Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: A posteriori validation of error statistics in observation space

J. Schwinger[1] and H. Elbern[1,2]

[1] Chemical state analyses of the atmosphere based on data assimilation may be degraded by inconsistent covariances of background and observation errors. An efficient method to calculate consistency diagnostics for background and observation errors in observation space is applied to analyses of the four-dimensional variational stratospheric chemistry data assimilation system SACADA (Synoptic Analysis of Chemical Constituents by Advanced Data Assimilation). A background error covariance model for the assimilation of Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) ozone retrievals is set up and optimized. It is shown that a significant improvement of the assimilation system performance is attained through the use of this covariance model compared to a simple covariance formulation, which assumes background errors to be a fixed fraction of the field value. The forecast skill, measured by the distance between the model forecast and MIPAS observations, is shown to improve. Further, an evaluation of analyses with independent data from the Halogen Observation Experiment (HALOE), the Stratospheric Aerosol and Gas Experiment II (SAGE II), and ozone sondes reveals that the standard deviation of ozone analyses with respect to these instruments is reduced throughout the middle stratosphere. Compared to the impact of background error variances on analysis quality, it is found that the precise specification of spatial background error correlations appears to be less critical if observations are spatially and temporally dense. Results indicate that ozone forecast errors of a state of the art stratospheric chemistry assimilation system are of the same order of magnitude as MIPAS observation errors.

Citation: Schwinger, J., and H. Elbern (2010), Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: A posteriori validation of error statistics in observation space, *J. Geophys. Res.*, *115*, D18307, doi:10.1029/2009JD013115.

1. Introduction

[2] A posteriori validation of error statistics is an indispensable means in order to control and adjust the error covariances employed in data assimilation (DA) algorithms. This is particularly true for variational assimilation schemes, which owe their efficiency to the fact, that the background error covariance matrix (BECM) is prescribed a priori, and that no analysis error covariance matrix of the analyzed fields is provided.

[3] Methods for a posteriori validation of assimilation algorithms do not provide error covariances directly, rather they offer a set of diagnostic constraints. Hence, background and observation error covariances can be iteratively adapted until the diagnostic's result is in agreement with the theoretically expected value (this iterative process is referred to as "tuning" or "optimization" of error covariances in this text). The most popular among these a posteriori validation tests is the χ^2 criterion, which is being widely used in all kind of inverse problems [see e.g., *Rodgers*, 2000]. In atmospheric data assimilation applications *Ménard et al.* [2000a], *Khattatov et al.* [2000], and *Chipperfield et al.* [2002] tuned free parameters of the BECM parameterization in their Kalman filter variants by χ^2 validation. Also, the same criterion has been used to monitor changes in observation errors [*Stajner et al.*, 2004], or to asses the overall consistency of the specified error statistics [*Baier et al.*, 2005; *Elbern et al.*, 2007].

[4] However, for the purpose of a differentiated validation of covariances, the χ^2 test remains insufficient, as it does not distinguish between contributions of observation and background errors to the χ^2 value. To overcome this deficiency, *Desroziers and Ivanov* [2001], *Chapnik et al.* [2004], and *Chapnik et al.* [2006] have extended the χ^2 validation for a variational DA setting, in order to validate individual components of the objective function. They demonstrated that their method is well suited to correct observation error variances for subsets of mutually uncorrelated observations.

[1]Rhenish Institute for Environmental Research, University of Cologne, Cologne, Germany.
[2]Also at Institute for Chemistry and Dynamics of the Geosphere-2 (Troposphere), Jülich Research Centre, Jülich, Germany.

However, as traces of large matrices have to be estimated during the course of this procedure, the Desroziers and Ivanov method involves considerable additional algorithmic expenses in terms of implementation and computational cost.

[5] Building on works of *Hollingsworth and Lönnberg* [1986] and *Lönnberg and Hollingsworth* [1986], the examination of innovation (observation-minus-forecast) vector statistics has been used to determine the error statistics of short-range forecasts in meteorological applications [e.g., *Xu et al.*, 2001; *Daley*, 1991]. Because the innovations are calculated anyway by an assimilation system, this technique causes practically no additional computational cost. As a prerequisite for applicability, the Hollingsworth and Lönnberg method requires that observation errors be horizontally strictly uncorrelated, in order to separate the contribution of observation and background errors to the innovation statistics. Therefore, this method is not very well suited for data from space-borne remote sensing devices, as the (usually unknown) horizontal correlation of observation errors is likely to degrade the results.

[6] *Desroziers et al.* [2005] have devised an extended set of diagnostic quantities, based on the statistics of the innovation vector, the observation-minus-analysis vector, and the analysis-minus-background vector in observation space. In contrast to the Hollingsworth and Lönnberg method, separate diagnostics are available for background and observation errors. Furthermore, a formula for analysis error (in observation space) is given. Recently, these diagnostic tools have been used by *Bonavita et al.* [2008] and *Li et al.* [2009] to adjust error covariances in their ensemble Kalman filter applications.

[7] The objectives of this study are the evaluation of the *Desroziers et al.* [2005] method (hereafter referred to as Desroziers et al. method) in the context of four-dimensional variational (4D-var) chemistry data assimilation, and the application of this technique to the optimization of background error covariances of the SACADA assimilation system, which has been introduced in a recent paper by *Elbern et al.* [2010], hereafter referred to as ESB10. In contrast to ESB10, the present work focuses on the assimilation of ozone only. The Desroziers et al. method is applied to adjust spatially averaged background error variances in four height ranges for the assimilation of MIPAS ozone retrievals. This approach improves the representation of the BECM compared to ESB10, where the background error variances were taken to be a fixed fraction of the field value. It must be noted that the flow-dependent extension of the SACADA BECM parameterization, which was described in ESB10, is not considered here. In this work it is attempted to derive an optimal setting for the background error correlation length scale under the assumption of isotropy and homogeneity.

[8] This paper is organized as follows: In section 2 the theory underlying the diagnostic tools, which have been used to evaluate and tune the parameters of the SACADA BECM model, is briefly reviewed. Section 3 provides an overview of the SACADA assimilation system and the observational data used for this work. The tuning strategy pursued for this study as well as the practical application of the Desroziers et al. method are outlined in section 4. Assimilation results produced with the optimized error covariances are compared to analyses obtained with the untuned BECM model in section 5. The potential and restrictions of the method are finally discussed in section 6.

2. A Posteriori Validation of Error Covariances

2.1. Diagnosis of Error Statistics in Observation Space

[9] Let $\mathbf{y}$, $\mathbf{x}^b$ and $\mathbf{x}^a$ be the set of p observations, the forecast (background) state and the analyzed state, respectively. *Desroziers et al.* [2005] have shown, that the differences between the analysis and background fields, projected into observation space by the generally nonlinear operator H, and the observations

$$\mathbf{d}_b^o := \mathbf{y} - H(\mathbf{x}^b),$$
$$\mathbf{d}_b^a := H(\mathbf{x}^a) - H(\mathbf{x}^b),$$
$$\mathbf{d}_a^o := \mathbf{y} - H(\mathbf{x}^a),$$

can be used to diagnose the consistency of covariance matrices. Their approach is based on the fact that, assuming a perfect assimilation system, the following relations hold:

$$E\left[\mathbf{d}_b^a (\mathbf{d}_b^o)^T\right] = \mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T, \tag{1}$$

$$E\left[\mathbf{d}_a^o (\mathbf{d}_b^o)^T\right] = \widetilde{\mathbf{R}}, \tag{2}$$

$$E\left[\mathbf{d}_b^a (\mathbf{d}_a^o)^T\right] = \mathbf{H}\widetilde{\mathbf{A}}\mathbf{H}^T, \tag{3}$$

$$E\left[\mathbf{d}_b^o (\mathbf{d}_b^o)^T\right] = \mathbf{H}\widetilde{\mathbf{B}}\mathbf{H}^T + \widetilde{\mathbf{R}}, \tag{4}$$

Here, $E[\;]$ is the expectation operator, and $\widetilde{\mathbf{R}}$, $\widetilde{\mathbf{B}}$, $\widetilde{\mathbf{A}}$ are the observation, background, and analysis error covariance matrices, respectively. As the observation operator H may be nonlinear, matrix $\mathbf{H}$ is the Jacobian of H at the analysis point, i.e.,

$$\mathbf{H} := \left.\frac{\partial H}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^a}.$$

In practice, the covariances $\widetilde{\mathbf{R}}$ and particularly $\widetilde{\mathbf{B}}$ are not well known and approximations $\mathbf{R}$ and $\mathbf{B}$ must be used within the data assimilation algorithm. As a consequence of equations (1) to (4), a necessary condition for $\mathbf{B}$ and $\mathbf{R}$ to be correctly specified is given by

$$\mathbf{B} \approx \widetilde{\mathbf{B}} \quad \text{and} \quad \mathbf{R} \approx \widetilde{\mathbf{R}}. \tag{5}$$

In the following text, $\widetilde{\mathbf{R}}$, $\widetilde{\mathbf{B}}$ and $\widetilde{\mathbf{A}}$ are referred to as "diagnosed" error covariances, as they are evaluated a posteriori based on assimilation results. In contrast, $\mathbf{R}$ and $\mathbf{B}$ denote the error covariances used to obtain the analyses, which can be optimized in order to fulfill condition (5).

[10] It should be noted that equations (1) to (4) are based on linear estimation theory. For variational methods to be equivalent to linear estimation schemes, it is required that background and observation errors are Gaussian distributed [e.g., *Talagrand*, 1997]. Further, in the nonlinear case, it has to be assumed that $H(\mathbf{x}^a) - H(\mathbf{x}^b) \approx \mathbf{H}(\mathbf{x}^a - \mathbf{x}^b)$, that is, the

linear approximation is sufficiently valid within the range spanned by the statistically probable realizations of analysis increments $\delta \mathbf{x}^a = \mathbf{x}^a - \mathbf{x}^b$. As a necessary condition for the former prerequisite to be true, the observation-minus-background differences $\mathbf{d}_b^o$ must show a Gaussian distribution. This condition has been tested prior to application of the Desroziers et al. method in section 4.2.

[11] Equation (4) is the basis of the well-known method of Hollingsworth and Lönnberg [*Hollingsworth and Lönnberg*, 1986; *Lönnberg and Hollingsworth*, 1986]. In order to extract information about $\mathbf{B}$ and $\mathbf{R}$ from (4), observation errors must be strictly uncorrelated in space [see e.g., *Daley*, 1991], which is the case for data from a radiosonde network. In contrast, it is unlikely that observations from orbiting remote sensing instruments fulfill this requirement. Therefore, the diagnostics conducted in this work relies on (1) and (2).

2.1.1. Limitations of the Method

[12] If the covariance diagnostic outlined above, is conducted in the framework of a 4D-var assimilation system, the observation operator H contains a model integration and, consequently, the calculation of $\tilde{\mathbf{B}}$ from $\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T$ as well as the calculation of $\tilde{\mathbf{A}}$ from $\mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^T$ would require tangent linear and adjoint model integrations. Therefore, it is not feasible to separate $\tilde{\mathbf{B}}$ and $\tilde{\mathbf{A}}$ from $\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T$ and $\mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^T$, and the method remains restricted to observation space. However, it has been shown by *Cohn* [1993] that, given a passive tracer field, the background error variances evolve in time according to the model dynamics, that is, the variances of the tracer field are simply advected. This result provides the basis for background error covariance parameterizations in numerous Kalman filter applications [e.g., *Ménard et al.*, 2000b; *Khattatov et al.*, 2000; *Fierli et al.*, 2002; *Segers et al.*, 2005; *Rösevall et al.*, 2007]. In the framework of this study, the diagonal elements of $\tilde{\mathbf{B}}$ are approximated as mean values of diagonal elements of $\mathbf{H}\hat{\mathbf{B}}\mathbf{H}^T$ located in the same region. In view of the findings cited above, this approach can be considered valid, as long as the chemical lifetime of the assimilated trace gases is large compared to the time span defined by the 4D-var assimilation window. It should be noted that these constraints are likely to hamper the applicability of the Desroziers et al. covariance diagnostic in a 4D-var system for short-lived species like stratospheric NO_2.

[13] Further, it should be stressed at this point that equations (1) and (2) provide a consistency check only: Given a set of background, analysis, and observation values together with respective error estimates, the diagnosed quantities $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{B}}$ can be compared to $\mathbf{R}$ and $\mathbf{B}$. However, a good match of diagnosed and specified covariances is no more than a necessary condition for $\mathbf{R}$ and $\mathbf{B}$ to be correct. Further, although it is an obvious idea to tune the error covariances by iteratively replacing $\mathbf{R}$ and $\mathbf{B}$ by matrices that come closer to their diagnosed counterparts, there is neither a general proof of convergence, nor a proof that there exists a unique solution. *Desroziers et al.* [2005] demonstrated success in tuning background and observation error covariances at the same time using a synthetic toy problem. In practice, however, linear estimation theory is an approximation only (nonlinear, nonperfect model, deviations from perfect Gaussian error statistics, etc.), and the evaluation of equations (1) to (4) requires some simplifying assumptions.

In section 4.3 of this study, it is shown that under these conditions a unique solution for tuning $\mathbf{B}$ and $\mathbf{R}$ simultaneously might not be found.

[14] For the purpose of this study, it was therefore decided to restrict the problem to tuning $\mathbf{B}$ only, unless consistency with a given matrix $\mathbf{R}$ is obtained. Since sophisticated methods are employed to derive the retrieval error statistics of remote sensing data [*Rodgers*, 2000; *Raspollini et al.*, 2006], and because extensive validation activities have taken place in recent time (see *Cortesi et al.* [2007] for information on the validation of MIPAS O_3 retrievals), it is supposed that the observation error covariance matrix $\mathbf{R}$ is considerably better known than the background error of the newly developed SACADA assimilation system. $\mathbf{R}$ is therefore taken as fixed in this study. Equations (1) and (2) are applied to adapt $\mathbf{B}$ until equation (5) is satisfied.

2.1.2. Treatment of Systematic Retrieval Errors

[15] Errors in the retrieval of atmospheric constituents, are traditionally classified as systematic or random. The random error is due to the propagation of measurement noise through the retrieval process, while systematic errors originate from uncertain spectral data, calibration parameters, errors in radiative transfer modeling and other sources. The classification of errors as systematic or random is somewhat vague, because systematic errors may vary in time and space and an error component considered as systematic may appear random on a different spatial or temporal scale. A comprehensive discussion of this issue can be found in the textbook of *Rodgers* [2000].

[16] In most cases, a systematic error will cause a bias of the retrieval against the true atmospheric state, which may be constant or vary slowly in space and/or time. Random, zero-mean errors of background and observations, however, are a prerequisite for the applicability of the theory outlined above. Likewise, most data assimilation algorithms including 4D-var are designed to correct random, zero-mean errors in the background field based on observations with random, zero-mean errors. Therefore, an "optimal" analysis in principle requires a removal of observational and/or model bias prior to or during the assimilation procedure. Several methods for bias correction in data assimilation schemes have been proposed and tested in recent time [see *Dee*, 2005, and references therein].

[17] However, since a bias correction of MIPAS retrievals is beyond the scope of this study, it must be questioned under which conditions an evaluation of the SACADA assimilation system using the Desroziers et al. method can be useful. As shown by *Dee* [2005], 4D-var, like most current data assimilation algorithms is "bias blind", and biases either originating from the model or the observations are linearly transferred to the analysis. In an operational setting, the background for the next assimilation cycle is then computed starting from a biased analysis. Now, suppose that (1) time variation of systematic errors is slow compared to the timescale established by the assimilation window, that is, any measurement bias is nearly constant in time; (2) the spatial structure of systematic errors is smooth compared to structures induced by atmospheric chemistry and dynamics; and (3) the biased atmospheric state represented by the observations is not forbidden by chemical and dynamical constraints encoded in the model. Consequently, after a few assimilation cycles, the model forecast

(background) is forced to nearly reproduce the bias pattern of the assimilated measurements. Whether this is actually the case, can be revealed by inspection of observation-minus-background residuals. A zero expectation

$$E\left[\mathbf{d}_b^o\right] \approx 0 \qquad (6)$$

confirms that measurement bias is readily forecasted by the assimilating model. Note that these considerations do not apply if observations from more than one source are assimilated simultaneously. In this case systematic errors are likely to show different spatial distributions, prohibiting the adaption of the model forecast to any of these bias patterns.

[18] In conclusion, as long as additional information, that is independent measurements, is absent, and if condition (6) is fulfilled, a measurement bias cannot be detected within the data assimilation scheme and the system behaves as if unbiased observations were assimilated. Consequently, systematic errors must not be included in $\mathbf{R}$ in this case. The central prerequisite for doing so, that is, a (near) zero bias in observation-minus-background residuals, has been examined in this study prior to the application of the Desroziers et al. method (see section 4.2). Although the analysis obtained with this setup will be biased against the (unknown) true atmospheric state, a number of useful evaluation tests can be conducted without limitation. For example, the quality of a model forecast can be evaluated with respect to observations from the same instrument, and it can be tested whether a specific setting for error covariances improves or degrades the forecast skill. It must be noted, however, that some systematic error sources may contribute to the random error rather than resulting in a retrieval bias and, consequently, should be included in $\mathbf{R}$. This issue is further discussed in section 3.2.

2.2. χ^2 Diagnostics and Forecast Skill

[19] The innovation vector $\mathbf{d}_b^o$ is a Gaussian random vector with covariance $\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}$. Thus, if the covariances $\mathbf{R}$ and $\mathbf{B}$ are correctly specified, the quantity

$$z := \mathbf{d}_b^{oT}\left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}\mathbf{d}_b^o \qquad (7)$$

is χ^2 distributed with p degrees of freedom. Also, it is well known [see, e.g., *Talagrand*, 2003] that the cost function J at its minimum can be expressed as $J(\mathbf{x}^a) = z/2$. Therefore, $J_p^a := J(\mathbf{x}^a)/p$ is χ^2-distributed, satisfying

$$E\left[J_p^a\right] = 1/2. \qquad (8)$$

Criterion (8) can be used to survey the overall consistency of the specified error covariances.

[20] A further and independent means to asses the quality of covariance matrices a posteriori, is to compare different covariance settings with respect to the forecast skill of the assimilating model. The cost function value prior to assimilation (at the first iteration), normalized by the number of observations p

$$J_p^f := J(\mathbf{x}^b)/p = \frac{1}{2p}\left[\mathbf{y} - H(\mathbf{x}^b)\right]^T \mathbf{R}^{-1}\left[\mathbf{y} - H(\mathbf{x}^b)\right] \qquad (9)$$

is a measure of forecast quality with respect to observations that are about to be assimilated. A well-behaved data assimilation system will allow for improving forecasts the closer $\mathbf{B}$ and $\mathbf{R}$ represent the true covariances.

3. Assimilation System and Observational Data

3.1. SACADA Assimilation System

[21] SACADA is a 4D-var assimilation system for stratospheric trace gas observations. Here, only a brief overview of the system will be given, for a comprehensive description and for first evaluation results the reader is referred to ESB10. A stratospheric global chemistry transport model (CTM) and its adjoint version is kernel of the new SACADA system. The German Weather Service's global forecast model GME [*Majewski et al.*, 2001] serves as the meteorological driver for the CTM, providing meteorological fields at the same temporal and spatial locations where they are needed by the chemistry module. GME utilizes an icosahedral grid structure and a semi Lagrange scheme for horizontal advection, which are adopted for the SACADA CTM. The stratospheric chemistry module accounts for 167 gas phase and 10 heterogeneous reactions involving 48 stratospheric constituents. Background error covariances are parameterized using a diffusion approach devised by *Weaver and Courtier* [2001]. A quasi-newton limited memory Broyden Fletcher Golfarb Shanno (L-BFGS) algorithm [*Nocedal*, 1980; *Liu and Nocedal*, 1989] is employed for the minimization of the cost function. The general assimilation system setup was the same as in ESB10: The meteorological simulation is initialized at the start of each 24 h assimilation window with fields from European Centre for Medium-Range Weather Forecasts (ECMWF) analyses. The analyzed chemical fields from the end of one assimilation window are used as the background for the next assimilation cycle.

3.2. MIPAS Data and Error Characterization

[22] MIPAS V4.61 offline level 2 data have been assimilated in this study. The MIPAS instrument is a Fourier transform spectrometer on board of the European Space Agency's (ESA) Environmental Satellite (EnviSat). EnviSat has been placed in a Sun-synchronous polar orbit with an inclination of 98.55°. The orbit is almost circular at about 800 km altitude resulting in a total of 14.3 orbits that are performed each day. MIPAS measures high-resolution emission spectra in the midinfrared from 4.1 to 14.7 μm wavelength in a limb viewing mode. A single limb scan covers an altitude range from approximately 6 to 68 km by 17 steps. A detailed description of the instrument design can be found in the work of *Fischer and Oelhaf* [1996].

[23] *Raspollini et al.* [2006] provide a comprehensive description of the ESA offline data processing along with its error budget. The data product contains estimates of the instrument noise (random) error covariance matrix for each retrieved MIPAS profile. For the purpose of this study, error covariances within the MIPAS profiles have been neglected ($\mathbf{R}$ was taken to be diagonal). MIPAS spatial resolution is estimated to be $3 \times 30 \times 300$ km in the vertical direction, perpendicular to, and along the line of sight, respectively [e.g., *Fischer et al.*, 2008]. This spatial resolution is comparable to the SACADA model grid resolution, which, in

the current model setup, is ≈250 km in the horizontal and ≈2 km in the vertical direction. Therefore, it can be expected that MIPAS observations are largely representative for the atmospheric state within a single model grid cell, and an error of representativeness was not included in **R**. Errors arising due to the interpolation of modeled ozone to the location of observation have been neglected. The impact of these simplifications is assumed to be small compared with other uncertainties in the observation error statistics. The given standard deviations σ^o have been increased in some cases, as unrealistic small error estimates severely hamper the performance of the minimization algorithm. Therefore, a minimum absolute error margin of 0.08 parts per million (ppm) and a minimum relative error of 5% have been prescribed for MIPAS observational errors.

[24] Systematic error estimates calculated according to *Dudhia et al.* [2002; see also *Raspollini et al.*, 2006] are available for five generic atmospheric conditions (data can be obtained from http://www-atm.physics.ox.ac.uk/group/mipas/err/). There exists one major systematic error source, the mapping of the pressure-temperature retrieval random error into the volume mixing ratio retrieval (referred to as "pT error" hereafter), which is uncorrelated between any two ozone profiles and should be included in **R**. The magnitude of this error is estimated to be 4% to 6% of the retrieved ozone value in the altitude range of interest. Tuning experiments without and with a 5% pT error contribution added to **R** have been conducted for this study. As the bias of observation-minus-background residuals remains small (see below), other systematic error sources have not been included in **R** following the rationale given in section 2.1.2.

3.3. Validation Data

[25] HALOE, SAGE II, and ozone sonde measurements have been used as independent (not assimilated) validation data for the BECM tuning experiments performed in this study. Both, HALOE and SAGE II, are solar occultation type spectrometers [e.g., *Nazaryan et al.*, 2005], which have been operational from 1991 to 2005, and from 1984 to 2005, respectively. The SAGE II instrument was carried by the Earth Radiation Budget Satellite (ERBS) spacecraft and operated in the UV, visible and near-infrared (385–1020 nm) wavelength region. Due to the orbit geometry, tangent point locations vary slowly from 70°N to 70°S within approximately one month. In this study data product version 6.2 has been employed, which, for ozone retrievals, is a minor update to version 6.1 [*Wang et al.*, 2002]. The HALOE sensor [*Russell et al.*, 1993] on board the Upper Atmosphere Research Satellite (UARS) recorded spectra in the infrared wavelength region of about 30 occultation events per day. As in the case of SAGE II, the latitude of tangent point location changes slowly from day to day, covering a range of 80°N to 80°S within approximately one month. The latest data product release (version 19) has been used for validation. Since the retrieval grids of HALOE and SAGE II data products have a finer vertical spacing than the SACADA model grid, data points within one model layer are averaged before they are used for validation of analyzed ozone.

[26] Finally, ozone sonde profiles have been obtained from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC, http://www.woudc.org/) for validation of

the BECM tuning procedure. Since ozone sondes offer a considerably higher vertical resolution than the SACADA model, sonde measurements within each model layer have been averaged prior to comparison with analyses.

4. Tuning of Background Error Covariances

4.1. Tuning Strategy

[27] The background error covariance matrix **B** can be decomposed into a correlation matrix **C** and a diagonal matrix $\mathbf{\Sigma}$: $= \mathrm{diag}(\sigma_1^b,\ldots,\sigma_n^b)$ containing the background error standard deviations:

$$\mathbf{B} = \mathbf{\Sigma}\mathbf{C}\mathbf{\Sigma}.$$

Hence, the problem of finding a suitable BECM formulation can be split into two separate tasks, namely, to specify correct standard deviation values σ_k^b for each model grid point $k = 1, \ldots, n$, and to find an appropriate spatial correlation structure to be encoded in the parameterization of **C**. Although the diffusion approach implemented in the SACADA assimilation system allows for a fully flexible anisotropic and inhomogeneous specification of background error correlations, it was decided, as a starting point for the current study, to use a relatively simple correlation structure, defined by a global horizontal and a vertical correlation length scale L_h and L_v, respectively. Furthermore, as the focus of this study has been placed on horizontal background error correlations, the parameter L_v has been kept fixed at a value of 1.5 km, which is the order of magnitude of the vertical grid spacing. This value has been chosen ad hoc in order to provide a weak smoothing in the vertical direction. For convenience, the subscript h is omitted in the following text and the parameter L denotes the horizontal correlation length scale of background errors.

[28] The tuning strategy pursued in this study is as follows: After a spin-up assimilation of three days (employing coarse estimates of $\mathbf{\Sigma}$ and L), the first step is to use equations (1) and (2) in order to adjust σ_k^b iteratively (see section 4.3 for details). After applying this procedure, the error standard deviations σ^b and σ^o are consistent, that is, condition (5) is fulfilled for mean diagonal elements of **R** and **B**. In order to avoid degradation of the results due to model errors and/or misspecification of the correlation length scale L (which is still untuned at this stage), the tuning is performed on a subset of observations only. This issue is extensively discussed in section 4.3.

[29] In a second step, a suitable correlation length scale L for the parameterization of **C** is estimated based on (1). To this end, a set of assimilation experiments has been conducted, using the tuned background variances, but different settings of L. Again, a special subset of observations, as detailed in section 4.5, has to be chosen in order to obtain meaningful results. The χ^2 criterion and an inspection of forecast skill are used to validate the optimized length scale.

[30] The assumption implicit to the strategy of tuning L independently of $\mathbf{\Sigma}$ is that the choice of L has a rather small influence on the model forecast (the background field, or more technically on $\mathbf{d}_b^o$) in comparison to $\mathbf{\Sigma}$. It is demonstrated in sections 4.3 and 4.7 that this assumption is justified for a range of reasonable values of L. It should be noted, that this statement apparently somewhat contradicts

Table 1. Definition of Model Layers for Tuning of Background Error Standard Deviations

Index i	Model Levels	Height Range (km)	Pressure Range (hPa)
1	25–28	16.5–11.2	96.1–221
2	20–24	23.9–16.5	30.0–96.1
3	15–19	32.4–23.9	8.3–30.0
4	10–14	41.3–32.4	2.4–8.3

the premise that the BECM plays a major role in obtaining meaningful analyses in data sparse regions. In fact, the technique applied here for the tuning of L in model space from data available in observation space, is feasible only if observational data is sufficiently dense. In the case of MIPAS data, where the distance between profiles is about 550 km along the orbit track, it seems to be possible to extract information on correlation length scales larger than ≈ 400 km, as discussed in section 4.5. If data is significantly sparser than the length scale established by L, a tuning of L based on the theory outlined above is not possible.

4.2. Experimental Setup

[31] A period of three weeks in 2003 (21 October to 10 November) was chosen for performing the tuning experiments. Note that this time span is also covered by the second case study presented in ESB10. Data coverage from the MIPAS sensor was dense within this time, and a large number of profiles from HALOE and SAGE II are available for validation. Although the BECM parameters derived in this study are possibly time dependent, the general results are not likely to be affected by the choice of a particular time interval. In order to test the capabilities of the method, the assimilation has been confined to ozone data only, and the model domain was restricted to a range of 221 to 2.4 hPa (approximately 11 to 41 km height). It must be noted that this pressure range contains the upper troposphere in the tropics. It is known that the quality of MIPAS data degrades in the tropopause region and below [e.g., *Cortesi et al.*, 2007], and also the SACADA model has been set up for stratospheric data assimilation. However, as tropospheric MIPAS observations account for only 2% of the absolute number of assimilated MIPAS measurements, it is not expected that the limitations of model and observations in the troposphere will significantly affect the global assimilation results.

[32] For the purpose of this study, the model domain was divided into four layers, as defined in Table 1, each layer comprising 4 to 5 model levels or a height range of approximately 8 km. A five-parameter background error covariance model was adjusted to the diagnosed background errors using the approach outlined above. The five parameters were taken to be the background error standard deviations σ_i^b, $i = 1, \ldots, 4$ at the center of the four layers and one correlation length scale L assumed to be valid for the whole model domain. To obtain background errors for an arbitrary model level, the SACADA BECM module linearly interpolates the enclosing values.

[33] An assimilation run over the whole three week period with untuned BECM parameters, using the coarse estimate of 30% relative background error and assuming a background error correlation length scale of $L = 600$ km, was conducted to serve as a control run. The distribution of control run observation-minus-background differences $\mathbf{d}_b^o$, compiled for the four altitude intervals $i = 1, \ldots, 4$ is shown in Figure 1. The $\mathbf{d}_b^o$ distributions are approximately Gaussian, except for the lowermost layer, where an excess of small difference values can be observed. This excess can be explained by the contribution of tropospheric O_3 observations, which account for 16% of the total number of measurements in this layer. The $\mathbf{d}_b^o$ distributions do not show a significant bias (<0.023 ppm), except for the uppermost layer, where a bias of ≈ 0.1 ppm is found.

[34] It could be argued that an inspection of, for example, observation-minus-background differences (Figure 1) reveals without any additional diagnostics, that a 30% background error is too large (30% error amounts to about 3 ppm in absolute values in the region of the ozone volume mixing ratio maximum). In order to evaluate the impact of an optimized BECM compared to a simple but more plausible estimate of background errors, a second control run, which switches the relative background error from 30% to 10% after a spin-up assimilation of three days, was conducted. For convenience, the two control runs employing 30% and 10% relative background error are referred to as CTR30 and CTR10, hereafter.
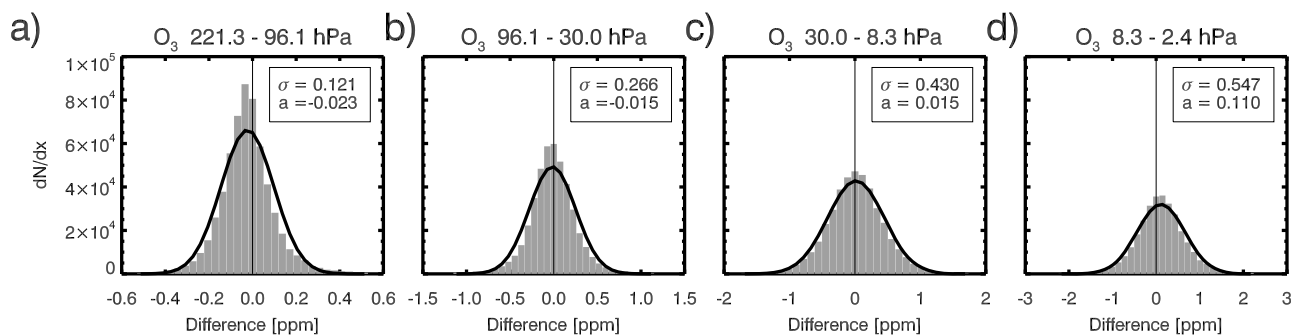

Figure 1. Distribution of observation-minus-background differences obtained from the untuned assimilation system. The differences are sampled over the time interval 24 October to 10 November 2003. A Gaussian fit to the data is given by the solid curve, with the parameters (standard deviation σ and mean a) of the fit indicated in the box at the top right of each plot.
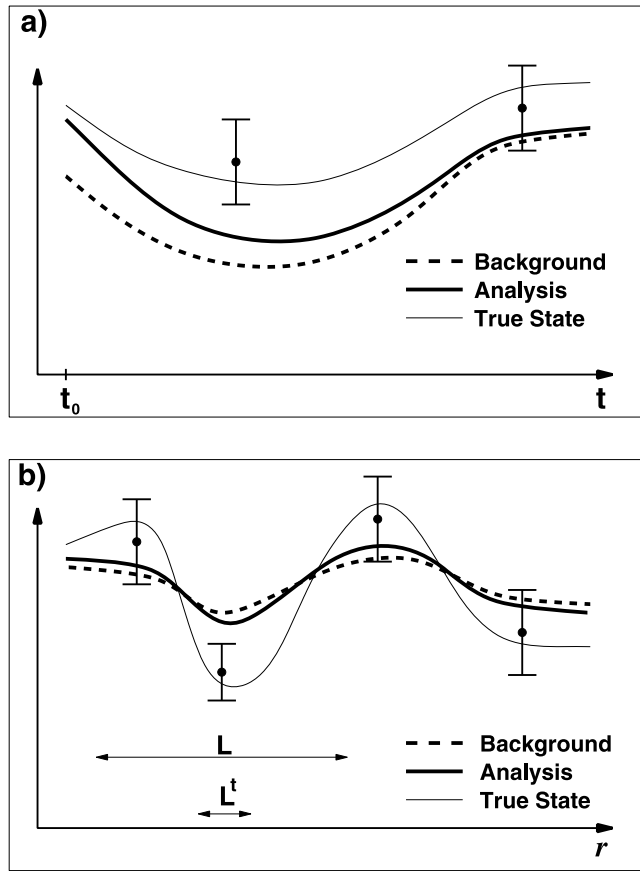
Figure 2. Illustration of two specific cases for which the straightforward application of the *Desroziers et al.* [2005] method leads to an underestimation of background errors and overestimation of observation errors. (a) Due to model error and (b) due to misspecification of the correlation length scale. See text for further explanation.

4.3. Tuning of σ^{b}

[35] The tuning of error standard deviations σ_i^{b} was performed for 24 October 2003, taking the first three days of the CTR30 control run as a spin-up assimilation. Following *Desroziers et al.* [2005], the diagnosed values of background and observation error variances can be estimated based on equations (1) and (2) by

$$\left(\tilde{\sigma}_i^{\mathrm{b}}\right)^2 = \frac{1}{p_i}\left(\mathbf{d}_{\mathrm{b}}^{\mathrm{a}}\right)_i^{\mathrm{T}}\left(\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}\right)_i = \frac{1}{p_i}\sum_{j=1}^{p_i}\left(y_j^{\mathrm{a}} - y_j^{\mathrm{b}}\right)\left(y_j^{\mathrm{o}} - y_j^{\mathrm{b}}\right) \qquad (10)$$

$$\left(\tilde{\sigma}_i^{\mathrm{o}}\right)^2 = \frac{1}{p_i}\left(\mathbf{d}_{\mathrm{a}}^{\mathrm{o}}\right)_i^{\mathrm{T}}\left(\mathbf{d}_{\mathrm{b}}^{\mathrm{o}}\right)_i = \frac{1}{p_i}\sum_{j=1}^{p_i}\left(y_j^{\mathrm{o}} - y_j^{\mathrm{a}}\right)\left(y_j^{\mathrm{o}} - y_j^{\mathrm{b}}\right) \qquad (11)$$

for a subset of observations i with a number of p_i observations. Here, y_j^{o} is the jth observation in subset i, whereas y_j^{a} and y_j^{b} are the corresponding analysis and background counterparts, respectively. The subsets i were taken to comprise all available observations within the model subdomain (altitude interval) i defined above.

[36] However, it turned out that the straightforward application of (10) leads to a significant underestimation of background errors. In order to avoid this fault, it is necessary to understand how equations (10) and (11) interact: For

given background and observation values, the term $\left(y_j^{\mathrm{o}} - y_j^{\mathrm{b}}\right)$ remains fixed while $\left(y_j^{\mathrm{a}} - y_j^{\mathrm{b}}\right)$ determines the contribution of the jth summand to $\left(\tilde{\sigma}_i^{\mathrm{b}}\right)^2$. If, on average, $\left(y_j^{\mathrm{a}} - y_j^{\mathrm{b}}\right)$ is large, then the value of $\tilde{\sigma}_i^{\mathrm{b}}$ is diagnosed to be large. Conversely, if the analysis stays close to the background, a small background error is diagnosed. If applied in a real world 4D-var assimilation system, the term $\left(y_j^{\mathrm{a}} - y_j^{\mathrm{b}}\right)$ can remain small due to the following reasons, which are beyond the specification of a proper background error:

[37] 1. Violation of the perfect model assumption: The model may not be able to reproduce some observations within the assimilation window properly due to imperfect model physics or due to model resolution. Hence, the analysis y^{a} often stays closer to y^{b} than appropriate, that is, closer to y^{b} than it would do if the assimilating model was perfect. A specific example for a kind of model error, which is likely to occur in chemical modeling is depicted in Figure 2a: Here, relaxation toward a chemical equilibrium different from the true equilibrium state leads to an analysis that approaches the background state with increasing time.

[38] 2. Background error correlations: Assume that observation j has several neighboring observations within a distance r. If the true correlation length L^{t} is small ($L^{\mathrm{t}} \ll r$), but the correlation length scale L, which is employed during the assimilation procedure, is wrongly specified to be large, i.e., $L \geq r$, the assimilation algorithm will smooth out spatial variability and only small analysis increments will emerge. Consequently, the analysis will stay close to the background state, as illustrated in Figure 2b.

[39] These considerations are not only of theoretical nature, but have a strong influence on the diagnosed background errors. As an example for the first case, Figure 3 shows profiles of diagnosed background errors calculated based on $\mathbf{y}^{\mathrm{b}}$, $\mathbf{y}^{\mathrm{a}}$ and $\mathbf{y}^{\mathrm{o}}$ vectors from different time slices within the 24 h assimilation window on 24 October 2003. The growing influence of model error leads to smaller $\tilde{\sigma}^{\mathrm{b}}$
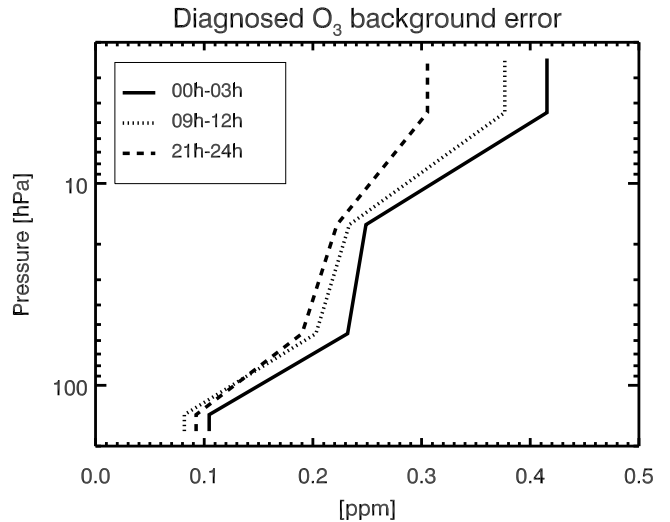

Figure 3. Diagnosed background error for different time slices within the 24 h assimilation window (solid, 0–3 h; dotted, 9–12 h; dashed, 21–24 h). Line styles for the different time intervals are also indicated in the box at the top left. All available MIPAS data on 24 October 2003 were included in these calculations.
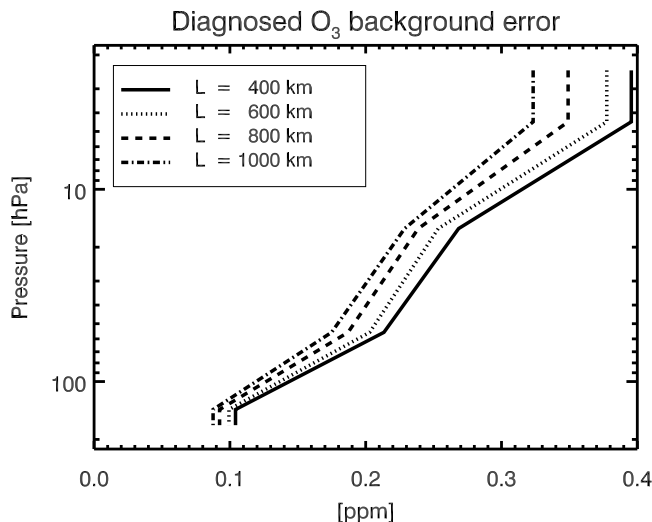
Figure 4. Diagnosed background error for different background error correlation length scales L (solid, $L = 400$ km; dotted, $L = 600$ km; dashed, $L = 800$ km; dash-dotted, $L = 1000$ km). The line styles for different values of L are also indicated in the box at the top left. All available MIPAS data on 24 October 2003 were included in these calculations.

values the farther the time interval is chosen from the model start time. In Figure 4 profiles of $\tilde{\sigma}^b$ obtained by assimilating all available MIPAS data for the same date are shown for different values of L. It can clearly be seen, that $\tilde{\sigma}^b$ is considerably smaller for large L.

[40] In oder to obtain a realistic estimate of σ^b, it was decided to perform the diagnostic (10) and (11) on a subset of observations only. This subset was chosen to be every second MIPAS profile ("thinning" of observations) [e.g.,

Dando et al., 2007] within the time interval 00:00 UTC–06:00 UTC on 24 October 2003. This choice ensures that underestimation of $\tilde{\sigma}^b$ due to the implications 1. and 2. given above is largely excluded. Furthermore, the consequences arising from the approximation that the diagonal elements of $\tilde{\mathbf{B}}$ are identified with the diagonal elements of $\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^T$ are reduced the less the analyzed observations are scattered in time (i.e., the shorter the time window for the tuning process is chosen). The geolocations of the 135 profiles selected for the tuning process are shown in Figure 5. Tuning experiments for two different settings of $\mathbf{R}$ have been performed. The first experiment was conducted with observation error standard deviations directly taken from the ESA data product files, without accounting for any systematic error sources. For the second experiment the observation errors were modified to account for an 5% contribution from the pT error (see section 3.2).

4.3.1. Tuning Experiment 1

[41] The tuning process was started by applying the covariance diagnostics to the untuned assimilation system, where a relative background error of 30% had been assumed. As shown in Figure 6a the values of the diagnosed quantities $\tilde{\sigma}^o$ and $\tilde{\sigma}^b$ significantly differ from σ^o and the first-guess background standard deviation σ^b. Particularly, σ^b and $\tilde{\sigma}^b$ show differences of up to 1.5 ppm, revealing that the coarse estimate of a 30% relative background error used for the CTR30 run is by far too large. Also shown in Figure 6b is the consistency diagnostics for a 10% relative background error, which was assumed subsequently for the CTR10 run. There is a very close agreement of $\tilde{\sigma}^b$ and σ^b as well as of $\tilde{\sigma}^o$ and σ^o up to about 50 hPa. Larger deviations of between $\tilde{\sigma}^b$ and σ^b up to 0.3 ppm occur above 50 hPa.

[42] Improved estimates of σ^b were then constructed by inserting the mean value of diagnosed and specified background error standard deviations and recomputing the analysis. This procedure was repeated until a satisfying
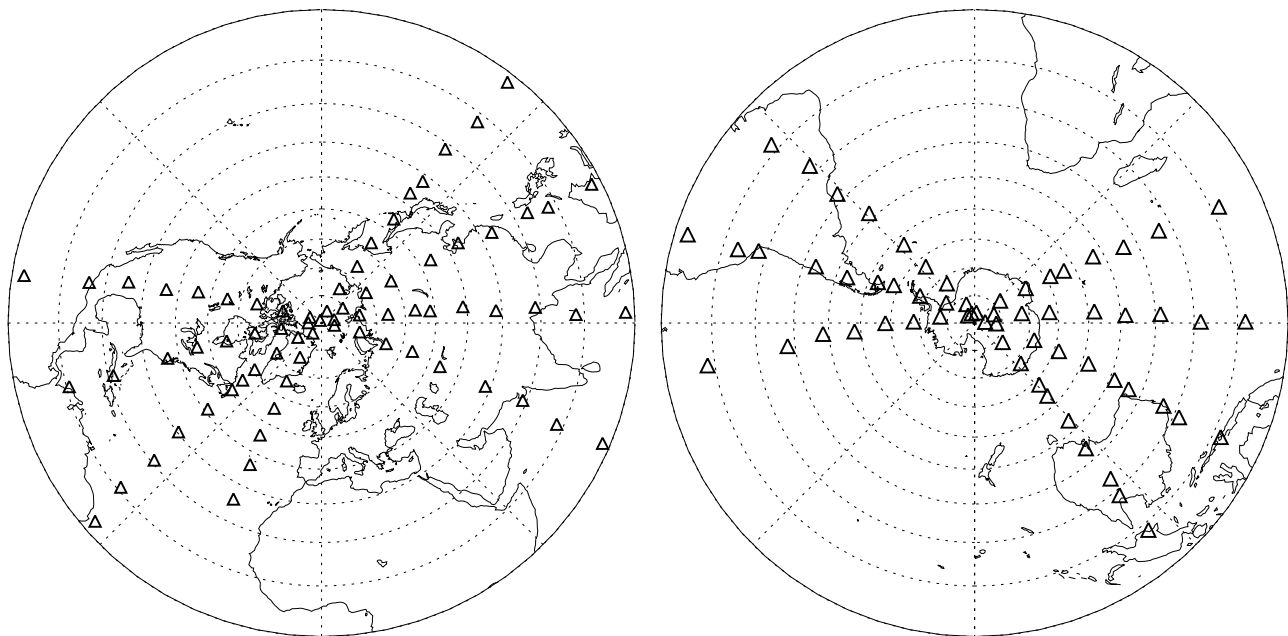

Figure 5. Geolocation of the 135 MIPAS profiles (triangles) selected for the tuning of background error covariances on 24 October 2003.
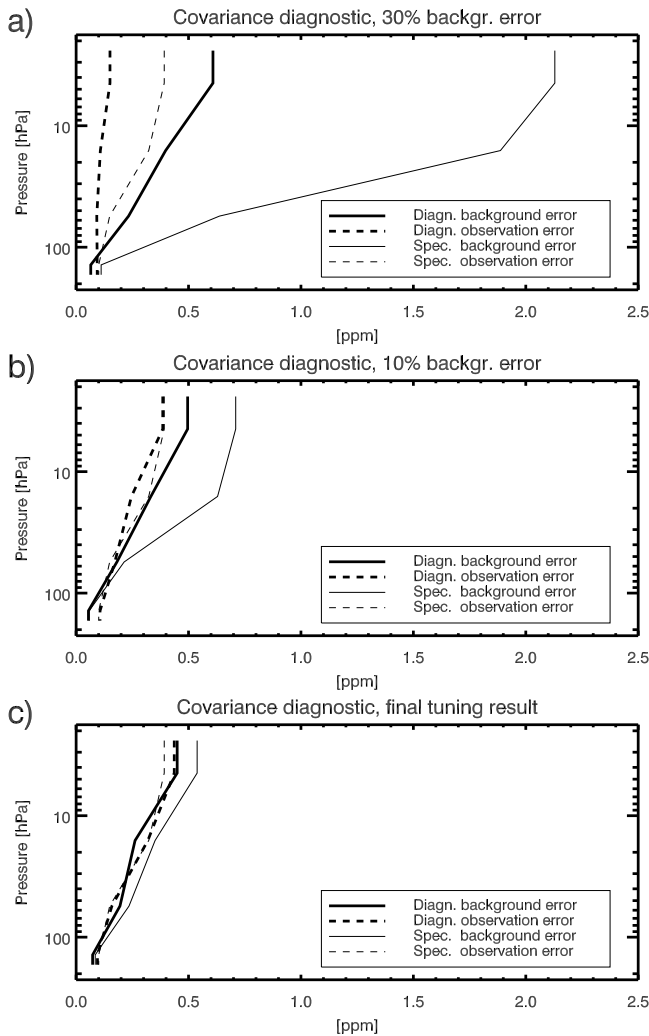
a)

b)

c)

Figure 6. Covariance diagnostics using a subset of MIPAS O_3 profiles only, tuning experiment 1. Shown are the specified background errors σ^b (thin solid line), diagnosed background errors $\tilde{\sigma}^b$ (thick solid line), specified observation errors σ^o (thin dashed line), and diagnosed observation errors $\tilde{\sigma}^o$ (thick dashed line) for (a) untuned background error variances assuming a 30% relative background error, (b) untuned background error variances assuming a 10% relative background error, and (c) the final tuning result. Line styles are also indicated in the box at the bottom right of each plot.

consistency between specified and diagnosed background error standard deviations was obtained ($\tilde{\sigma}^b \approx \sigma^b$) after three to four iterations. In this study no formal convergence criterion was applied. Development of such criteria as well as automatization of the described tuning procedure for operational application is ongoing work.

[43] The final result of the tuning procedure is presented in Figure 6c, and the values of tuned background error standard deviations can be found in Table 2 (third column). Compared to the untuned CTR30 (CTR10) setting, the maximum difference between $\tilde{\sigma}^b$ and σ^b shrank by a factor of 30 (6) to about 0.05 ppm. This small inconsistency between diagnosed and specified error standard deviations is

still present in the final tuning result, and the associated analysis cost function value $J_p^a = 0.55$, which is 10% larger than the expected value of 0.5, reflects this inconsistency. In the first place this is certainly caused by the choice of the relatively simple background error covariance model, which does not allow for latitudinal variations of σ^b. Furthermore, a closer inspection of the MIPAS data revealed that about 5% of the assimilated profiles show strong oscillations in the region above approximately 20 hPa. An example is shown in Figure 8. It can be speculated that these kind of oscillations are artifacts originating from the ESA retrieval algorithm due to the fact that no regularization has been applied [*Raspollini et al.*, 2006]. If these profiles are sorted out manually, the tuning result further improves (not shown) and the associated J_p^a value reduces to 0.52. However, the development of objective criteria for the decision to reject a certain profile, which probably shows unphysical behavior or not (i.e., stricter quality control criteria) is beyond the scope of this study. Consequently, no additional selection of MIPAS profiles took place for the subsequent assimilation experiments.

[44] To check whether the tuned background error variances remain valid for dates other than 24 October, the tuning procedure was repeated for 27 October and 6 November. The differences found in the tuning results did not exceed 0.04 ppm, which is within the range of the aforementioned remaining inconsistency between $\tilde{\sigma}^b$ and σ^b. Therefore, it can be hoped that the background error variances vary slowly enough to use the tuning result obtained at a specific date for a period of a few weeks at least.

[45] Further, it has been verified that the tuning result does not depend notably on the value of L used for the spin-up assimilation (21 to 23 October). To this end, the spin-up was repeated using $L = 400$ km, $L = 800$ km, $L = 1000$ km and $L = 1400$ km, but the differences in diagnosed background error variances $\tilde{\sigma}^b$ relative to the $L = 600$ km setting were found to be small (<0.025 ppm, not shown). At this point, it is important to recognize the different impact of L on the background field on one hand and on the tuning procedure on the other hand: The latter has been shown to be substantial, with smaller background error estimates emerging for larger values of L (see Figure 4). The model forecast (background), however, seems to depend only weakly on the exact value of L, as demonstrated by the small differences in the tuning result due to background fields obtained with different settings of L in the preceding spin-up assimilation. This constitutes a first justification for the strategy of tuning σ^b independently of L, as discussed above in section 4.1. A more detailed evaluation of forecast skill and its dependence on L is presented in section 4.7.

Table 2. Values of Background Error Standard Deviations Obtained With the Tuning Procedure and Increased Values Used for Assimilation Over a 24 h Assimilation Window

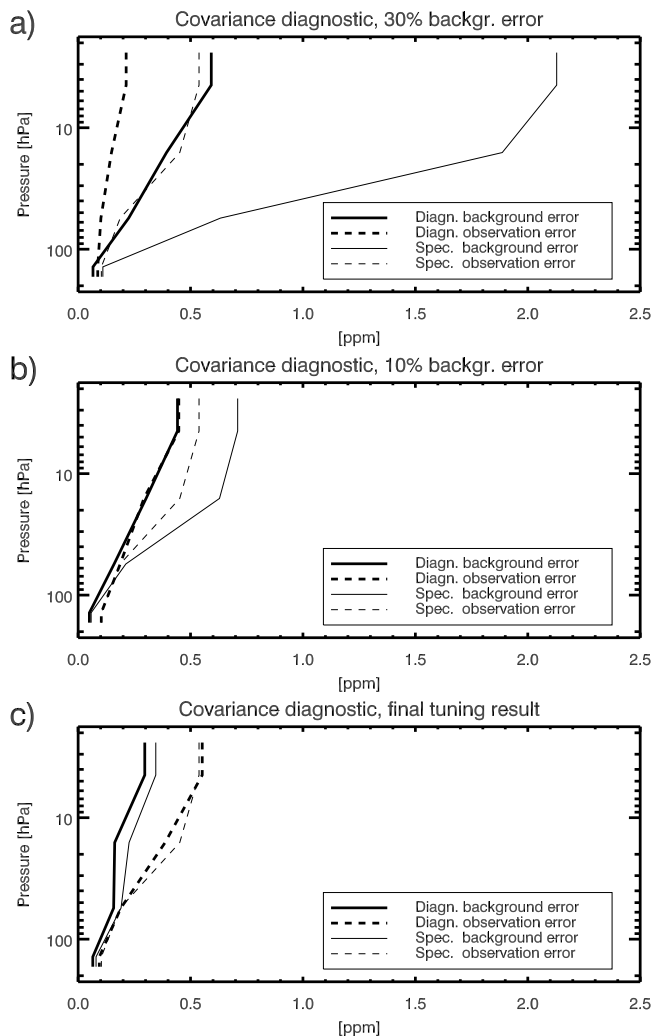| Index i	Height (km)	Final Tuning Result σ_i^b (ppm)	Values $f\sigma_i^b$ Used for 24 h Assimilation Window (ppm)
1	13.9	0.055	0.06875
2	20.2	0.23	0.2875
3	28.1	0.32	0.40
4	36.9	0.56	0.70

a)

b)

c)

Figure 7. As in Figure 6, but for tuning experiment 2, with a 5% contribution of systematic error sources added to **R**.

4.3.2. Tuning Experiment 2

[46] The second tuning experiment has been conducted considering a 5% pT error component included in **R**. As shown in Figures 7a and 7b, the assumption of an 30% or 10% relative background error leads, similar to experiment 1, to large deviation between diagnosed and specified quantities. Interestingly, it is possible to obtain a tuning result with a remaining inconsistency < 0.05 ppm, which is not larger than for tuning experiment 1 (Figure 7c). However, due to the increased observation error margins, the diagnosed background error is significantly smaller than the MIPAS observation error above 60 hPa in this case.

[47] As a conclusion, it appears that, in practice, the Desroziers et al. diagnostics allows for more than one consistent specification of error covariances, or possibly a range of consistent solutions. To decide which tuning result is more realistic, additional diagnostic is necessary. For example, it was found that the condition $J_p^a = 0.5$ cannot be met with the 5% pT error added to **R**. Further, the forecast skill did not improve significantly relative to the case where a 30% background error is assumed (not shown). These findings point toward the conclusion that the background

errors obtained in tuning experiment 2 are unrealistically small.

[48] Finally, it must be noted, that the assumption of a global 5% pT error is a simplification. In principle, it is possible to calculate the pT error component for individual profiles [see *Raspollini et al.*, 2006], and a refinement of background error estimates might be a possible outcome. However, for this first evaluation of the SACADA BECM formulation, it was decided not to consider the pT error contribution to the total random error for the remainder of this study. This can bee seen as a precautionary measure in order to avoid an underestimation of the model forecast (background) error.

4.3.3. Extension to a 24 h Assimilation Window

[49] The background error standard deviations σ^b derived above are consistent for a short-assimilation ("tuning") window of 6 h. It is assumed that the perfect model assumption inherent to the strong constraint 4D-var method is valid to a reasonable degree over this short time span. If it is intended to use a longer-assimilation window of 24 h, additional considerations are necessary: A weak constraint 4D-var system [e.g., *Trémolet*, 2006] would, as it takes into account the presence of model errors, assign less weight to the model trajectory relative to the observations, thereby avoiding that the optimal solution sticks too close to the background. In a strong constraint data assimilation framework the only possibility to give less weight to the model trajectory is to increase the background error instead. Therefore, it was decided to apply a multiplicative inflation factor f to the tuned σ^b for use over a 24 h assimilation window. In order to find a suitable value for f, some experimentation was necessary: Assimilation runs over the case study period with $f = 1.10$, $f = 1.20$, $f = 1.30$, and finally $f = 1.25$ were conducted, and J_p^a was examined as described in section 4.6. Fortunately, since the diagnostics conducted to find an optimal correlation length scale L was
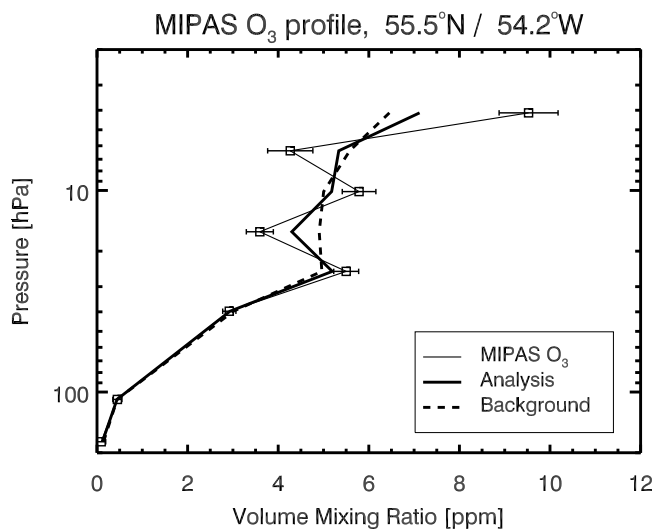

Figure 8. Example of one MIPAS profile (squares with error bars, connected with thin solid line) showing oscillations at altitudes above approximately 20 hPa. The background and analysis are given by the thick dashed and solid line, respectively. Line styles are also indicated in the box at the bottom right.
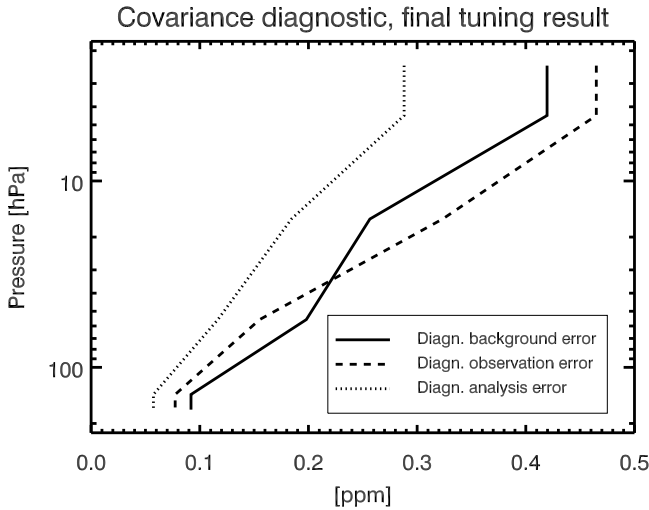
Covariance diagnostic, final tuning result

Figure 9. Diagnosed analysis error in observation space $\tilde{\sigma}^{a}$ (dotted), diagnosed background error $\tilde{\sigma}^{b}$ (solid), and the diagnosed observation error $\tilde{\sigma}^{o}$ (dashed) for the final tuning result. Line styles are also indicated in the box at the bottom right.

found to be not significantly influenced by relatively small changes in background error specification, it was possible to perform a preliminary diagnostic of L first. For $L = 600$ km, which turned out to be apparently the most consistent choice (see sections 4.5 and 4.7), a value of $f = 1.25$ was determined such that $J_{p}^{a} \approx 0.5$. The final tuning result σ^{b} (6 h tuning window) and the values used for subsequent assimilation experiments $f\sigma^{b}$ are listed in Table 2.

4.4. Diagnosed Analysis Error

[50] Given that $\mathbf{B} \approx \tilde{\mathbf{B}}$ and $\mathbf{R} \approx \tilde{\mathbf{R}}$, a diagnosed analysis error valid in observation space can be approximated based on equation (3) by

$$\left(\tilde{\sigma}_{i}^{a}\right)^{2} = \frac{1}{p_{i}}\left(\mathbf{d}_{b}^{a}\right)_{i}^{T}\left(\mathbf{d}_{a}^{o}\right)_{i} = \frac{1}{p_{i}}\sum_{j=1}^{p_{i}}\left(y_{j}^{a} - y_{j}^{b}\right)\left(y_{j}^{o} - y_{j}^{a}\right), \qquad (12)$$

for the same subset of observations i as used in (10) and (11). The analysis error calculated by means of (12) is shown together with the diagnosed background and observation errors in Figure 9. It should be emphasized, however, that this approximation is valid only at the location of observations, as it estimates the diagonal elements of $\mathbf{H}\tilde{\mathbf{A}}\mathbf{H}^{T}$. It must not be misunderstood as an estimate of diagonal elements of $\mathbf{A}$ itself. In contrast to the BECM, where it is acceptable to derive diagonal entries of $\tilde{\mathbf{B}}$ using $\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{T}$ under the conditions discussed in the last paragraph of section 2.1, the analysis error will generally be closer to the background error (i.e., larger) in air parcels not influenced directly (or not influenced at all) by observations. Therefore, equation (12) cannot provide an analysis error estimate in model grid space, as needed, e.g., for the purpose of cross-validating data from different sensors. Rather, (12) can be considered as a lower limit of analysis error in model grid space, and it can be concluded that the true analysis error will be in between $\tilde{\sigma}^{a}$ and $\tilde{\sigma}^{b}$.

4.5. Diagnostics and Tuning for Off-Diagonal Elements of B

[51] Based on equation (1), the off-diagonal elements $\tilde{b}_{ij}$ of $\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{T}$ can be estimated in analogy to (10) by

$$\tilde{b}_{ij} = \frac{1}{p_{ij}}\sum_{\substack{i,j=1 \\ i\neq j}}^{p_{ij}}\left(y_{i}^{a} - y_{i}^{b}\right)\left(y_{j}^{o} - y_{j}^{b}\right), \qquad (13)$$

for a subset of observations with p_{ij} members. Because equation (13) refers to observation space, it can only be expected to deliver useful information for the parameterization of matrix $\mathbf{B}$ in model space if the time difference between pairs of observations i and j is kept as small as possible. Therefore, only those pairs i,j have been included into the calculation, which are not separated by more than one model time step or 1200 seconds. This corresponds to approximately 1/3 of an ENVISAT orbit or a maximum spatial distance of about 13,500 km.

[52] In order to check the consistency of the correlation length parameter L, a set of six assimilation experiments with $L = 200$ km, $L = 400$ km, $L = 600$ km, $L = 800$ km, $L = 1000$ km, and $L = 1400$ km has been conducted, with the diagonal elements of $\mathbf{B}$ kept fixed at the values derived in section 4.3. Additionally, on three days (24 October to 26 October 2003) dedicated assimilation runs have been performed using observations within the 6 h tuning window. Based on these results, the off-diagonal elements of $\mathbf{H}\tilde{\mathbf{B}}\mathbf{H}^{T}$ have been evaluated by selecting pairs of observations in the same model level and by applying (13). The results have been normalized to correlation coefficients, accumulated into bins of 550 km extent and plotted as bar diagrams as shown in Figure 10 for model levels 15–19 (8.3–30 hPa, results for $L = 200$ and $L = 1400$ km not shown) and in Figure 11 for model levels 20–24 (30–96 hPa, results for $L = 200$ and $L = 1400$ km not shown). The bin size of 550 km has been chosen to match the distance of MIPAS profiles along the orbit track. The solid line is the Gaussian correlation function

$$c_{ij} = \frac{b_{ij}}{a_{0}} = \exp\left\{-\frac{r^{2}}{2L^{2}}\right\}, \qquad (14)$$

with r being the great circle distance between grid point i and j, and L the correlation length scale, which has actually been employed during the assimilation procedure. The normalizing parameters a_{0}, the values of which are given in Figures 10 and 11, are the diagnosed background variances for each layer and each setting of L. The dashed line depicts the same correlation function but with L replaced by its diagnosed counterpart $\tilde{L}$, which was obtained by a least squares fit to the data. It must be emphasized that the parameters a_{0} differ from the diagnosed background error variances $(\tilde{\sigma}_{i}^{b})^{2}$ derived in section 4.3. This is due to the fact that the diagnostic is performed on the full set of observations here. A thinning of measurements is not applicable, as this would remove too much information about background error correlations on short distances. As a result, the normalizing variances a_{0} are smaller than the corresponding $\tilde{\sigma}^{b}$ obtained for the same model region in section 4.3. Also, it can be observed that a_{0} decreases with increasing L. This behavior is consistent with the findings presented in
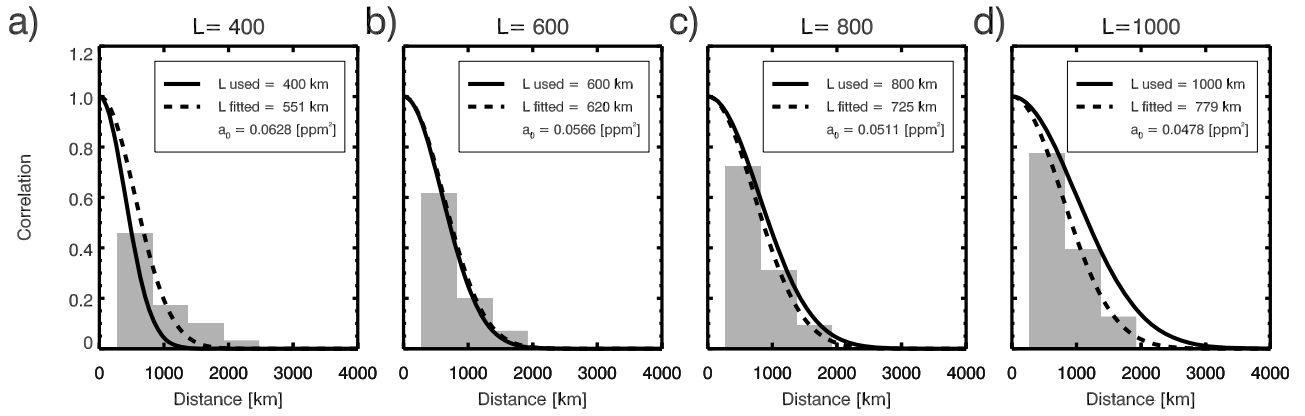
Figure 10. Tuning of the correlation length parameter L for model levels 15–19 (8.3–30 hPa). The bar diagrams give the correlation calculated according to equation (13) for four assimilation experiments with (a) $L = 400$ km, (b) $L = 600$ km, (c) $L = 800$ km, and (d) $L = 1000$ km. Data of three consecutive days (24–26 October 2003) have been sampled. The solid lines represent the correlation function with the prescribed length scale L (the length scale actually encoded in **B**), while the dashed lines show the correlation functions with the diagnosed correlation length scale $\tilde{L}$, where $\tilde{L}$ was obtained by a least squares fit to the respective data.

section 4.3, namely the loss of diagnosed variance with growing L (see Figure 4). For the purpose of gaining information on the consistency of L, the relative change of covariance with distance is important, and it is supposed here that this information can be derived, if the individual normalization a_0 is taken into account for each setting of L.

[53] The results show, that L seems to be inconsistent with $\tilde{L}$ for $L = 400$ km, $L = 1000$ km, and, to a lesser extent, for $L = 800$ km. The closest match between the specified correlation length scale L and its diagnosed counterpart $\tilde{L}$ is found for $L = 600$ km. Results for the other altitude ranges look very similar, although for the upper model domain the aforementioned small bias between observations and model background hampers the applicability of the method. Further, it must be mentioned that, due to the sampling of points i, j along the satellite orbit, the zonal structure of background error correlations cannot be estimated by means of this technique. Therefore, any anisotropy in background error correlations cannot be detected without accepting larger time differences between pairs of observations.

4.6. Examination of χ^2 Diagnostics

[54] In Figure 12a time series of normalized analysis cost function values J_p^a are shown for the period 24 October to 10 November 2003 for the six assimilation experiments, employing correlation length scales L in the range between 200 and 1400 km. It can be seen, that the value of J_p^a is highly sensitive to the specified correlation length scale. A high J_p^a mean value is found for the short ($L = 200$ km) length scale. A minimum mean value of 0.58 can be observed for $L = 400$ km followed by an increase of J_p^a with increasing correlation length scale up to a value of 0.83 for $L = 1400$ km.

[55] It should be noted that the increase of normalized analysis cost function values for very short correlation length scales is somewhat unexpected at first glance. In fact, it should be supposed that small values of L allow for a closer fit to observations (eventually leading to an over-fitting of measurements as L approaches zero). However, this is only true in a 3D-var context. In a 4D-var system, a degraded analysis obtained due to an unrealistic small
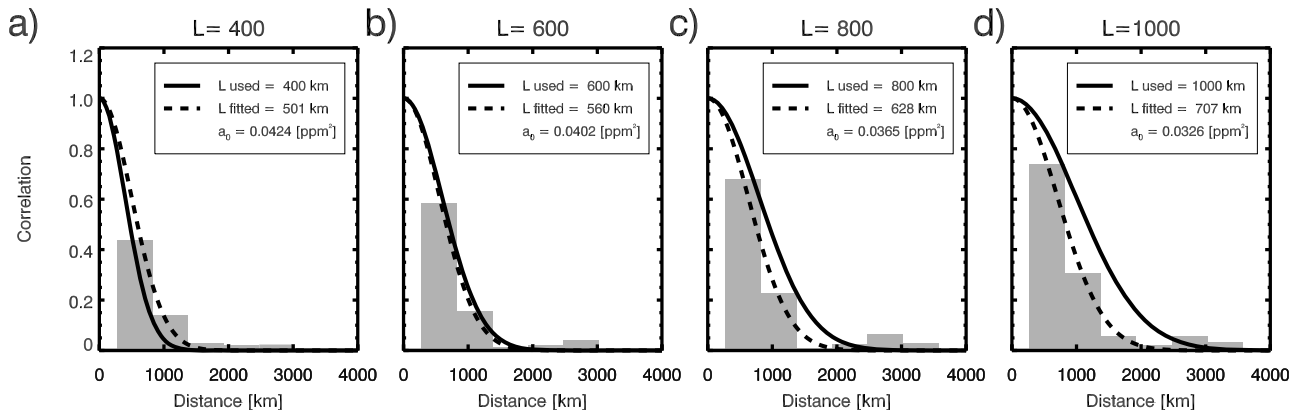

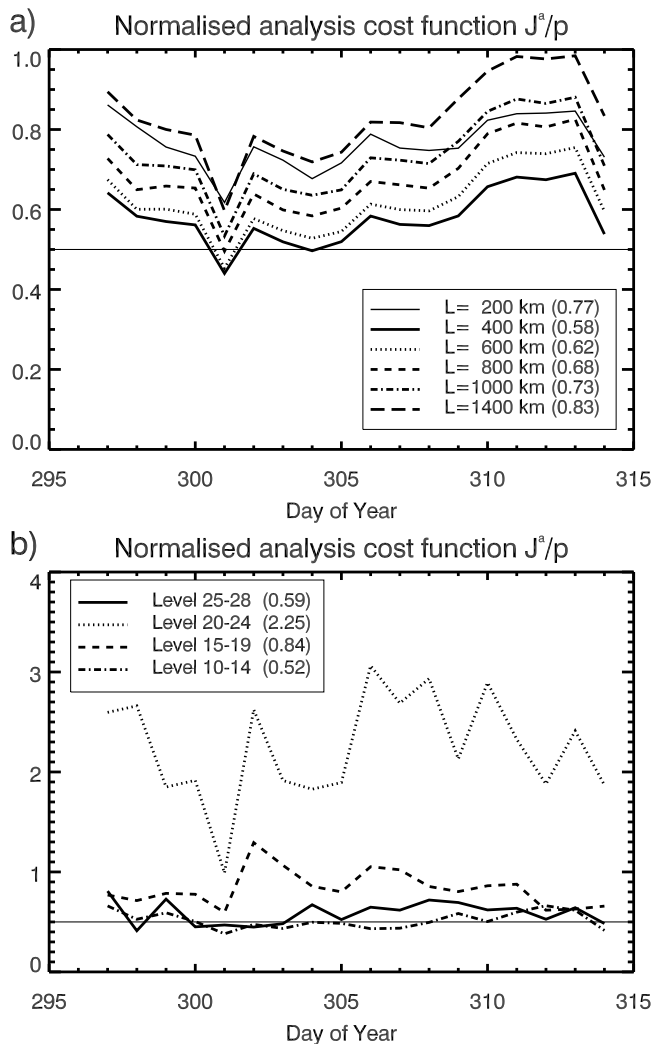**Figure 11.** As in Figure 10, but for model levels 20–24 (30–96 hPa).

a) Normalised analysis cost function J^a_p/p

Legend:
- L= 200 km (0.77)
- L= 400 km (0.58)
- L= 600 km (0.62)
- L= 800 km (0.68)
- L=1000 km (0.73)
- L=1400 km (0.83)

b) Normalised analysis cost function J^a_p/p

Legend:
- Level 25-28 (0.59)
- Level 20-24 (2.25)
- Level 15-19 (0.84)
- Level 10-14 (0.52)

Figure 12. (a) Normalized analysis cost function J^a_p for the six assimilation experiments with $L = 200$ km (thin solid), $L = 400$ km (solid), $L = 600$ km (dotted), $L = 800$ km (short dashes), $L = 1000$ km (dash-dotted), and $L = 1400$ km (long dashes) over the period 24 October to 10 November 2003 (day of year 297–314) for the whole model domain. Line styles and time averages for the different values of L are indicated in the box at the bottom right. (b) J^a_p inside the Antarctic region (90°S–60°S) for 221–96 hPa (model level 28–25, solid), 96–30 hPa (level 24–20, dotted), 30–8.3 hPa (level 19–15, dashed), and 8.3–2.4 hPa (level 14–10, dash-dotted). Values are valid for $L = 600$ km but are similar for other settings of L. Line styles and time averages for each height range are indicated in the box at the top left.

inspection of the data. It turned out, that at high southern latitudes between 96 and 30 hPa the values of J^a_p exceed the expected value by a factor of about 5 (Figure 12b). Interestingly, this region contains the polar vortex, which was, although already weakening, still present at this time of the year 2003 (see *Lahoz et al.* [2006] for a comprehensive description of the southern hemisphere polar vortex evolution in 2003). Modeled potential vorticity with its distinct high values in magnitude southward of 60°S together with corresponding ozone analyses can be found in ESB10 (Figure 8) for 23 October 2003 at the 55 hPa level.

[57] There are two possible reasons for the observed inconsistency of error statistics in the Antarctic polar vortex region:

[58] 1. The model skill might be lower within and at the edge of the polar vortex due to insufficient model resolution, numerical diffusion of the transport scheme, errors in modeling the correct position of the vortex, or other reasons. This is not reflected in the BECM model used in this study, as a latitudinal variation of background error is not taken into account.

[59] 2. MIPAS retrieval quality under polar vortex conditions might be lower and error estimates of the observations could possibly be too optimistic. Also, the assumptions made about systematic retrieval errors, as discussed in section 2.1.2, may be violated in this region.

[60] As a comprehensive clarification of this issue is beyond the scope of this paper, data from the Antarctic region have been excluded from the analysis of J^a_p and also J^f_p statistics, hereafter. It should be noted, that the described inconsistency between the model and observations could have been detected with the Desroziers et al. method, if the diagnosis had been performed for individual latitude bands. Due to the design of the tuning procedure (thinning of
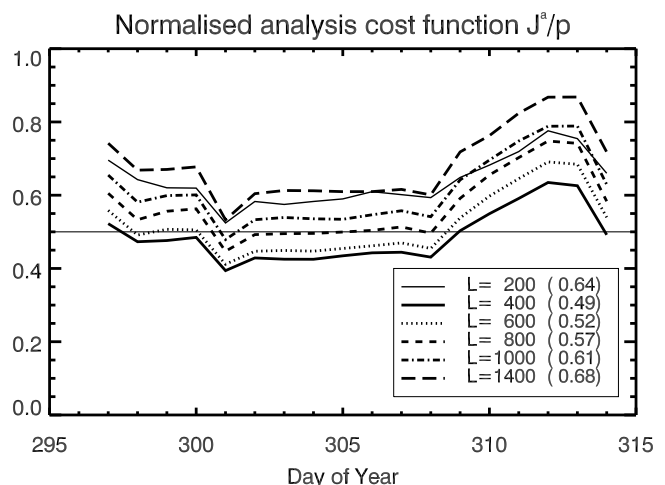
background error correlation directly affects the subsequent model integration over the assimilation window. As a consequence, observations are not as closely matched as it would have been the case if the analysis were obtained with a more realistic value of L. This is also the reason why the tuning of background error variances was not performed with L set to zero in this study, and a thinning of observations has been applied instead.

[56] The fact that the expected value of 0.5 is not met for either of the specified values of L gave reason to closer

Normalised analysis cost function J^a/p

Legend:
- L= 200 (0.64)
- L= 400 (0.49)
- L= 600 (0.52)
- L= 800 (0.57)
- L=1000 (0.61)
- L=1400 (0.68)

Figure 13. Normalized analysis cost function J^a_p for the six assimilation experiments with $L = 200$ km (thin solid), $L = 400$ km (solid), $L = 600$ km (dotted), $L = 800$ km (short dashes), $L = 1000$ km (dash-dotted), and $L = 1400$ km (long dashes) over the period 24 October to 10 November 2003. Profiles from latitudes between 50°S and 90°S have been excluded. Line styles and time averages for each value of L are indicated in the box at the bottom right.
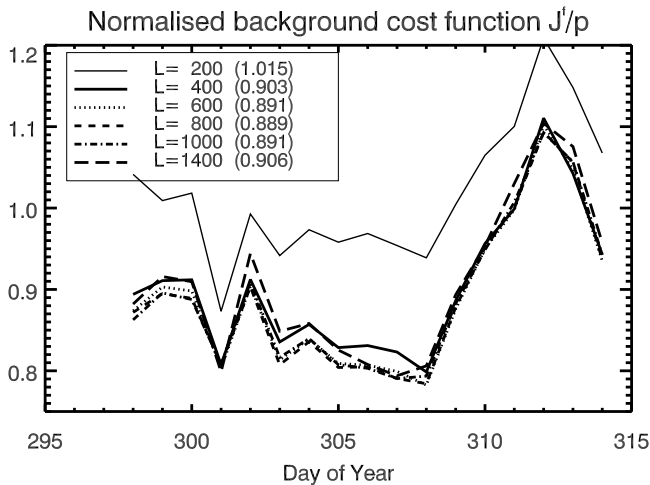
Figure 14. Normalized background cost function J_p^f for the six assimilation experiments with $L = 200$ km (thin solid), $L = 400$ km (solid), $L = 600$ km (dotted), $L = 800$ km (short dashes), $L = 1000$ km (dash-dotted), and $L = 1400$ km (long dashes) over the period 25 October to 10 November 2003. Profiles from latitudes between 50°S and 90°S have been excluded. Line styles and time averages for each value of L are indicated in the box at the top left.

observations, short tuning window) it is not likely however, that the tuning results obtained in section 4.3 have been significantly influenced.

[61] A plot similar to Figure 12a, but with the contribution of the domain 50°S–90°N to J_p^a only, is given in Figure 13. The time evolution of all J_p^a graphs is very similar, but mean values are lower now by about 15% with the value of 0.49 for $L = 400$ km being slightly below and the value 0.52 for $L = 600$ being slightly above 0.5.
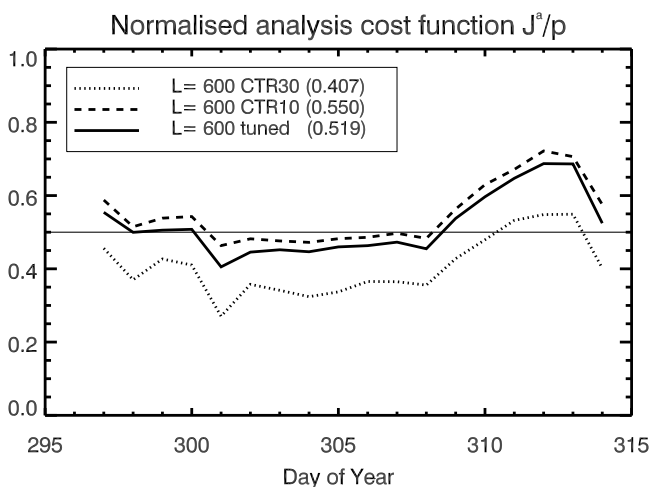
4.7. Examination of Forecast Skill

[62] The forecast skill as measured by J_p^f (with lower values indicating a better forecast) for the domain 50°S–90°N is displayed in Figure 14. As expected, less skillful forecasts are obtained with a very weak background error correlation ($L = 200$ km). The forecast quality is lower for $L = 400$ km at 9 out of 16 days, comparable to cases with larger settings of L on 6 days, and slightly better on one day. There is no significant difference between the forecast skill for the $L = 600$ km, $L = 800$ km and $L = 1000$ km settings. A further extension of background error correlations seems to degrade the forecast quality, as seen for the assimilation experiment employing $L = 1400$ km. Although longer time series would be necessary to gain a better statistical basis, it can be tentatively concluded that the forecast skill is largely unaffected by the choice of L over a broad range of values between 600 and 1000 km. In this situation it seems reasonable to select the shortest background error correlation length scale, which enables a skillful analysis and subsequent forecast. Also, the findings obtained with the Desroziers et al. method point toward the conclusion that the value $L = 600$ km can be considered as the most consistent choice.

[63] The fact that the differences in forecast skill as displayed in Figure 14 are relatively small for L within a range between 400 and 1400 km is an a posteriori justification for separately tuning σ_i^b and L: If the impact of L on the model forecast were large, separate sets $\sigma_{i,L}^b$ for each value of L had to be derived in order to account for the different quality of the forecast (=background), before proceeding with the analysis outlined in section 4.5.

5. Validation

[64] The performance of the assimilation system using the optimized BECM parameterization compared to the untuned (CTR30 and CTR10) configuration with respect to the χ^2 criterion and forecast skill is shown in Figures 15 and 16. It

Figure 15. Normalized analysis cost function J_p^a for the untuned (CTR30, dotted; CTR10, dashed) and tuned (solid) assimilation system (24 October to 10 November 2003). Profiles from latitudes between 50°S and 90°S have been excluded. Time averages for each assimilation experiment are indicated in the box at the top.
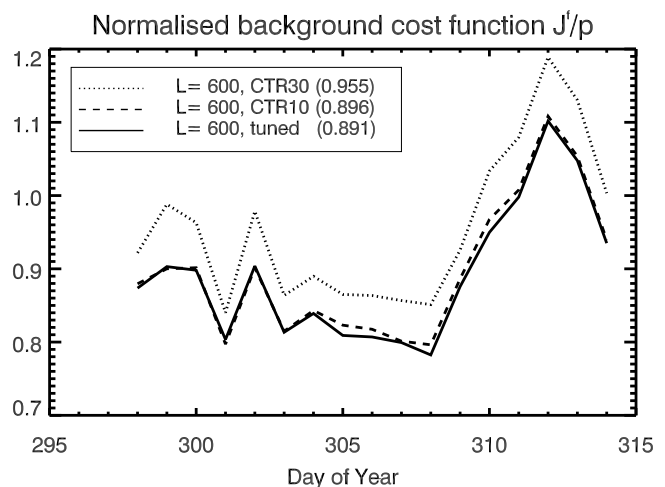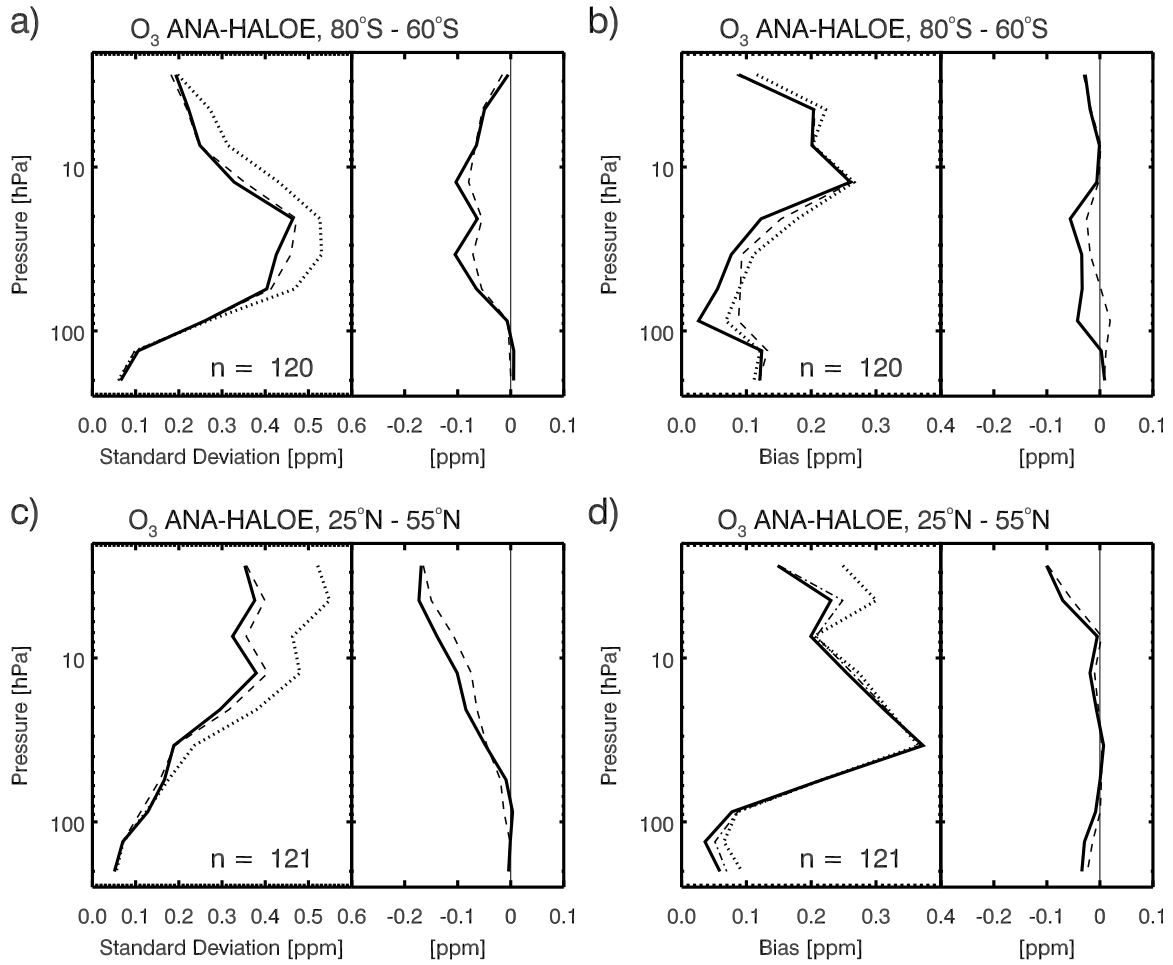

Figure 16. Normalized background cost function J_p^f for the untuned (CTR30, dotted; CTR10, dashed) and tuned (solid) assimilation system (25 October to 10 November 2003). Profiles from latitudes between 50°S and 90°S have been excluded. Time averages for each assimilation experiment are indicated in the box at the top.

Figure 17. (a) Standard deviation and (b) bias of analyzed ozone with respect to HALOE observations between 80°S and 60°S latitude (sunset scans). The analysis has been obtained by assimilating MIPAS profiles with untuned (CTR30, dotted; CTR10, dashed) and tuned (solid) background error covariances. The change relative to the CTR30 setting is given in the right section of each plot. (c) Standard deviation and (d) bias of analyzed ozone with respect to HALOE observations between 25°N and 55°N latitude (sunrise scans). These statistics include all available HALOE observations during the period 24 October to 10 November 2003.

is not surprising that the system employing the tuned BECM better meets the expected analysis cost function value of 0.5 (Figure 15), as this criterion was partly used during the tuning procedure (determination of the inflation factor f). More encouraging is the result, that the forecast skill, measured by the background (forecast) cost function J_p^f, improves, as demonstrated by the 6.7% reduction of J_p^f from 0.955 for CTR30 to 0.891 shown in Figure 16. It can be noted that the number of 6.7% does not reflect the fact that the lowest possible value of J_p^f is 0.5 (for an optimal forecast $J_p^f \approx J_p^a \approx 0.5$). Therefore, the improvement observed in this study corresponds to a 12% reduction relative to the possible gain. Only a slight improvement (0.5%) in forecast skill is observed relative to the CTR10 setting.

[65] The quality of analyses obtained with the tuned and untuned system, has been further evaluated by comparison with HALOE and SAGE II data. A total of 241 HALOE profiles collected from 30 October to 7 November 2003 is available for the case study period. About half of these profiles (sunset events) are located in a latitude band

between 60°S and 80°S, while the sunrise events took place between 25°N and 55°N. The bias and standard deviation of HALOE observations with respect to MIPAS based analyses (vertically interpolated to the location of HALOE profile points) have been calculated for these two latitude bands. The results presented in Figure 17 indicate that, relative to the CTR30 assimilation run, there is a substantial reduction of standard deviation between the 60 and the 2.4 hPa pressure level for both latitude bands. The decrease is about 0.05 to 0.1 ppm over large parts of the profiles, reaching values of 0.2 ppm in the uppermost model domain within the northern latitude band. The corresponding percentage decrease (not shown) is about 20% in the middle stratosphere. Relative to the CTR10 assimilation run the standard deviation with respect to HALOE profiles is decreased between 50 and 10 hPa in the southern and above 30 hPa in the northern latitude band. Below there is no significant difference between the CTR10 and the tuned BECM setting, although in the northern latitude band, CTR10 performs slightly better between 100 and 50 hPa. This is in line with
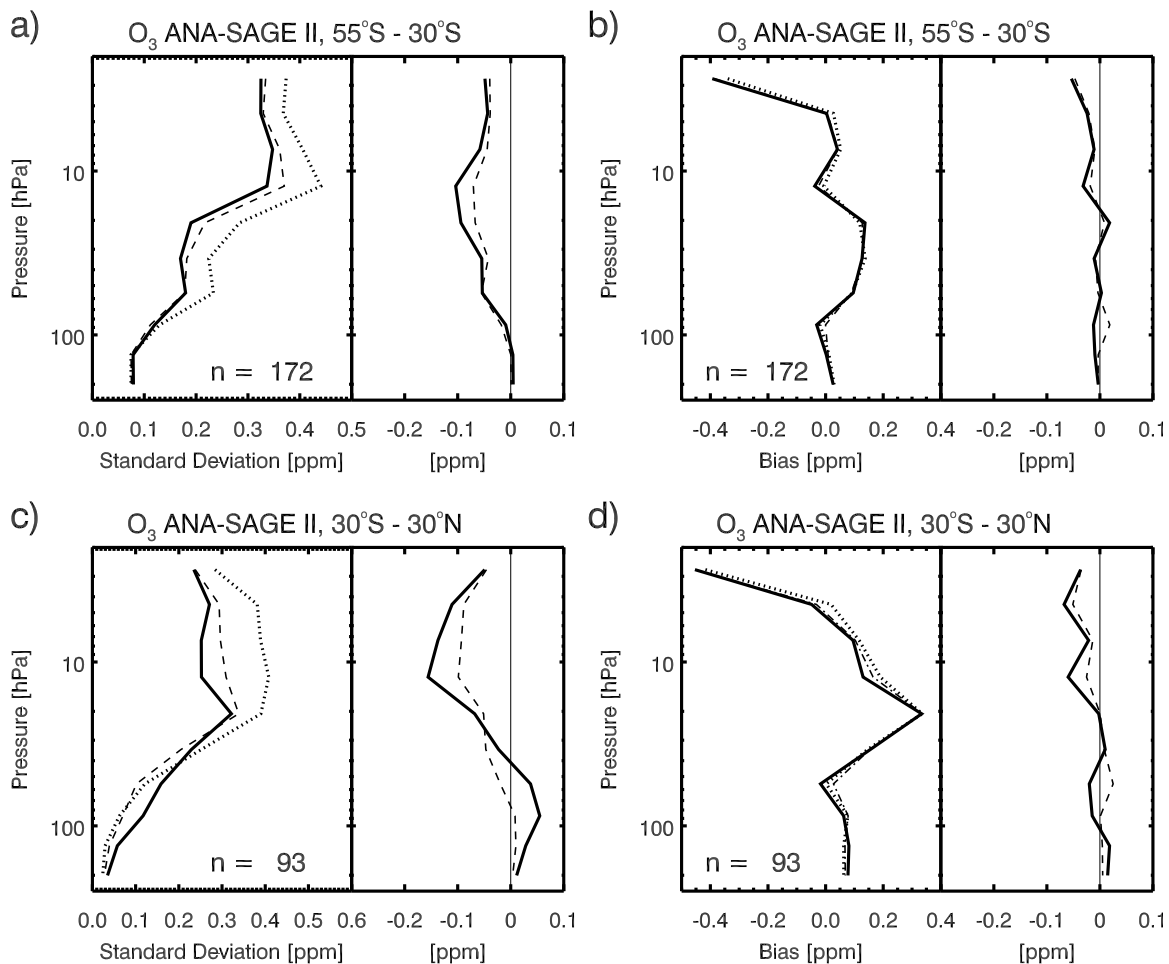
Figure 18. (a) Standard deviation and (b) bias of analyzed ozone with respect to SAGE II observations between 55°S and 30°S latitude. The analysis has been obtained by assimilating MIPAS profiles with untuned (CTR30, dotted; CTR10, dashed) and tuned (solid) background error covariances. The change relative to the CTR30 setting is given in the right section of each plot. (c) Standard deviation and (d) bias of analyzed ozone with respect to SAGE II observations between 30°S and 30°N latitude. These statistics include all SAGE II observations available during the period 24 October to 10 November 2003.

the finding that the 10% relative background error leads to a very good consistency in error statistics below 50 hPa (compare Figure 6). The bias is less affected by the tuning procedure, although a reduction can be observed at southern latitudes.

[66] SAGE II data is available for the whole case study period. The tangent point locations of SAGE II profiles vary slowly in latitude, covering a range from 55°S to 35°N. Standard deviation and bias of ozone analyses with respect to the SAGE II retrievals are shown in Figure 18 for southern midlatitudes (55°S to 30°S) and for tropical latitudes (30°S to 30°N). A reduction of standard deviation as a result of using the tuned background error covariances can be observed for SAGE II midlatitude profiles. The standard deviation profiles' shape and magnitude is very similar to the HALOE northern midlatitude profiles. For tropical SAGE II profiles, an increase of standard deviation due to the tuning of BECM parameters can be seen in the region below ≈ 50 hPa. It was found that this is a consequence of the relatively simple covariance setup used for this study, which does not allow for latitudinal variation of background

errors. The Desroziers et al. diagnostics applied to data in specific latitude ranges (not shown), confirms that the 10% relative background error used for CTR10 performs better (is more consistent) in the tropical region than the tuned σ^b obtained as an average over all latitudes. The bias relative to SAGE II profiles remains practically unaffected by the tuning procedure.

[67] For further validation, there are 67 sonde ascents from 29 locations available for the period 24 October to 10 November (excluding five ascents that did not reach a minimum burst height of 20 hPa). Since there are only nine ascents for southern extratropical stations, the analysis has been focused on tropical and northern latitudes. The results are shown in Figure 19. Generally, standard deviations are larger compared to Figures 17 and 18. This can be explained by the fact that ozone sondes offer a higher spatial resolution than the remote sensing devices HALOE and SAGE II, even after smoothing to the model grid with approximately 2 km vertical grid spacing. Nevertheless, some common features can be observed: In the tropics, a positive impact of the optimized BECM is found only above the 30 hPa level,
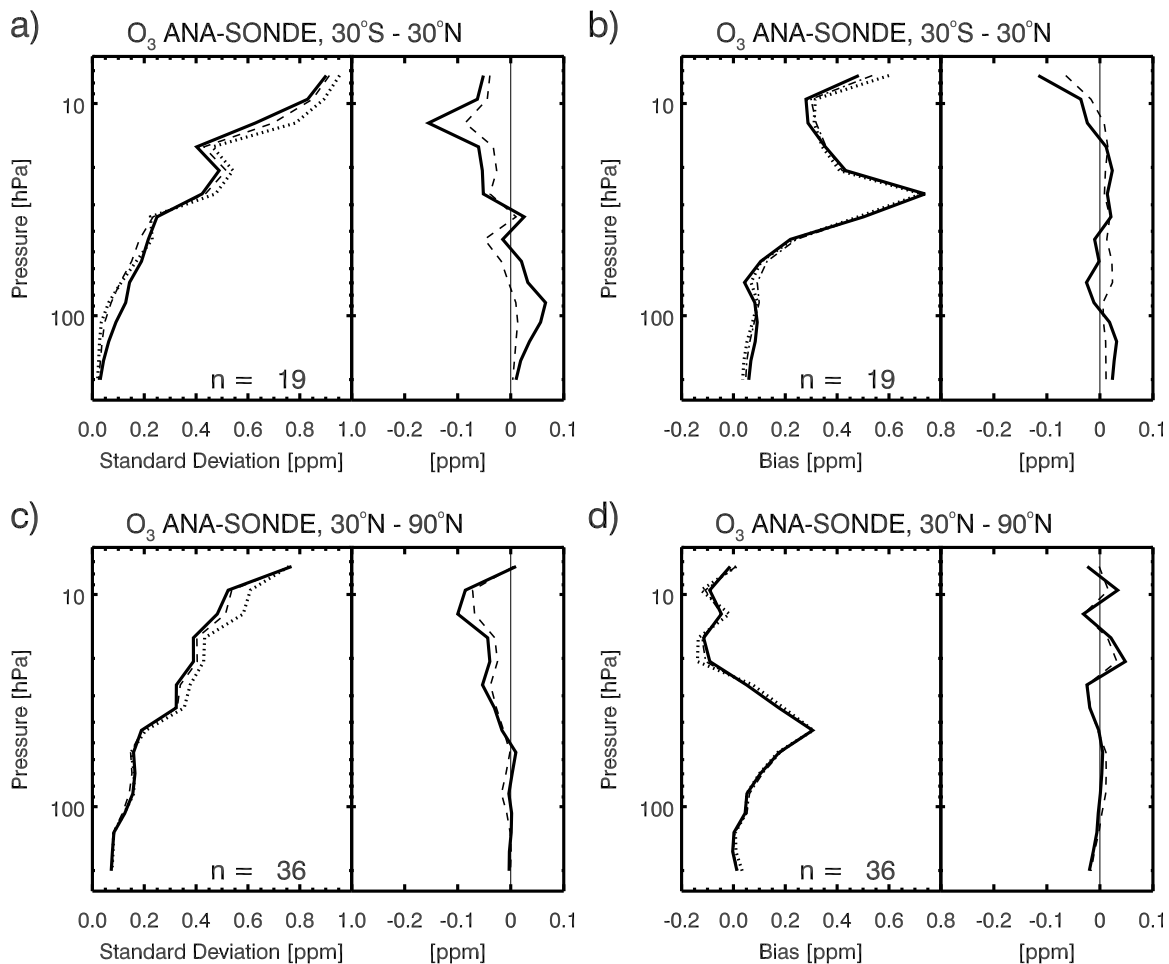
Figure 19. (a) Standard deviation and (b) bias of analyzed ozone with respect to ozone soundings at tropical stations (30°S–30°N). The analysis has been obtained by assimilating MIPAS profiles with untuned (CTR30, dotted; CTR10, dashed) and tuned (solid) background error covariances. The change relative to the CTR30 setting is given in the right section of each plot. (c) Standard deviation and (d) bias of analyzed ozone with respect to ozone sonde measurements in the latitude range 30°N–90°N.

while the untuned BECM appears to perform better below. This effect, similar in magnitude, is also seen in the comparison with SAGE II data (see discussion above). A reduction of standard deviation due to the use of the tuned BECM is detected in the latitude range 30°N to 90°N above the 50 hPa level, although this effect is slightly weaker than the reduction seen in the HALOE and SAGE II validation cases.

6. Discussion and Conclusions

[68] A set of a posteriori diagnostics in observation space, which was devised by *Desroziers et al.* [2005], has been applied for the first time in the context of atmospheric chemistry data assimilation. A five-parameter background error covariance model for the 4D-var assimilation system SACADA was set up and iteratively optimized. The five parameters have been chosen to be the (globally averaged) ozone background error variances for four altitude ranges and a background error correlation length scale. The tuning has been performed by assimilating MIPAS ozone profiles provided by ESA for a period from 21 October to

10 November 2003. As the results depend to some degree on the error characteristics of the assimilated data, it can be expected that the optimized parameters will be slightly different for other instruments.

[69] Control assimilation runs were performed with the assumption that the ozone background error is a fixed fraction of the ozone value. A coarse estimate of a 30% ozone background error, and a more plausible but ad hoc estimate of 10% were selected as control cases. The coarse 30% estimate was found to be by far too large, while the 10% estimates leads to a consistent specification of background and observation errors below 50 hPa but to increasing inconsistency for higher altitudes. The optimized background error variances were inferred to be of the same order of magnitude as the mean MIPAS observation error variances. A length scale of $L = 600$ km assuming a Gaussian correlation function was found to be the optimal setting for the SACADA system. In contrast to the large impact of background error variances on forecast quality, the forecasts skill appears to be only weakly dependent on the precise value of L for a range of $L = 400$ to $L = 1400$ km. It should be noted that observations providing less spatial

coverage than, for example, MIPAS profiles, cannot be used to derive information on spatial background error correlations with the Desroziers et al. method.

[70] The assimilation system performance was enhanced considerably through the use of optimized BECM parameters compared to the control run using the 30% error estimate. The forecast skill, which was defined as the squared observation-minus-forecast differences weighted by observation error, improved by 6.7%. A statistical validation with HALOE and SAGE II data reveals, that the standard deviation of the analyzed ozone field with respect to HALOE and SAGE II is reduced by approximately 20% between 60 and 2 hPa in the extratropics. For tropical latitudes a similar improvement is found above the 30 hPa level. However, the untuned BECM settings appear to be superior in the tropical lower stratosphere and upper troposphere region. Here, further improvement can be expected by making the tuned variances and the correlation length scale also dependent on latitude. Expectedly, the impact of the tuned BECM compared to the 10% background error setting is smaller, yet still positive. The validation with ozone sonde profiles confirms the positive impact of the improved background error covariances on stratospheric ozone analyses. These findings underline the importance of giving the appropriate weight to the background field in data assimilation schemes.

[71] The Desroziers et al. diagnostic is easy to implement and causes practically no additional computational cost. However, the application of this method must be conducted with care, as the diagnostic of background and observation error variances can be hampered by misspecification of spatial background error correlations, and, in a 4D-var context, the presence of model errors. Therefore, it is necessary to perform the optimization of variances on a subset of observations only (reducing the influence of the BECM's off-diagonal elements) and with a shortened assimilation window (reducing the influence of model errors).

[72] Further, it was found that a simultaneous tuning of background and observation errors is not feasible, since the Desroziers et al. method appears to be nonunique, that is, it allows for a range of consistent background and observation error specifications. Therefore, this study assumes that the observation error variances provided with the MIPAS ozone profiles are correct. For the region containing the Antarctic polar vortex it was found that an unresolved inconsistency in error statistics remains. This inconsistency may be caused by the treatment of MIPAS retrieval error, which might be not appropriate in this region, or by a lower model skill in and at the edge of the vortex, which would not have been reflected in the BECM setup used for this study.

References

Baier, F., T. Erbertseder, O. Morgenstern, M. Bittner, and G. Brasseur (2005), Assimilation of MIPAS observations using a three-dimensional global chemistry-transport model, *Q. J. R. Meteorol. Soc.*, *131*, 3529–3542.

Bonavita, M., L. Torrisi, and F. Marcucci (2008), The ensemble Kalman filter in an operational regional NWP system: Preliminary results with real observations, *Q. J. R. Meteorol. Soc.*, *134*, 1733–1744.

Chapnik, B., G. Desroziers, F. Rabier, and O. Talagrand (2004), Properties and first application of an error-statistics tuning method in variational assimilation, *Q. J. R. Meteorol. Soc.*, *130*, 2253–2275.

Chapnik, B., G. Desroziers, F. Rabier, and O. Talagrand (2006), Diagnosis and tuning of observational error in a quasi-operational data assimilation setting, *Q. J. R. Meteorol. Soc.*, *132*, 543–565.

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

Cohn, S. E. (1993), Dynamics of short-term univariate forecast error covariances, *Mon. Weather Rev.*, *121*, 3123–3149.

Cortesi, U., et al. (2007), Geophysical validation of MIPAS-ENVISAT operational ozone data, *Atmos. Chem. Phys.*, *7*, 4807–4867.

Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge Univ. Press, Cambridge, U.K.

Dando, M. L., A. J. Thorpe, and J. R. Eyre (2007), The optimal density of atmospheric sounder observations in the Met Office NWP system, *Q. J. R. Meteorol. Soc.*, *133*, 1933–1943.

Dee, D. P. (2005), Bias and data assimilation, *Q. J. R. Meteorol. Soc.*, *131*, 3323–3343.

Desroziers, G., and S. Ivanov (2001), Diagnosis and adaptive tuning of observation error parameters in a variational assimilation, *Q. J. R. Meteorol. Soc.*, *127*, 1433–1452.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis error statistics in observation space, *Q. J. R. Meteorol. Soc.*, *131*, 3385–3396.

Dudhia, A., V. L. Jay, and C. D. Rodgers (2002), Microwindow selection for high-spectral-resolution sounders, *Appl. Opt.*, *41*(18), 3665–3673.

Elbern, H., A. Strunk, H. Schmidt, and O. Talagrand (2007), Emission rate and chemical state estimation by 4-dimensional variational inversion, *Atmos. Chem. Phys.*, *7*, 1–59.

Elbern, H., J. Schwinger, and R. Botchorishvili (2010), Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: System configuration, *J. Geophys. Res.*, *115*, D06302, doi:10.1029/2009JD011953.

Fierli, F., A. Hauchecorne, S. Bekki, B. Théodore, and O. F. d'Andon (2002), Data assimilation of stratospheric ozone using a high-resolution transport model, *Geophys. Res. Lett.*, *29*(10), 1381, doi:10.1029/2001GL014272.

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

Fischer, H., et al. (2008), MIPAS: An instrument for atmospheric and climate research, *Atmos. Chem. Phys.*, *8*, 2151–2188.

Hollingsworth, A., and P. Lönnberg (1986), The statistical structure of short-range forecast errors as determined from radiosonde data: Part I. The wind field, *Tellus, Ser. A*, *38*, 111–136.

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

Lahoz, W. A., A. J. Geer, and A. O'Neill (2006), Dynamical evolution of the 2003 southern hemisphere stratospheric winter using Envisat trace gas observations, *Q. J. R. Meteorol. Soc.*, *132*, 1985–2008.

Li, H., E. Kalnay, and T. Miyoshi (2009), Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter, *Q. J. R. Meteorol. Soc.*, *135*, 523–533.

Liu, D. C., and J. Nocedal (1989), On the limited memory BFGS method for large-scale optimization, *Math. Program.*, *45*, 503–528.

Lönnberg, P., and A. Hollingsworth (1986), The statistical structure of short-range forecast errors as determined from radiosonde data: Part II. The covariance of height and wind errors, *Tellus, Ser. A*, *38*, 137–161.

Majewski, D., D. Liermann, P. Prohl, B. Ritter, M. Buchhold, T. Hanisch, G. Paul, W. Wergen, and J. Baumgardner (2001), The operational global icosahedral-hexagonal gridpoint model GME: Description and high-resolution tests, *Mon. Weather Rev.*, *130*, 319–338.

Ménard, R., S. E. Cohn, L.-P. Chang, and P. M. Lyster (2000a), Assimilation of stratospheric chemical tracer observations using a Kalman Filter: Part II. χ^2-validated results and analysis of variance and correlation dynamics, *Mon. Weather Rev.*, *128*, 2672–2686.

Ménard, R., S. E. Cohn, L.-P. Chang, and P. M. Lyster (2000b), Assimilation of stratospheric chemical tracer observations using a Kalman Filter: Part I. Formulation, *Mon. Weather. Rev.*, *128*, 2654–2671.

Nazaryan, H., M. P. McCormick, and J. M. Russel III (2005), New studies of SAGE II and HALOE ozone profile and long-term change comparisons, *J. Geophys. Res.*, *110*, D09305, doi:10.1029/2004JD005425.

Nocedal, J. (1980), Updating quasi-newton matrices with limited storage, *Math. Comput.*, *35*(151), 773–782.

Raspollini, P., et al. (2006), MIPAS level 2 operational analysis, *Atmos. Chem. Phys.*, *6*, 5605–5630.

Rodgers, C. D. (2000), *Inverse Methods for Atmospheric Sounding*, World Sci., Hackensack, N. J.

Rösevall, J. D., D. P. Murtagh, J. Urban, and A. K. Jones (2007), A study of polar ozone depletion based on sequential assimilation of satellite data from the ENVISAT/MIPAS and Odin/SMR instruments, *Atmos. Chem. Phys.*, *7*, 899–911.

Russell, J. M., et al. (1993), The halogen occultation experiment, *J. Geophys. Res.*, *98*, 10,777–10,797, doi:10.1029/93JD00799.

Segers, A. J., H. J. Eskes, R. J. van der A, R. F. van Oss, and P. F. J. van Velthoven (2005), Assimilation of GOME ozone profiles and a global chemistry-transport model using a Kalman filter with anisotropic covariance, *Q. J. R. Meteorol. Soc.*, *131*, 477–502.

Stajner, I., N. Winslow, R. B. Rood, and S. Pawson (2004), Monitoring of observation errors in the assimilation of satellite ozone data, *J. Geophys. Res.*, *109*, D06309, doi:10.1029/2003JD004118.

Talagrand, O. (1997), Assimilation of observations: An introduction, *J. Meteorol. Soc. Jpn.*, *75*(1B), 191–209.

Talagrand, O. (2003), A posteriori validation of assimilation algorithms, in *Data Assimilation for the Earth System*, edited by R. Swinbank, V. Shutyaev, and W. A. Lahoz, pp. 85–95, Kluwer Acad., Dordrecht, Netherlands.

Trémolet, Y. (2006), Accounting for an imperfect model in 4D-Var, *Q. J. R. Meteorol. Soc.*, *132*(621), 2483–2504.

Wang, H. J., D. M. Cunnold, L. W. Thomason, J. M. Zawodny, and G. E. Bodeker (2002), Assessment of SAGE II version 6.1 ozone data quality, *J. Geophys. Res.*, *107*(D23), 4691, doi:10.1029/2002JD002418.

Weaver, A., and P. Courtier (2001), Correlation modelling on the sphere using a generalized diffusion equation, *Q. J. R. Meteorol. Soc.*, *127*, 1815–1846.

Xu, Q., L. Wei, A. van Tuyl, and E. H. Barker (2001), Estimation of three-dimensional error covariances: Part I. Analysis of height innovation vectors, *Mon. Weather Rev.*, *129*, 2126–2135.

H. Elbern and J. Schwinger, Rhenish Institute for Environmental Research, University of Cologne, Aachener Str. 209, 50931 Cologne, Germany. (he@riu.uni-koeln.de; js@riu.uni-koeln.de)