fig. 6. A consequence of this difference is that the assimilation minus Osiris standard error appears larger in Fig. 6 than in Fig. 8.

There is a clear improvement between $N_{\text{eq}}=48$ and 96, and a small but still significant improvement when the resolution is increased again to 192. After that point, however, there is no significant gain in accuracy as measured by the $\mathcal{SE}_{\text{inst}}$ validation statistic.

This invariance to changes in resolution also shows that the smoothing term is not having a major influence on the solution for $N_{eq} \geq 192$. The coefficient of the smoothing term varies as the 6th power of the resolution, so there is a reduction by a factor of $4^6=4096$ in the amplitude of this term between the $N_{eq}=192$ and the $N_{eq}=768$ analyses.

3.7 Dependence on weighting of the process model

Figure 9 shows how $\mathcal{SE}_{\text{inst}}$ varies as the weighting of the process model, w_{ap} is varied. There is remarkably little sensitivity over a wide range of values, but when w_{ap} is decreased below 10^{-2} there is a rapid increase in the residuals. The loss of accuracy is less severe, but still significant, if a small second order smoothing term is also included (symbols in Fig. 9). The high order smoothing on its own tends to produce over- and under-shoots: the inclusion of a second order prevents this.

It may be tempting to think of the process model term in the cost function as being "like" the background term in the widely used strong constraint version of 4D-VAR formulated
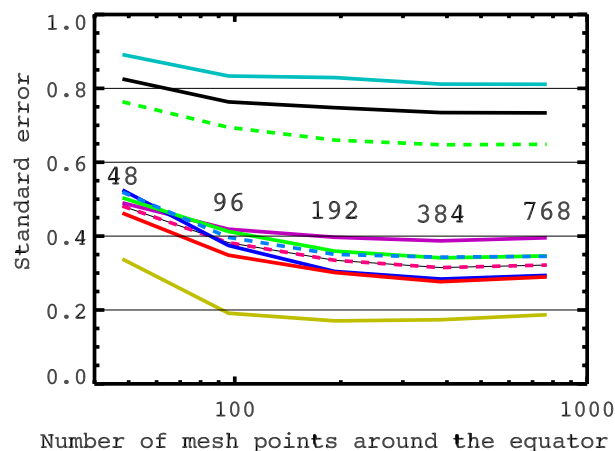

Fig. 8. $\mathcal{SE}_{\text{inst}}$ (Eq. 19) for varying resolution, given in terms of the number of mesh points around the equator. The colouring and line styles are as for Fig. 5. Evaluated from assilmilations for July, 2003, on the 850K surface.

by Talagrand (1997). In the latter, the cost function consists of two terms, the non-observation term being a background term. Thus, in so far as it is a non-observation term, the process model term here is "like" the background term. However, unlike the background term used in 4D-VAR, the process model term here does not involve any empirical correlation structures. A further important point, which may explain the insensitivity of the results to w_{ap}, is that the process model term has a large null space: that is, there is a large family of solutions of $D\chi/Dt=0$ for which $A_{\text{ap}}\equiv0$. Increasing w_{ap} will move the analysis closer to the null space of $\mathcal{A}_{\text{ap}}$. However, this might not be a very large change in absolute terms if the distance between χ and that null space is already small. With the standard form of 4D-VAR the background term has no null space – it can specify the solution completely. If the weighting of this term is increased the contribution of the observations will eventually become insignificant. This cannot happen here: there will always be a significant input of information from the observations, no matter how large w_{ap} is made. It is, however, found that, with the present iteration algorithm, increasing w_{ap} slows the convergence.

4 Conclusions

The direct inversion method for data assimilation, which applies the evolution equations as a weak constraint, has been demonstrated to work for isentropic analyses of stratospheric tracers. In the context of global geophysical data sets, this is a moderate sized system: much smaller than operational meteorological analysis systems but large enough for some of the problems associated with large systems to be relevant. For the Kalman filter, for example, the cost scales with the square of the number of control variables held at each time
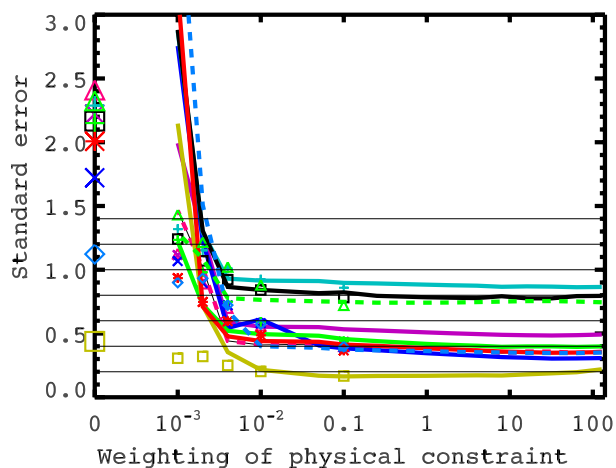
Fig. 9. As for Fig. 8, except that the x-axis is w_{ap}, logarithmic from $w_{ap}=10^{-3}$ to 140, with an extra point at $w_{ap}=0$. The lines show results for $w_2=0$, the symbols for $w_2=10^{-4}$.

level. With the present method the cost is nearer to being linearly dependent on the number of control variables.

The efficiency of the algorithm makes it possible to increase the spatial resolution to the point were the solution becomes essentially independent of the resolution (Sect. 3.6).

The resulting analyses have been compared with radiosondes and a wide range of other satellite instruments. The bias relative to the ECC sonde data is extremely small, less than 0.5% below 650 K. Agreement with HALOE, SAGE II, POAM III and SAGE III lunar occultation retrievals is good up to around 40 km. Above that height the MIPAS analysis develops a significant negative bias. SAGE III measures high relative to the analysis throughout the stratosphere, as does SBUV above 25 km.

The random error estimates of 0.23 ppmv (about 4%) at 850 K derived from comparison of overlapping analysis periods in Sect. (3.2) are consistent with the standard error of differences from other instruments. This is closer than agreement found by analysing close-located profiles from other instruments, demonstrating the added value gained by imposing the isentropic advection constraint on the observations to construct the analysis.

The typical e-folding time for the decay of influence of an observation in the analysis is found to be 2 to 3 days. This time-scale depends on the assumed skill of the model, as reflected in the weighting given to the model constraint in the analysis.

The gridded ozone fields produced in this study are available from http://home.badc.rl.ac.uk/mjuckes/mista/ as Netcdf files (CF compliant).

Appendix A

The instruments

There are three main classes of instruments in this study. Four occultation instruments (HALOE, POAM III, SAGE II & III) measure the absorption of solar radiation by viewing the sun through the atmospheric limb as the satellite moves into and out of darkness. This gives good vertical resolution and accurate results, but is restricted to around 28 profiles per day. This is not enough to represent the spatial structure in the fields. Two instruments measure scattered solar radiation. SBUV is nadir viewing, measuring backscattered UV. This provides good spatial coverage in the sunlit hemisphere, but the vertical resolution is poor. Osiris measures limb scattered UV, giving better vertical resolution. Lastly, MIPAS measures the emitted infrared spectrum with limb viewing geometry. This provides global coverage, day and night. The vertical resolution is moderate, better than the nadir viewing instruments but not as good as that achieved by the occultation instruments

SAGE III data were obtained from the NASA Langley Research Center EOSDIS Distributed Active Archive Center. A few details of the instruments are listed in Table A1, together with the web-sites which contain further information and access to the data.

The ozone sonde data used comprises profiles from 37 stations (Table A2). WOUDC is one of five World Data Centres which are part of the Global Atmosphere Watch programme of the World Meteorological Organization. The WOUDC is operated by the Experimental Studies Division of the Meteorological Service of Canada.

Appendix B

Noise correlation structure functions and their inverses

This appendix derives the form of the process model error covariance used in the analysis equation. If the problem is discretized, the steps below would be essentially trivial, amounting to no more than assuming that the process model error covariance is diagonal under the assumed discretization. In the present work it is advantageous to delay the discretization of the problem, so the equivalent steps must be worked through in the continuous form of the problem. This does not produce any surprises, but it helps to clarify the relation between the continuous noise process in Eq. (1) and the discretized form.

The correlation structure is first approximated by a simple analytic form which can be inverted, in the sense described below. The inverse can then be used to construct the Bayesian likelihood distribution function.

The Bayesian expression for the joint likelihood function is made up of terms containing the inverses of error

Table A1. Full names of instruments which have provided data for this study and web sites where further information can be obtained.

Acronym	Level 2 Version	Full name	Web site	Satellite	Launch
HALOE	19	Halogen Occultation Experiment	http://haloedata.larc.nasa.gov/	UARS	1991
MIPAS	4.61	Michelson Interferometer for Passive Atmospheric Sounding	http://envisat.esa.int/mipas	ENVISAT	2002
POAM III	3	Polar Ozone and Aerosol Measurement	http://wvms.nrl.navy.mil/POAM/	SPOT-4	1998
SAGE II	6.2	Stratospheric Aerosol and Gas Experiment II	http://www-sage3.larc.nasa.gov/	ERBS	1984
SAGE III	3	Stratospheric Aerosol and Gas Experiment III	http://www-sage2.larc.nasa.gov/	Meteor-3M	2001
SBUV II	61610	Solar Backscatter Ultraviolet	orbit-net.nesdis.noaa.gov/crad/sit/ozone/	NOAA 16	series
OSIRIS	1.2	Optical Spectrograph and Infrared Imaging System	http://www.osiris.yorku.ca	ODIN	2001
WOUDC	–	World Ozone and Ultraviolet Radiation Data Centre	http://www.woudc.org	–	–
Winds:					
ECMWF	–	European Centre for Medium-Range Weather Forecasts	www.ecmwf.int (distributed by http://badc.nerc.ac.uk)	–	–

covariances. Here, the error covariance has been expressed as a continuous function, f_{ap}, rather than as a matrix. The inverse of the function may then be expressed as a differential operator with the property that:

$$L\left[f_{ap}\right] = \cos^{-1}(\phi)\delta(t - t_0)\delta(\lambda - \lambda_0)\delta(\phi - \phi_0).$$

As the spherical geometry introduces some algebraic complexity which is not relevant here, the rest of this appendix will deal with Cartesian geometry. Let

$$\tau = c_1 t$$

and suppose that the covariance structure function depends only on the space-time distance between two points given by:

$$r^2 = (x - x_0)^2 + (y - y_0)^2 + (\tau - \tau_0)^2.$$

The Bayesian formulation requires the inverse of the error covariance. In the continuous representation the inverse should be interpreted as an operator which produces a delta function when applied to the covariance function. The following three equations show examples of the covariance functions associated with differential operators of different orders:

$$F_1 = \frac{1}{4\pi r l_{ap}^2} \exp\left(\frac{-r}{l_{ap}}\right); \left(1 - l_{ap}^2\nabla_*^2\right) F_1 = \delta_*, \quad \text{(B1a)}$$

$$F_2 = \frac{1}{8\pi l_{ap}^3} \exp\left(\frac{-r}{l_{ap}}\right); \left(1 - l_{ap}^2\nabla_*^2\right)^2 F_2 = \delta_*, \quad \text{(B1b)}$$

$$F_3 = \frac{l_{ap} + r}{32\pi l_{ap}^4} \exp\left(\frac{-r}{l_{ap}}\right); \left(1 - l_{ap}^2\nabla_*^2\right)^3 F_3 = \delta_*, \quad \text{(B1c)}$$

where

$$\nabla_*^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial \tau^2}$$

$$\delta_* = \delta(x - x_0)\delta(y - y_0)\delta(\tau - \tau_0),$$

and l_{ap} is a correlation length scale. In Eq. (B1a) the differential operator is a Laplacian, but w explained earlier. the corresponding structure function is unbounded as $r \to 0$. Equation (B1b) uses a higher order Laplacian and the corresponding structure function is finite at the $r \to 0$, but has a discontinuity in the gradient. Higher order operators can be used (Eq. B1c) to generate functions with more smoothness: here a continuous gradient at $r = 0$.

If

$$f_{ap} = \sigma_{ap}^2 F_2,$$

then the inverse error covariance has the form:

$$C_{ap}^{-1} = \sigma_{ap}^{-2} \left(-l_{ap}^2\nabla_*^2 + 1\right)^2.$$

In order to arrive at the analysis equation which will be solved it is now assumed that

$$l_{ap} \ll l_{fd},$$

where l_{fd} is the minimum spacing of the discretization grid. With this assumption the inverse error covariance can be approximated:

$$C_{ap}^{-1} \approx \sigma_{ap}^{-2}.$$

Note that the variance of the model error is $f_{ap}(0) = \sigma_{ap}^2/(8\pi l_{ap}^3)$. The above calculation shows

Table A2. Name, position, country and instrument type of sondes used in this study. The total number of ascents in the 6 month period presented here is 860.

Number	Name	# Profiles	Latitude	Longitude	Country	Type
191	Samoa	12	−14.22	−170.55	American Samoa	ECC
323	Neumayer	40	−70.64	−8.25	Antarctica	ECC
450	Davis	10	−68.57	77.97		ECC
328	Ascension Island	19	−7.97	−14.41	Ascension Island	ECC
29	Macquarie Island	28	−54.49	158.95	Australia	ECC
394	Broadmeadows	21	−37.67	144.95		ECC
18	Alert	31	82.5	−62.32	Canada	ECC
21	Stony Plain	24	53.55	−114.1		ECC
24	Resolute	15	74.71	−94.96		ECC
76	Goose Bay	25	53.31	−60.35		ECC
77	Churchill	16	58.74	−94.06		ECC
315	Eureka	28	79.99	−85.93		ECC
242	Praha	37	50.02	14.45	Czech Republic	ECC
434	San Cristobal	18	−0.91	−89.59	Ecuador	ECC
438	Suva	14	−18.12	178.4	Fiji	ECC
99	Hohenpeissenberg	62	47.8	11.02	Germany	Brewer-Mast
174	Lindenberg	27	52.21	14.12		ECC
316	De Bilt	24	52.1	5.18	Holland	ECC
10	New Delhi	11	28.65	77.22	India	Indian-sonde
187	Poona	7	18.53	73.85		Indian-sonde
205	Thivandrum	8	8.48	76.95		Indian-sonde
437	Watukosek (Java)	26	−7.49	112.6	Indonesia	ECC
336	Isfahan	1	32.51	51.7	Iran	ECC
318	Valentia Observatory	7	51.93	−10.24	Ireland	ECC
7	Kagoshima	22	31.6	130.6	Japan	Carbon-Iodine
12	Sapporo	24	43.1	141.3		Carbon-Iodine
14	Tateno (Tsukuba)	22	36.1	140.1		Carbon-Iodine
101	Syowa	37	−68.99	39.6		Carbon-Iodine
190	Naha	21	26.2	127.7		Carbon-Iodine
443	Sepang Airport	12	2.73	101.7	Malaysia	ECC
256	Lauder	20	−45.03	169.68	New Zealand	ECC
89	Ny-Ålesund	32	78.93	11.95	Norway	ECC
221	Legionowo	29	52.4	20.97	Poland	ECC
265	Irene	12	−25.89	28.22	South Africa	ECC
401	Santa Cruz	17	28.46	−16.25	Spain	ECC
435	Paramaribo	25	5.81	−55.2	Suriname	ECC
156	Payerne	76	46.49	6.57	Switzerland	ECC

that, under the assumption that l_{ap} is small compared to the grid spacing, the value of the variance needed in the discretized equations is σ_{ap}^2. This is equal to the volume integral of the error covariance structure function.

Appendix C

The binary-thinned latitude-longitude spherical grid

The spherical grid used here is a compromise between the convenience of a simple latitude longitude grid and the efficiency of grids with near uniform node distributions. The nodes lie on a set of equidistant latitudes, including both poles and the equator.

The grid is defined by five parameters listed in Table C1. The number of latitudes, including both poles, is $(N_{lat0}-1)\times 2^{N_{grid}-1}+1$, and they are equi-spaced. The number of grid points around the equator is $N_{eq}=N_{eq0}\times 2^{N_{grid}-1}$. Moving polewards from the equator, the number of nodes around a latitude circle is halved every time the node spacing falls below γ_{ref} times the node spacing at the equator. The thinning stops before the number of points falls below $N_{lon\,min}$. It can be shown that each finer grid includes all the points of the coarser grids in the hierarchy.

Table C1. Parameters describing the grid used in this study, with variable longitudinal spacing. See text for details.

Parameter	values used	description
N_{grid}	3,4,5,6,7	The number of grid refinements
N_{lat0}	7	The number of latitudes on the coarsest grid
N_{eq0}	12	The number of nodes around the equator on the coarsest grid
γ_{ref}	0.7	Determines latitudes at which node spacing is doubled
$N_{\text{lon min}}$	12	Minimum number of longitudinal nodes

Grid sophistication can be measured by the ratio of the maximum to minimum node spacing, γ_{mesh}. Thuburn and Li (2000) and references therein achieve a value of near the optimal value of unity. A simple latitude/longitude grid, on the other hand, has $\gamma_{\text{mesh}} \approx N_{\text{lat}}$, which grows as the resolution increases. The present method has $\gamma_{\text{mesh}} = 2$. Having a high value of γ_{mesh} is inefficient, because accuracy is limited by the largest grid spacing and having more densely packed nodes at the poles adds to computational cost without contributing to improved accuracy. The indirect effects of the inhomogeneous grid spacing are generally more significant than the direct effects of having more nodes than necessary. With $\gamma_{\text{mesh}} = 2$ the inhomogeneity introduced by the mesh is less than the inhomogeneity in typical satellite observations, where the ratio of along track to across track observational spacing may be a factor of 20. This suggests that the mesh defined above will be sufficiently homogeneous for present purposes.

A key factor in the implementation of the multi-grid algorithm is the consistency of the transformation between mesh refinements with the discretization of the Laplace operator. It is easily shown that $\int \nabla^2 \psi \, dA = 0$, for any field ψ, where the integral is taken over the surface of the sphere. Once a finite difference representation of ∇^2 has been chosen, the representation of integrals should be chosen to respect the above constraint. This constrains the choice of weights given to each node. This may imply, for instance, using $\cos(\theta - d\theta/2) - \cos(\theta + d\theta/2)$ instead of $d\theta \sin \theta$ (where θ is co-latitude). The two expressions give the same formal accuracy, but the first also satisfies the integral condition exactly. It has been found that using such weights in the mesh refinement process leads to an efficient multi-grid algorithm, whereas using weights with the same formal accuracy but not satisfying the integral condition slows convergence substantially.

Appendix D

The relaxation algorithm

The relaxation algorithm is a standard iterative solution method for second order elliptical partial differential equations. In its simplest form, if A is the discretization of the analysis equation and S the source term, the iteration is defined by:

$$Au^{(k)} - S = r^{(k)}, \tag{D1}$$

$$u^{(k+1)} = u^{(k)} - A_{\text{diag}}^{-1} r^{(k)}, \tag{D2}$$

where A_{diag} is the diagonal matrix obtained by setting the off-diagonal components of A to zero. The (k) superscript here denotes the kth approximation in an iterative approach to the solution. If A is diagonally dominant (that is, if every diagonal term is greater than the sum of absolute values of off diagonal terms in the corresponding row) it can be shown that $|r^{(k)}| \to 0$ as k increases. Unfortunately, this condition cannot be fulfilled by any discretization of the high order smoothing term in the analysis equations being solved here.

To overcome this problem the differential equation is split into a multi-component system. Auxiliary variables $u_a^{(m:k)}$, $m = 0, 1, 2$, are introduced such that

$$\begin{pmatrix} \rho + w_{\text{ap}}\mathcal{A}_{\text{ap}} & 0 & -w_{\text{num}}\nabla^2 \\ -\nabla^2 & 1 & 0 \\ 0 & -\nabla^2 & 1 \end{pmatrix} \begin{pmatrix} u_a^{(0:k)} \\ u_a^{(1:k)} \\ u_a^{(2:k)} \end{pmatrix} = \\ \begin{pmatrix} r^{(0:k)} + src \\ r^{(1:k)} \\ r^{(2:k)} \end{pmatrix}, \tag{D3}$$

where $r^{(m:k)}$ is an analysis residual and the superscript now includes both the index of the auxiliary variables and of the iteration number. $u_a^{(0:k)} \equiv u^{(k)}$. The source term is

$$src = \sum_i u_{\text{obs}:i} \delta(x - x_i^{(\text{obs})}) \delta(y - y_i^{(\text{obs})}).$$

and the observation density is $\rho = w_{\text{obs}} \sum_i \delta(x - x_i^{(\text{obs})}) \delta(y - y_i^{(\text{obs})})$. The aim is to construct an iteration such that $r^{(\cdot:k)} \to 0$ as $k \to \infty$ and hence $u_a^{(0:k)} \to u_a$ and $u_a^{(1:k)} \to \nabla^2 u_a$. This is done by defining an adjustment $a^{(m:k)}$ by solving an approximate form of (Eq. D3) obtained by neglecting all non-local coefficients. This leaves a 3×3 matrix equation at each mesh node. Let

$$A_{\text{block}-\text{diag}} \equiv \begin{pmatrix} \rho_0 + w_{\text{ap}}\gamma_{\text{ap}} & 0 & -w_{\text{num}}\gamma \\ \gamma & 1 & 0 \\ 0 & \gamma & 1 \end{pmatrix},$$

where ρ_0 is the discretized observation density and $\gamma = \frac{2}{c_1^2 dt^2} + \frac{2}{dx^2} + \frac{2}{dy^2}$ for a Cartesian grid with spacing dx and dy in the x and y directions respectively. This 3×3 matrix can easily be inverted, and the solution estimate is then updated as follows:

$$u_a^{(m:k+1)} = u_a^{(m:k)} - \sum_{n=0,2} \left(A_{\text{block}-\text{diag}}^{-1} \right)_{mn} r^{(n:k)}. \tag{D4}$$

Appendix E

Multi-gridding

A multi-gridding algorithm has been employed to speed up the relaxation algorithm. This involves a hierarchy of grids with successively reduced resolution. This approach is motivated by the fact that the relaxation algorithm can handle small scale structures very efficiently but is slow to adjust large scale structures. The lower resolution grids provide information about the large scale structures which can be used to speed up convergence towards a solution on the finest grid. Let

$$u_a^{(.:k:g)} \equiv (u_a^{(0:k:g)}, u_a^{(1:k:g)}, u_a^{(2:k:g)})$$

be an approximation to u_a and associated ancillary variables, where k is an iteration number and g is a grid number, with $g=0$ corresponding to the highest resolution grid.

The analysis operator must then be available in a form which can be applied to any of the grids. If this operator is defined in differential form this does not pose any problems.

We define an analysis residual on the analysis grid:

$$r_a^{(.:k:0)} = \mathcal{A}\left[u_a^{(.:k:0)}\right]$$

and introduce a smoothing operator to transfer the residual fields onto the coarser grids:

$$r_a^{(.:k:g)} = \mathcal{S}\left[r_a^{(.:k:g-1)}\right], \qquad g=1, ...g_{max}.$$

The relaxation algorithm can now be applied to generate an adjustment on the coarsest grid:

$$a^{(.:k:g_{max})} = A_{block-diag}^{-1}\left[r^{(.:k:g_{max})}\right].$$

We then introduce an interpolation operator $\mathcal{U}$ which moves fields from coarse to finer grids and define a preliminary adjustment:

$$a_*^{(m:k:g)} = \mathcal{U}\left[a^{(m:k:g+1)}\right],$$

The adjustment on finer grids is then defined by

$$a^{(m:k:g)} = \mathcal{R}\left[r^{(m:k:g)} - \mathcal{A}\left[a_*^{(m:k:g)}\right]\right] + a_*^{(m:k:g)}.$$

Edited by: M. Dameris

References

Bennett, A. F.: Inverse methods in physical oceanography, Cambridge University Press, Cambridge, UK, 1992.

Bennett, A. F., Chua, B. S., Harrison, D. E., and McPhaden, M. J.: Generalized inversion of tropical-atmosphere-ocean (TAO) data and a coupled model of the tropical pacific, J. Climate, 11, 1768–1792, 1998.

Bloom, S. C., Takacs, L. L., Silva, A. M. D., and Ledvina, D.: Data Assimilation Using Incremental Analysis Update, Mon. Weather Rev., 124, 1256–1271, 1996.

Brühl, C., Drayson, S. R., II, J. M. R., Crutzen, P. J., McInerney, J. M., Purcell, P. N., Claude, H., Gernandt, H., McGee, T. J., McDermid, I. S., and Gunson, M. R.: Halogen Occultation Experiment ozone channel validation, J. Geophys. Res., 101, 10 217–10 240, 1996.

Chipperfield, M., Khattatov, B., and Lary, D. J.: Sequential assimilation of stratospheric chemical observations in a three-dimensional model, J. Geophys. Res., 107(D21), 4585, doi:10.1029/2002JD002110, 2002.

Cohn, S. E., Sivakumaran, N. S., and Todling, R.: A fixed-lag Kalman smoother for retrospective data assimilation, Mon. Wea. Rev., 122, 2838–2867, 1994.

Courtier, P.: Dual formulation of four-dimensional variational assimilation, Quart. J. Roy. Meteorol. Soc., 123, 2449–2462, 1997.

Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., and Fisher, M.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var), I: Formulation, Quart. J. Roy. Meteorol. Soc., 124, 1783–1807, 1998.

Danilin, M. Y., Ko, M. K. W., Bevilacqua, R. M., Lyjak, L. V., Froidevaux, L., Santee, M. L., Zawodny, J. M., Hoppel, K. W., Richard, E. C., Spackman, J. R., Weinstock, E. M., Herman, R. L., McKinney, K. A., Wennrg, P. O., Eisele, F. L., Stimpfle, R. M., Scott, C. J., Fahey, D. W., Elkins, J. W., Webster, C. R., Wilson, J. C., Perkins, K. K., and Bui, T. V.: Comparison of ER-2 Aircraft and POAM-III, MLS, and SAGE-II Measurements during SOLVE Using Traditional Correlative Analysis and Trajectory Hunting Technique, J. Geophys. Res., 107(D20), 4420, doi:10.1029/2001JD002012, 2002.

Dragani, R., Redaelli, G., Visconti, G., Mariotti, A., Rudakov, V., MacKenzie, A. R., and Stefanutti, L.: High resolution stratospheric tracer fields reconstructed with lagrangian techniques: a comparative analysis of predictive skill, J. Atmos. Sci., 59, 1943–1958, 2002.

El Serafy, G. Y. and Kelder, H. M.: Near-real-time approach to assimilation of satellite-retrieved 3D ozone fields in a global model using a simplified Kalman filter, Quart. J. Roy. Meteorol. Soc., 129, 3099–3120, 2003.

Eskes, H., van Velthoven, P. F. J., Valks, P. J. M., and Kelder, H.: Assimilation of GOME total ozone satellite observations in a three dimensional tracer transport model, Quart. J. Roy. Meteorol. Soc., 129, 1663–1681, 2003.

Fischer, H. and Oelhaf, H.: Remote sensing of vertical profiles of atmospheric trace constituents with MIPAS limb-emission spectrometers, Appl. Opt., 35, 2787–2796, 1996.

Gill, A. E.: Atmosphere-Ocean Dynamics, Academic Press, New York, 1982.

Glatthor, N., von Clarmann, T., Fischer, H., Funke, B., Grabowski, U., Höpfner, M., Kellmann, S., Kiefer, M., Linden, A., Milz, M.,

Steck, T., Stiller, G. P., Tsidu, G. M., and Wang, D.-Y.: Mixing Processes during the Antarctic Vortex Split in September October 2002 as Inferred from Source Gas and Ozone Distributions from ENVISAT/MIPAS, J. Atmospheric Sci., 62, 787–800, 2005.

Jeuken, A. B. M., Eskes, H. J., van Velthoven, P. F. J., Kelder, H. M., and Hólm, E. V.: Assimilation of total ozone satellite measurements in a three-dimensional tracer transport model, J. Geophys. Res., 104(D5), 5551–5564, 1999.

Juckes, M. and Lawrence, B.: Data Assimilation for Reanalyses: potential gains from full use of post-analysis-time observations, Tellus A, 58 (2), 171–178, 2006.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, Cambridge, UK, 2002.

Khattatov, B. V., Lamarque, J.-F., Lyjak, L. V., Menard, R., Levelt, P. F., Tie, X. X., Gille, J. C., and Brasseur, G. P.: Assimilation of satellite observations of chemical species in global chemistry-transport models, J. Geophys. Res., 105(D23), 29 135–29 144, 10.1029/2000JD900466, 2000.

Lorenc, A. C.: Development of an Operational Variational Assimilation Scheme, J. Meteorol. Soc. Japan, 75, 339–346, 1996.

Lumpe, J. D., Fromm, M., Hoppel, K., Bevilacqua, R. M., Randall, C. E., Browell, E. V., Grant, W. B., McGee, T., Burris, J., Twigg, L., Richard, E. C., Toon, G. C., Sen, B., Boesch, H., Fitzenberger, R., and Pfeilsticker, K.: Comparison of POAM III ozone measurements with correlative aircraft and balloon data during SOLVE, J. Geophys. Res., 108(D5), 8316, doi:10.1029/2001JD000472, 2003.

Lyster, P. M., Cohn, S. E., Menard, R., Chang, L.-P., Lin, S.-J., and Olsen, R.: An Implementation of a Two Dimensional Kalman Filter for Atmospheric Chemical Consituent Assimilation on Massively Parallel Computers, Mon. Wea. Rev., 125, 1674–1686, 1997.

Mengistu Tsidu, G., Kiefer, M., von Clarmann, T., Fischer, H., Funke, B., Grabowski, U., Hase, F., Höpfner, M., López-Puertas, M., and Stiller, G. P.: Validation of MIPAS/ENVISAT Level-1B data products, edited by: Huang, H.-L., Lu, D., and Sasano, Y., Optical Remote Sensing of the Atmosphere and Clouds III, Proc. of SPIE, 4891, 483–496, 2003.

Migliorini, S., Piccolo, C., and Rodgers, C.: Intercomparison of direct and indirect measurements: Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) versus sonde ozone profiles, Journal Geophysical Research, 109(D19), D19 316, doi:10.1029/2004JDOO4988, 2004.

Pierce, R. B., Al-Saadi, J., Fairlie, T. D., Natarajan, M., Harvey, V. L., Grose, W. L., Russell, J. M., Bevilacqua, R., Eckermann, S. D., Fahey, D., Popp, P., Richard, E., Stimpfle, R., Toon, G. C., Webster, C. R., and Elkins, J.: Large-scale chemical evolution of the arctic vortex during the 1999–2000 winter: HALOE/POAM3 Lagrangian photochemical modelling for the SAGE III ozone loss and validation experiment (SOLVE) campaign, J. Geophys. Res., 108(D5), 8317, doi:10.1029/2001JD001063, 2003.

Planet, W. G., Miller, A. J., Horvath, K., Nagatani, R., Flynn, L., Hilsenrath, E., Kondragunta, S., Cebula, R. P., and DeLand, M. T.: Total ozone determinations from NOAA operational SBUV/2 observations: an update, J. Geophys. Res., 106(D15), 17 471–17 478, 2001.

Rabier, F., Thépaut, J. N., and Courtier, P.: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system, Quart. J. Roy. Meteorol. Soc., 124, 1861–1888, 1998.

Riishøjgaard, L. P.: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, Tellus, 50A, 42–57, 1998.

Riishøjgaard, L. P.: A method for estimating the analysis error variance in a physical space data assimilation system, Quart. J. Roy. Meteorol. Soc., 126, 1367–1385, 2000.

Riishøjgaard, L. P.: Deterministic prediction of the ensemble variance for a barotropic vorticity-equation model, Quart. J. Roy. Meteorol. Soc., 127, 1761–1773, 2001.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific Publishing, London, 2000.

Smit, H. and Kley, D.: JOSIE: The 1996 WMO International intercomparison of ozonesondes under quasi flight conditions in the environmental simulation chamber at Jülich, WMO/IGAC-Report, WMO Global Atmosphere Watch report series, No. 130 (Technical Document No. 926), World Meteorological Organization, Geneva, 1998.

Struthers, H., Brugge, R., Lahoz, W. A., O'Neill, A., and Swinbank, R.: Assimilation of ozone profiles and total column measurements into a global general circulation model, J. Geophys. Res., 107(D20), 4438, doi:10.1029/2001JD000957, 2002.

Talagrand, O.: Assimilation of Observations, an Introduction, J. Meteor. Soc. Japan, 75, 191–209, 1997.

Thomason, L. W. and Taha, G.: SAGE III Aerosol Extinction Measurements: Initial Results, Geophysical Research Letters, 30, 33–1 – 33–4, doi:10.1029/2003GL017 317, 2003.

Thuburn, J. and Li, Y.: Numerical simulations of Rossby-Haurwitz waves, Tellus, 52A, 181–189, 2000.

von Savigny, C., Haley, C. S., Sioris, C. E., McDade, I. C., Llewellyn, E. J., Degenstein, D., Evans, W. F. J., Gattinger, R. L., Griffoen, E., Lloyd, N., McConnell, J. C., McLinden, C. A., Murtagh, D. P., Solheim, B., and Strong, K.: Stratospheric Ozone Profiles Retrieved from Limb Scattered Sunlight Radiance Spectra Measured by the OSIRIS Instrument on the Odin Satellite, Geophys. Res. Lett., 30, 1755–1758, doi:10.1029/2002GL016401, 2003.