

Comparison of the ensemble Kalman filter and 4D-Var assimilation methods using a stratospheric tracer transport model

Sergey Skachko¹, Quentin Errera¹, Richard Ménard², Yves Christophe¹, and Simon Chabrillat¹

¹Belgian Institute for Space Aeronomy, BIRA-IASB, Brussels, 1180 Belgium

²Air Quality Research Division, Environment Canada, Dorval, Canada

Correspondence to: Sergey Skachko
(Sergey.Skachko@aeronomie.be)

Abstract. An Ensemble Kalman filter (EnKF) assimilation method is applied to the tracer transport using the same stratospheric transport model as in the 4D-Var assimilation system BASCOE. This EnKF version of BASCOE was built primarily to avoid the large costs associated with the maintenance of an adjoint model. The EnKF developed in BASCOE accounts for two adjustable parameters: a parameter α controlling the model error term and a parameter r controlling the observational error. The EnKF system is shown to be markedly sensitive to these two parameters, which are adjusted based on the monitoring of a χ^2 -test measuring the misfit between the control variable and the observations. The performance of the EnKF and 4D-Var versions was estimated through the assimilation of Aura-MLS ozone observations during an 8-month period which includes the formation of the 2008 Antarctic ozone hole. To ensure a proper comparison, despite the fundamental differences between the two assimilation methods, both systems use identical and carefully calibrated input error statistics. We provide the detailed procedure for these calibrations, and compare the two sets of analyses with a focus on the lower and middle stratosphere where the ozone lifetime is much larger than the observational update frequency. Based on the Observation-minus-Forecast statistics, we show that the analyses provided by the two systems are markedly similar, with biases smaller than 5 % and standard deviation errors smaller than 10% in most of the stratosphere. Since the biases are markedly similar, they have most probably the same causes: these can be deficiencies in the model and in the observation dataset, but not in the assimilation algorithm nor in the error calibration. The remarkably similar performance also shows that in the context of stratospheric transport, the choice of the assimilation method can be based on application-dependent factors, such as CPU cost or the ability to generate an ensemble of forecasts.

1 Introduction

Two of the most important and widely used data assimilation methods are the Four-Dimensional Variational method (4D-Var: Talagrand and Courtier, 1987) and the Ensemble Kalman Filter (EnKF: Evensen, 1994; Houtekamer and Mitchell, 1998; Evensen, 2003). Although they solve similar estimation problems, they are built around different constraints and

thus have different strengths and weaknesses. The BASCOE (Belgian Assimilation System for Chemical Observations) system was originally developed with the 4D-Var assimilation method applied to a stratospheric CTM (Errera et al., 2008; Errera and Ménard, 2012). This variational method determines the initial conditions which optimize the fit between model
25 forecast and observations over a period, i.e. an assimilation window. In atmospheric chemistry, an assimilation window of 12 h (Flemming et al., 2009) or 24 h (Errera et al., 2008; Elbern et al., 2010) is typically used. The 4D-Var provides an accurate solution, but requires the development and maintenance of an adjoint model, which may be a time consuming task in the CTM context.

The most popular alternative to the 4D-Var is the EnKF which consists in a Monte Carlo method (Evensen, 1994).
30 As the 4D-Var, the EnKF is built on the assumption of Gaussian-distributed observation errors to estimate the minimum variance in the misfit between model forecast and observations. But the EnKF computes this minimum variance estimate at each time step of the model by computing explicitly its error covariances. It does not require an adjoint model but assumes that the forecast errors are Gaussian-distributed. In the 4D-Var scheme, the evolution of forecast error within the assimilation window is computed by the model (whether it is accurate and appropriate or not) and is used generally
35 as a strong constraint. By contrast the EnKF relaxes this assumption into a weak constraint by adding a model error covariance to the analysis error covariance which becomes dynamically-propagated (for more details, see Lorenc, 2003; Ménard and Daley, 1996). Hence, the model error covariance is of great importance for the filter performance. Moreover, the uncertainty of the EnKF analysis is directly provided by the spread of the ensemble of analyses.

The 4D-Var and the EnKF have comparable computational costs. The advantages and disadvantages of each method in
40 the meteorological context have been discussed in several papers (e.g. Hamill, 2006; Kalnay et al., 2007). A rigorous intercomparison was also presented by Buehner et al. (2010b) in the context of global NWP (Numerical Weather Prediction) system with real observations. In this context, it was shown that the EnKF error variance is larger than with the 4D-Var. In their intercomparison paper, Buehner et al. (2010a) also conducted different variational experiments using static covariances with horizontally homogeneous and isotropic correlations as well as flow-dependent EnKF covariances with
45 spatial localization. The authors went further and made a hybrid system called Ensemble-4DVar using flow-dependent EnKF covariances without the need of the tangent-linear or adjoint versions of the model. An overall conclusion obtained by Miyoshi et al. (2010) with the Japanese weather prediction system is that both systems have essentially comparable performance.

In the context of chemical modelling, Lahoz and Errera (2010) and Sandu and Chai (2011) reviewed different assimila-
50 tion methods and challenges in chemical data assimilation. Data assimilation systems based on a CTM are often developed within the variational approach (Khattatov et al., 1999; Errera et al., 2008), but also with sequential filtering (Khattatov et al., 2000; Ménard et al., 2000; Miyazaki et al., 2012). Recently Sekiyama et al. (2011) constructed a total ozone assimilation system on the basis of a four-dimensional local ensemble transform Kalman filter (LETKF). Nakamura et al. (2013) applied the EnKF to stratospheric ozone data assimilation using a multi-model approach. Meanwhile some developments
55 based on the EnKF have begun to address tropospheric composition (Constantinescu et al., 2007a; Liu et al., 2012).

In the context of chemical data assimilation, few studies have been devoted to the comparison of the 4D-Var and EnKF methods. Constantinescu et al. (2007a) have compared the EnKF with an operational-like 4D-Var setting using common background errors modelled by autoregressive processes applied to a tropospheric chemistry model. Wu et al. (2008)

presented an intercomparison of four assimilation methods including the 4D-Var and the EnKF. The paper was organized as a sensitivity study with respect to different model and assimilation parameters. The experiments were conducted over a short periods of typically one or two days. One of their conclusions was that the EnKF is superior to the 4D-Var, but also that Optimum Interpolation is superior to the EnKF. However from the study, it was unclear whether each assimilation system was tuned to provide its best performance. Hence these conclusions were not entirely convincing as the individual systems may have performed differently with different parameter values.

In the present study, the EnKF and the 4D-Var were both tuned to provide their best performance while using the same spectral formulation for the prescribed background error covariance. First of all, the background error covariance is calibrated within the 4D-Var using the NMC method. The calibrated errors are passed to the EnKF to generate the initial ensemble and the model error term. The EnKF is then tuned to provide its best results with χ^2 diagnostics close to one (Ménard et al., 2000) by calibration of the observation and model error covariance. The 4D-Var uses the observation covariance error calibrated within the EnKF experiments. We have not attempted to introduce a localization of error covariances in the 4D-Var because the localizations in a 4D-Var and EnKF are not strictly equivalent (Buehner et al., 2010a). However, the prescribed correlation length-scales in the EnKF were adjusted to match, after localization, those prescribed in the 4D-Var.

The next section describes the configurations of the EnKF and the 4D-Var data assimilation systems used in this study. The Section 3 describes the experimental set-up and specifically the calibration of the error variances in the two systems. Section 4 compares their results. Finally, some conclusions are given in Sect. 5.

2 Description of the EnKF and 4D-Var data assimilation systems

2.1 Configuration of the 3D-CTM

The comparison of EnKF and 4D-Var is performed using a tracer version of the BASCOE CTM. The model in its usual configuration includes 57 chemical species with a full description of stratospheric chemistry (Errera et al., 2008). All species are advected via the Flux-Form Semi-Lagrangian scheme (Lin and Rood, 1996). For the purposes of this study as well as to reduce the CPU time, the chemistry is turned off as in Errera and Ménard (2012) and only the advection of ozone (O_3) is considered. The CTM is driven by winds and temperatures obtained from the ECMWF ERA-Interim reanalysis (Dee et al., 2011). The horizontal resolution of the model grid is 3.75° longitude by 2.5° latitude. Vertically, the model uses a subset of 37 levels of the ERA-Interim 60 levels which excludes most tropospheric levels. The vertical domain extends from 0.1 hPa down to the surface. Hence the model state is described by the vector $\mathbf{x} \in \mathfrak{R}^n$ of length $n = 96 \times 73 \times 37 \approx 2.6 \times 10^5$. Finally, the model time step is set to 30 minutes.

2.2 The 4D-Var system

The detailed description of the BASCOE-4DVar data assimilation system is provided in Errera and Ménard (2012). Here we give only the features relevant to the aims of this study. The evolution of the model state vector between the time step

$k-1$ and k is computed by the model operator:

$$\mathbf{x}(t_k) = M_{k-1,k}(\mathbf{x}(t_{k-1})), \quad k \in [0, K], \quad (1)$$

where k is the time index, $M_{k-1,k}$ is the model operator between t_{k-1} to t_k and K is the number of time steps within the assimilation window. In the 4D-Var experiments performed in this study, the assimilation window is set to 24 hours such
 95 that, considering the model time step of 30 minutes, $K = 48$.

4D-Var data assimilation is carried out by minimizing the so-called cost function (Talagrand and Courtier, 1987)

$$J = \frac{1}{2}[\mathbf{x}(t_0) - \mathbf{x}^b(t_0)]^T \mathbf{B}_0^{-1}[\mathbf{x}(t_0) - \mathbf{x}^b(t_0)] + \frac{1}{2} \sum_{k=0}^K \left(\mathbf{H}_k M_{0,k}(\mathbf{x}(t_0) - \mathbf{x}^b(t_0)) - \mathbf{d}_k \right)^T \mathbf{R}_k^{-1} \left(\mathbf{H}_k M_{0,k}(\mathbf{x}(t_0) - \mathbf{x}^b(t_0)) - \mathbf{d}_k \right), \quad (2)$$

where $\mathbf{x}^b(t_0) \in \mathfrak{R}^n$ is the background model state and $\mathbf{B}_0 \in \mathfrak{R}^{n \times n}$ is the background error covariance matrix; \mathbf{H}_k is the
 100 observation operator at time t_k ; the vector $\mathbf{d}_k = \mathbf{y}(t_k) - \mathbf{H}_k M_{0,k} \mathbf{x}^b(t_0)$ is the first-guess innovation vector at time t_k ; the $\mathbf{y}(t_k) \in \mathfrak{R}^{m_k}$ and $\mathbf{R}_k \in \mathfrak{R}^{m_k \times m_k}$ represent the observational vector and its associated error covariance matrix at time t_k , respectively; m_k is the number of observations assimilated during time step k .

The BASCOE system has been, up to now, designed to assimilate observational profiles delivered by limb-scanning instruments. Hence the observation operator \mathbf{H}_k simply consists in a linear interpolation of the model value at the ob-
 105 servation tangent point. We assume that the observation errors are uncorrelated both horizontally and vertically. The observation error covariance matrix \mathbf{R}_k is thus defined diagonal:

$$\mathbf{R}_k(i,j) = \begin{cases} (r \sigma_y(i)|_{t_k})^2, & \text{if } i = j \\ 0, & \text{if } i \neq j, \end{cases} \quad (3)$$

where r is an *adjustable observation error parameter* and $\sigma_y(i)|_{t_k}$ is the measurement error at level i and time t_k . The observations and their errors are described in Sect. 3.1 while the adjustment of r is described in Sect. 2.5. Note that the
 110 parameter r governing the observational error matrix \mathbf{R}_k is introduced into the BASCOE 4D-Var system to allow the comparison with the EnKF.

The dimension n of the matrix \mathbf{B}_0 makes the computation of the background term of Eq. (2) unfeasible by current computers. To avoid the inversion of \mathbf{B}_0 , a control variable transform is introduced:

$$\mathbf{L}\xi = \mathbf{x}_0 - \mathbf{x}_0^b \equiv \delta\mathbf{x}_0, \quad (4)$$

115 where ξ is a new control variable, $\delta\mathbf{x}_0$ is the analysis increment and \mathbf{L} is the square root of \mathbf{B}_0 :

$$\mathbf{B}_0 = \mathbf{L}^T \mathbf{L}. \quad (5)$$

Hence, the cost function is then re-written as:

$$J(\xi) = \frac{1}{2} \xi^T \xi + \frac{1}{2} \sum_{k=0}^K \left(\mathbf{H}_k M_{0,k}(\mathbf{L}\xi) - \mathbf{d}_k \right)^T \mathbf{R}_k^{-1} \left(\mathbf{H}_k M_{0,k}(\mathbf{L}\xi) - \mathbf{d}_k \right). \quad (6)$$

The method used to formulate the operator \mathbf{L} is discussed in Sect. 2.4 and additional information of this incremental form
 120 of 4D-Var may be found in Errera and Ménard (2012). The present study used BASCOE 4D-Var version b07.27.

2.3 The EnKF system

In this section, we describe a specific variant of the EnKF algorithm as implemented into the BASCOE system (BASCOE EnKF version b08.06). The general algorithm follows the theoretical formulation of the EnKF with perturbed observations (Houtekamer and Mitchell, 1998; Evensen, 2003). Here, we provide only details that are essential to understand the performed experiments.

An ensemble of initial states is produced by adding, to a model state, a set of spatially correlated perturbations according to the prescribed initial error covariance. The details of the procedure are described in details in Sect. 2.4. The ensemble of model states is propagated forward in time using the same tracer version of the BASCOE CTM as used in the 4D-Var system (see Sect. 2.1). In a practical implementation, the model error covariance is represented by the addition of a stochastic noise $\boldsymbol{\eta}_i$ to each ensemble member at each model time step:

$$\mathbf{x}_i^f(t_k) = M_{k-1,k}(\mathbf{x}_i^a(t_{k-1})) + \boldsymbol{\eta}_i(t_k), \quad i \in [1, N], \quad (7)$$

where N is the size of the ensemble, the superscripts 'f' and 'a' stand for model forecast and analysis, respectively. All other symbols have the same meaning as in the previous section, and the procedure to simulate the model noise $\boldsymbol{\eta}_i$ is discussed in the next section.

To derive the analysis equation, we define first the matrix holding the ensemble members at time t_k , $\mathbf{x}_i(t_k) \in \mathfrak{R}^n$

$$\mathbf{X}(t_k) = (\mathbf{x}_1(t_k), \mathbf{x}_2(t_k), \dots, \mathbf{x}_N(t_k)) \in \mathfrak{R}^{n \times N}. \quad (8)$$

In practice, the ensemble size N is much smaller than the dimension of the model state vector n . The ensemble mean is stored in the vector $\bar{\mathbf{x}}(t_k) \in \mathfrak{R}^n$:

$$\bar{\mathbf{x}}(t_k) = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i(t_k) \quad (9)$$

Let us note the perturbation of an ensemble member as:

$$\tilde{\mathbf{x}}_i(t_k) = \mathbf{x}_i(t_k) - \bar{\mathbf{x}}(t_k), \quad i \in [1, N]. \quad (10)$$

The ensemble perturbation matrix $\mathbf{X}'(t_k)$ is then written as

$$\mathbf{X}'(t_k) = (\tilde{\mathbf{x}}_1(t_k), \tilde{\mathbf{x}}_2(t_k), \dots, \tilde{\mathbf{x}}_N(t_k)) \in \mathfrak{R}^{n \times N}. \quad (11)$$

The ensemble forecast error covariance matrix $\mathbf{B}_e(t_k) \in \mathfrak{R}^{n \times n}$ is obtained from this ensemble perturbation matrix:

$$\mathbf{B}_e(t_k) = \frac{\mathbf{X}'(t_k)(\mathbf{X}'(t_k))^T}{N-1}. \quad (12)$$

The matrix $\mathbf{B}_e(t_k)$ can be also rewritten in terms of individual perturbations as:

$$\mathbf{B}_e = \frac{1}{N-1} \sum_{i,j} \tilde{\mathbf{x}}_i \tilde{\mathbf{x}}_j^T. \quad (13)$$

Using the same notation as in the previous section, we define the matrix of perturbed observations as:

$$\mathbf{Y}_k = (\mathbf{y}(t_k) + \boldsymbol{\epsilon}_1(t_k), \mathbf{y}(t_k) + \boldsymbol{\epsilon}_2(t_k), \dots, \mathbf{y}(t_k) + \boldsymbol{\epsilon}_N(t_k)) \in \mathfrak{R}^{m_k \times N}, \quad (14)$$

150 where $\epsilon_j(t_k) \in \mathfrak{R}^{m_k}$ are observation perturbation vectors at time t_k generated by random Gaussian numbers characterized by a zero mean distribution and a standard deviation equal to the observational error $(r\sigma_y(t_k))^2 \in \mathfrak{R}^{m_k}$ at time t_k :

$$\epsilon_j(t_k) \sim \mathcal{N}(0, (r\sigma_y(t_k))^2), \quad j \in [1, N]. \quad (15)$$

The observation error covariance matrix \mathbf{R}_k is defined as in the 4D-Var version by Eq. (3).

155 The analysis equation in the ensemble Kalman filter stochastic formulation, i.e. with perturbed observations (Houtekamer and Mitchell, 2001; Evensen, 2003), is written as

$$\mathbf{X}^a(t_k) = \mathbf{X}^f(t_k) + \mathbf{B}_e(t_k) \mathbf{H}_k^T [\mathbf{H}_k \mathbf{B}_e(t_k) \mathbf{H}_k^T + \mathbf{R}_k]^{-1} \mathbf{D}_k, \quad (16)$$

where $\mathbf{X}^a(t_k)$ is the analysis ensemble matrix, $\mathbf{X}^f(t_k)$ is the forecast ensemble matrix and $\mathbf{D}_k = \mathbf{Y}_k - \mathbf{H}_k \mathbf{X}^f(t_k)$ is the ensemble innovation matrix at time t_k .

A widely known issue with the EnKF method is its tendency to produce analyses with noisy spatial correlations at 160 large distances in the analysis covariance. This is due to the finite and relatively small size of the ensemble compared to the size of the model state vector (Houtekamer and Mitchell, 2001). To filter out this noise, we follow the method proposed by S.E. Cohn and R. Ménard in 1997 and applied in many EnKF systems (e.g. Houtekamer and Mitchell, 2001; Hamill et al., 2001; Constantinescu et al., 2007b; Fertig et al., 2007; Sakov et al., 2010). The method consists in using the Schur (element-wise) product of the ensemble covariance matrix with a compact support correlation function, here 165 denoted ρ . The function ρ used in this study is the 5th-order piecewise rational function of Gaspari and Cohn (1999) which is isotropic and decreases monotonically with distance depending on the correlation length scale L_{loc} . The function ρ is positive only for distances that are less than $2L_{loc}$ and zero otherwise. We applied this procedure to both horizontal and vertical correlations, using the compact support correlation functions ρ_h and ρ_v , with correlation length scales L_{loc}^h and L_{loc}^v , respectively. The choice of these parameters is discussed in Sect. 3.2.2.

170 The actual implementation of the analysis equation is thus written as follows (omitting the time index):

$$\mathbf{X}^a = \mathbf{X}^f + \rho_v^m \circ \rho_h^m \circ \mathbf{B}_e \mathbf{H}^T [\mathbf{H}(\rho_v^o \circ \rho_h^o \circ \mathbf{B}_e) \mathbf{H}^T + \mathbf{R}]^{-1} \mathbf{D}, \quad (17)$$

where the notation $\mathbf{A} \circ \mathbf{B}$ denotes the Schur product between two matrices \mathbf{A} and \mathbf{B} . The indexes m and o are introduced to show that the dimension of the matrix ρ corresponds to the model and observation space dimensions, when the Schur product is applied to the matrix $\mathbf{B}_e \mathbf{H}^T$ and $\mathbf{H} \mathbf{B}_e \mathbf{H}^T$, respectively. The observational operator \mathbf{H} involves the vertical 175 and horizontal interpolations on the model grid. And the function ρ has a length scale which is much broader than the interpolation distances. Hence, the order of the interpolation and the Schur product can be interchanged without significant loss of accuracy. So, Eq. (17) is written approximately as:

$$\mathbf{X}^a \approx \mathbf{X}^f + \rho_v^m \circ \rho_h^m \circ \mathbf{B}_e \mathbf{H}^T [\rho_v^o \circ \rho_h^o \circ \mathbf{H} \mathbf{B}_e \mathbf{H}^T + \mathbf{R}]^{-1} \mathbf{D}, \quad (18)$$

The application of the Schur product to the ensemble covariances has several advantages. First, the correlation function 180 filters out small and noisy correlations related to observations at large distances. Second, it allows the EnKF to perform reasonably well even with a small number of ensemble members. Houtekamer and Mitchell (2001) stated that the use of the Schur product improves the conditioning of the matrices $\mathbf{B}_e \mathbf{H}^T$ and $\mathbf{H} \mathbf{B}_e \mathbf{H}^T$. They also argued that the Schur product tends to reduce and smooth the effect of observations at intermediate distances.

In practice, the forecast error covariance matrix \mathbf{B}_e is never computed explicitly. The ensemble representation (12) is used instead:

$$\mathbf{X}^a = \mathbf{X}^f + \rho_v^m \circ \rho_h^m \circ \mathbf{X}' (\mathbf{H}\mathbf{X}')^T [\rho_v^o \circ \rho_h^o \circ (\mathbf{H}\mathbf{X}') (\mathbf{H}\mathbf{X}')^T + \mathbf{R}]^{-1} \mathbf{D}. \quad (19)$$

In our system the number of observations per model time step is rather small, allowing the inversion of the innovation matrix $[\mathbf{H}\mathbf{B}_e\mathbf{H}^T + \mathbf{R}]$ for a reasonable CPU cost.

2.4 Ensemble Initialization and Model Error Generation

Several authors have reported the problem of the EnKF divergence: the decreasing ability of the filter to correct the ensemble state towards the observations after a certain number of assimilation cycles (Houtekamer and Mitchell, 1998; Hamill, 2006). The exact cause of this filter divergence is not entirely clear but two main reasons have been raised: (i) the variance of the ensemble forecast error becomes too small when the effect of model error in the prediction is not considered (Lorenç, 2003); and (ii) the finite sample size causes a mismatch between estimated and true error variance (Houtekamer and Mitchell, 1998). A common method preventing the filter divergence is to increase artificially the ensemble covariance. In our system, the error covariance is increased by adding a state-wide model error η_i (Eq. 7) at every model time step to each ensemble forecast.

Let us first provide a short description of the method to formulate the variational background error covariance matrix, as proposed by Courtier et al. (1998) and adopted to the 4D-Var version by Errera and Ménard (2012). In this study, the method is used not only to compute the matrix \mathbf{B}_0 in the 4D-Var system (Eq. 5) but also to compute the initial ensemble and the model error in the EnKF system. The EnKF uses flow-dependent ensemble forecast error covariance (Eq. 12) evolving in time with the ensemble. On the contrary, 4D-Var reinitializes the background error covariance every 24 h.

As stated in Errera and Ménard (2012), the formulation of the background error covariance matrix is crucial for any variational data assimilation system. The matrix \mathbf{B}_0 should be sufficiently compact to be implemented numerically and sufficiently complex to represent adequately realistic error covariances of the first guess field. To achieve this goal, there are several approaches. The proposed method expresses the spatial correlations on a spherical harmonic basis (Courtier et al., 1998). It is based on the fact that on such basis, homogeneous and isotropic horizontal correlations are represented by a diagonal matrix with repeating values on the diagonal (for the same zonal wave number).

In this case, the operator \mathbf{L} introduced in Eq. (4) is defined by:

$$\mathbf{L} = \mathbf{\Sigma}\mathbf{S}\mathbf{\Lambda}^{1/2}, \quad (20)$$

where $\mathbf{\Sigma}$ is the (diagonal) background error standard deviation matrix; $\mathbf{\Lambda}^{1/2}$ is the spatial correlation matrix defined on a spherical harmonic basis hence diagonal; and \mathbf{S} is the spectral transform operator from the spectral space to the model space.

In the present study, the spatial correlation matrix considers Gaussian correlations in the horizontal and in the vertical directions with correlation length scales fixed to $L_0^h=800$ km horizontally and $L_0^v = 1$ level vertically. The vertical profile of the background standard deviation matrix is estimated using the NMC method (Parrish and Derber, 1992) and is shown on Fig. 1. This profile is used for every point of the horizontal grid.

The operator \mathbf{L} is also used to generate the initial deviation $\tilde{\mathbf{x}}_i(t_0)$ and the model error $\boldsymbol{\eta}_i(t_k)$ of the EnKF system. In the case of the initial deviation, this ensures that at the initial time, both EnKF and 4D-Var systems have the same error statistics. For an initial deviation, we have:

$$\tilde{\mathbf{x}}_i(t_0) = \mathbf{L}\boldsymbol{\zeta}_i(t_0), \quad i \in [1, N], \quad (21)$$

while for a model error, we have

$$\boldsymbol{\eta}_i(t_k) = \alpha \mathbf{L}\boldsymbol{\psi}_i(t_k), \quad i \in [1, N], \quad (22)$$

where $\boldsymbol{\zeta}_i(t_0)$ and $\boldsymbol{\psi}_i(t_k)$ are normally distributed random numbers with zero mean and variance equal to 1, defined in the spectral space; and where α is a *model error parameter* smaller than 1.

Normalizing to a normal distributed random deviate is exactly what should also be done for the simulation of the model error term. In the theory of the Kalman filter (Kalman, 1960), the model error $\boldsymbol{\eta}_i$ is uncorrelated with the observation error and with the initial condition error. The model error and the analysis error must remain uncorrelated at later times. Hence, the perturbations $\boldsymbol{\psi}_i$ should be different at each model time step.

The algorithm to generate EnKF state perturbations is then identical to the algorithm of the 4D-Var background error covariance generation. However, the operator \mathbf{L} is applied to the normally distributed random deviate $\boldsymbol{\zeta}_i$ (21) rather than to the control variable $\boldsymbol{\xi}$ (4).

2.5 Methodology of tuning the parameters r and α

As stated before, the EnKF system has two adjustable parameters: r and α . The observation error parameter r and the model error parameter α are adjusted statistically using a χ^2 diagnostic introduced by Ménard and Chang (2000) for the Kalman filter. This diagnostic compares the innovation vector \mathbf{d} with the innovation covariance matrix $\mathbf{S} = \mathbf{H}\mathbf{B}_e\mathbf{H}^T + \mathbf{R}$ (Eq. 16) using a Mahalanobis norm (Talagrand, 2010). Specifically, at every analysis time step k , the value of χ_k^2 is computed as follows:

$$\chi_k^2 = \mathbf{d}_k^T \mathbf{S}_k^{-1} \mathbf{d}_k. \quad (23)$$

An assimilation system is said to be optimal when $\langle \chi_k^2 \rangle$ is equal to the number of observations m_k at time t_k where $\langle \cdot \rangle$ denote the statistical expectation. Since the number of observations per time step is relatively large, i.e. about 1100 in our case, we can approximate $\langle \chi_k^2 \rangle$ by a realization of χ_k^2 for a given set of observed values (i.e. for a realization of the observation error).

As shown by Ménard and Chang (2000), modifying the model error parameter changes the trend (or slope) of χ_k^2 over time, while modifying the observational error parameter r changes the mean value of χ_k^2 . Since these two parameters have distinguishable effects on the time series of χ_k^2 (mean and trend), they can be tuned separately, as summarized by Khattatov et al. (2000):

1. Run the assimilation system and monitor χ_k^2/m_k . If its value increases (decreases) consistently with time, increase (decrease) α . This procedure is repeated until the mean value of χ_k^2/m_k does not show a trend in its time series.
2. If the average value of χ_k^2/m_k is larger (smaller) than 1, increase (decrease) the observation error scaling factor r .

3 Experimental set-up

3.1 Observations

The data used in this study are ozone profiles given by EOS Aura-MLS version 2.2 (Froidevaux et al., 2008). The observations of ozone cover the latitude range between 82°S and 82°N with an along-track separation of around 165 km
255 between consecutive scans. Around 3500 vertical scans are performed every day. Ozone profiles have a vertical resolution around 3 km in the stratosphere and they are valid for scientific studies between 215 and 0.02 hPa. However, ozone data are not assimilated above 1 hPa because the tracer assumption is not valid above this pressure level. The observational error σ_y (Eq. 3) is set from the instrumental error provided with each observation and increased if necessary to represent at least 5% of the observation value. This accounts for the representativeness error because smaller errors would give too
260 large a weight to observations.

3.2 Calibration of the systems

To perform a proper comparison between the 4D-Var and EnKF, we must calibrate both systems in such a way that they use the same error statistics. Our starting point is the calibration of the error covariance matrix \mathbf{B}_0 used by the 4D-Var system. This is realized through a calibration of the spatial correlation operator \mathbf{L} , i.e. the background error spatial
265 correlation matrix $\mathbf{\Lambda}$ and the background error standard deviation matrix $\mathbf{\Sigma}$ (Eq. 20). The calibrated operator \mathbf{L} is then used in the EnKF system, where the model error parameter α and the observation error parameter r are estimated using the χ^2 -diagnostic. Once the parameter r is estimated, its value is passed to the 4D-Var system for a final test of performance.

3.2.1 4D-Var

The matrix $\mathbf{\Sigma}$ of the 4D-Var system has been calibrated using the NMC method (Parrish and Derber, 1992; Rabier et al.,
270 1998; Bannister, 2008). For this purpose, a 6-month assimilation experiment (May-October 2008) has been performed assuming a matrix $\mathbf{\Sigma}$ setup as 30% of the background field and a matrix $\mathbf{\Lambda}$ assuming Gaussian correlations with correlation length scales $L_0^h = 800$ km horizontally and $L_0^v = 1$ level vertically. The NMC method assumes that the \mathbf{B}_0 matrix may be estimated by the difference between pairs of forecasts of different lead times but same validity times. In meteorology, the forecast pairs have typically 24 and 48 hours lead times. In our case the forecast pairs have 0 and 24 hours lead times, i.e.
275 the difference between the forecast pairs is equivalent to the analysis increments of the 4D-Var system. Indeed contrary to the meteorological case, there is no need in chemistry to perform a 24 h forecast to balance the model fields.

The calibration of \mathbf{B}_0 with the NMC method has been computed for several periods in 2008: May-July, August-October and May-October. No significant differences in the estimated $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$ have been found. So the period May-October is used in this study. Moreover, to parametrize the diagonal values of the matrix $\mathbf{\Sigma}$, two variants of it have been tested
280 using the NMC method. They assume that the background error standard deviations are defined by (1) a one-dimensional pressure profile and (2) a two-dimensional latitude-pressure field. The 4D-Var assimilation experiments using these two parameterizations of $\mathbf{\Sigma}$ have not shown important differences in results. So the one-dimensional profile of $\mathbf{\Sigma}$ (see Fig. 1) has been used to compare the 4D-Var and EnKF systems. We have also estimated the correlation matrix $\mathbf{\Lambda}$ with the NMC method. But the differences between the 4D-Var assimilation considering the NMC $\mathbf{\Lambda}$ and the Gaussian $\mathbf{\Lambda}$ (where both

285 experiments use the NMC Σ) have not shown an important difference in results. So the Gaussian Λ has been kept to ease its implementation in the EnKF system – specifically its explicit formulation of compact support correlation length scales.

3.2.2 EnKF

Once the matrices Λ and Σ are calibrated in the 4D-Var system, the resulting operator L is passed to the EnKF system. As explained in Sect. 2.3, the EnKF uses as localization method a Schur product with a compact support correlation
 290 function. The use of Schur product reduces the resulting correlation length scales. In order to maintain the correlations of the EnKF analysis comparable to those of the 4D-Var system, a different setting of the correlation length scales is adopted to generate the model error (Eq. 22). Let C be a matrix resulting from the Schur product of two matrices A and B : $C = A \circ B$. If the correlation length scales of A and B are, respectively L_A and L_B , the correlation length scale of C is given by (Gaspari and Cohn, 1999):

$$295 \quad \frac{1}{L_C^2} = \frac{1}{L_A^2} + \frac{1}{L_B^2}. \quad (24)$$

In our case, L_A corresponds to the correlation length scale L_{loc} of the compact support correlation function ρ and L_B corresponds to the correlation length scale of the forecast ensemble covariance matrix B_e , denoted in the following by L_e . Similarly, L_C corresponds to the correlation length scale of the analysis ensemble covariance matrix, denoted in the following by the effective correlation length scale L_{eff} . As we would like to maintain the L_{eff} equal to the
 300 Gaussian correlation length scales used in the 4D-Var (i.e. $L_0^h=800$ km and $L_0^v=1$ level), we need to set L_{loc} and L_e such that $L_{eff} = L_0$. First, reasonable values for the localized correlation length scales were chosen: $L_{loc}^h = 2000$ km and $L_{loc}^v = 1.5$. In this configuration, the correlation length scales used to generate the model error in the EnKF are defined by $L_e^h = 872$ km and $L_e^v = 1.3$ model level.

The next step in the calibration of the EnKF is the tuning of α and r using a χ^2 -diagnostic (see Sect. 2.5). Figure 2
 305 shows the time evolution of χ_k^2/m_k for three EnKF runs. The first run assumed $r = 1$ and $\alpha = 0$, resulting in χ_k^2/m_k at ~ 2.8 initially and growing quickly during the next days. The model error parameter α was then adjusted by trial and error until the time-series of χ_k^2/m_k displayed no trend, a condition met by a run using $\alpha = 0.025$. This second run still resulted in χ_k^2/m_k too large, around 3. A second series of trial and error adjustments for the observation error parameter r led to the final run for the EnKF calibration: setting $r = 1.65$ resulted in analyses with χ_k^2/m_k close to 1.

310 Figure 3 displays the Observation minus Forecast (OmF) statistics, biases and standard deviations with respect to the assimilated MLS data, for these three EnKF experiments with $[\alpha = 0, r = 1]$, $[\alpha = 0.025, r = 1]$, and $[\alpha = 0.025, r = 1.65]$. A clear improvement is found after the tuning of α , i.e. the presence of the model error term is essential for the EnKF to function properly. The impact of the tuning of r is not so visibly marked, however, the final EnKF experiment using α and r parameters both tuned shows systematically better results than the experiment where only α is tuned.
 315 Overall, this illustrates an important sensitivity of the EnKF to the adjustable error parameters.

The value of $r = 1.65$ has been passed to the 4D-Var system for a final experiment to ensure the use of common observation error statistics. Note that no significant differences in the OmF statistics have been found between the 4D-Var analysis with $r = 1$ and $r = 1.65$ (not shown). Thus, while the 4D-Var requires important work to develop an adjoint operator, the tuning of error parameters does not require large efforts in the context of stratospheric chemistry.

320 3.3 Numerical performance

We tried to configure the EnKF and 4D-Var systems to allow comparable total CPU costs. Preliminary experiments with the 4D-Var system show that the 4D-Var performs reasonably well using about 20 iterations. Accounting for the adjoint model integration in the 4D-Var, we have chosen for the EnKF an ensemble size of 40 members.

In terms of numerical performance, the 4D-Var requires about 750 s on a single processor to integrate one assimilation
325 window of 24 h. The EnKF algorithm consists in two separate phases: the ensemble propagation and the analysis (Eq. 19). The analysis phase of the EnKF requires 550 s to perform 48 analyses, covering the period of 24 h, (48 analyses correspond to the model time step of 0.5 h) and 500 s to propagate the ensemble during the same period on a single processor. The actual EnKF configuration allows to solve Eq. 19 on multiple processors, which helps to gain an important wallclock time: the analysis phase requires 100 s on 16 processors. Note that the computation of the Kalman gain in our EnKF is performed
330 using Cholesky decomposition where the full observation vector is considered at a given time step. No simplification is used to compute the inversion of the innovation matrix $[\mathbf{H}\mathbf{B}_e\mathbf{H}^T + \mathbf{R}]$ or the matrix $\mathbf{B}_e\mathbf{H}^T$. The CPU cost of the EnKF can be improved using local domain decomposition and integration of the ensemble members on different processors in parallel. These two tasks will be a subject of our future work.

4 Results and discussion

335 In this section, we discuss the performance of the EnKF and 4D-Var after calibration as described in the previous section. The performance is evaluated using standard OmF statistics i.e. the average of the differences (bias) between observations and forecasts, as well as their standard deviation. Here we use 24 h forecasts and the assimilated MLS O₃ profiles as observation data. This statistical diagnostic is performed in five different latitudinal bands covering the globe. Fig. 4 shows the OmF statistics for the period September-October 2008. The OmF errors are computed in percents from the observation
340 values. In order to evaluate if the difference between OmF errors provided by EnKF and 4D-Var are statistically significant for confidence interval of 95%, the Student test (T-test) and the Fisher test (F-test) are computed for biases and standard deviation errors, respectively. The application of these statistical tests are explained in the Appendix. Their results are depicted on Fig. 4 by green/red stars, which correspond to significant/not significant differences for the chosen confidence interval.

345 As seen from the figure, both systems exhibit a small OmF bias, generally less than 2 % in the pressure range 1-100 hPa. The OmF standard deviations are generally less than 10 % in the range 1-50 hPa and increase to a maximum of ~50 % in the tropical upper troposphere (200 hPa). In the Upper Troposphere-Lower Stratosphere (UTLS), the limb sounding observations by Aura-MLS are less precise and less accurate than in the mid-stratosphere (Froidevaux et al., 2008) which explains the increase in the bias and OmF standard deviation below 50 hPa. According to the T- and F-tests, the OmF
350 of the EnKF and 4D-Var 24 h forecasts are statistically equivalent at almost all levels and in all latitude bands. We note however that decreasing the value of the confidence interval from 95 % to 90 %, for example, reveals that following the T and F-tests, the OmF of the two systems are different at pressure levels around 4 and 70 hPa in the South Pole region. We note also that in this same region the OmF bias is much larger for both systems in the lower and upper stratosphere (but still smaller than 6 %) . The biases of two systems concord to be of the same sign and roughly of the same magnitude.

355 This indicates that the origin of the bias is common for both systems.

Fig. 5 shows a time series of the OmF bias and standard deviation in the upper stratosphere (1–10 hPa) above the South Pole region. The bias and standard deviations of both systems are very similar and remain stable between June and September, with a negligible bias and standard deviations around 6–7 %. From September to November, these values increase up to 7 % for the bias and 12 % for the standard deviation. During that period, the dynamics becomes relatively
360 active in the upper stratosphere above the South Pole due to the breakup of the vortex. The ERA-Interim wind fields used to drive the transport model may be either insufficiently accurate or their update period (6 hours) may be too long to allow the systems to provide accurate 24 hours forecasts.

Fig. 6 shows the time series of the OmF bias and standard deviation in the lower stratosphere (10–100 hPa) above the South Pole region. Again, both systems have similar OmF departures. Between May and September, the biases are
365 generally less than 2 % from the observation values, and the standard deviations are less than 10 % until the end of August. The standard deviations increase quickly during the first days of September and reach maximum values – around 15–17 % – in mid-September, the 4D-Var providing values slightly lower than those from the EnKF. At the beginning of November, the standard deviations have decreased back to pre-September levels. In September, the bias of both experiments also slightly increases up to 4 % where 4D-Var shows again values slightly lower than those by the EnKF. In mid-October,
370 the bias becomes negative but with values lower than -2 %. Since the months of September and October are precisely the period of photochemically-driven ozone destruction, we attribute these degradations in the OmF series to the absence of chemistry in our simplified model: during the ozone hole period, the tracer transport approximation is clearly not adequate. In this situation, the 4D-Var delivered a slightly better performance than the EnKF. This may be due to the assimilation window of 24 hours used by the 4D-Var, compared with the sequential assimilation of the EnKF.

375 Outside the ozone hole period/region and based on the OmF statistics between MLS and the analysis, no significant differences have been found between the systems. OmF statistics have also been computed against independent observations by Envisat/MIPAS (Raspollini et al., 2013). Although the OmF statistics differ slightly, no statistical differences between the EnKF and 4D-Var systems have been found either (not shown). Hence the slightly different OmF statistics are only due to the differences between the MLS and MIPAS datasets. We thus conclude that both systems are statistically
380 equivalent in the observation space except during the ozone hole period, where the 4D-Var delivers analyses with slightly smaller OmF departures than the EnKF.

However, individual analyses provided by the two systems can exhibit larger differences than those given by OmF statistics. Figure 7 (upper row) shows the ozone distribution at 54.6 hPa on September 15, 2008 by both systems and their differences. Although both analyses display similar patterns, they are clearly not identical. But looking at the monthly
385 averaged maps (Fig. 7, lower row) these difference become small. Hence while there are some differences between the analyses by the two systems, these are not systematic and can be attributed to noise in the analyses.

Finally, let us come back to the time series of the χ^2 diagnostic for the EnKF system (Fig. 2) A small but sharp increase occurs during the first days of September. We attribute this jump to the onset of photochemically-driven ozone depletion. Figure 8 shows the formation of the ozone hole through ozone analyses delivered by the EnKF system at 54.6 hPa (model
390 level 22) in the southern hemisphere from August 29 to September 20 (one snapshot every two days). While it takes several days to see a clear ozone hole above Antarctica (even on September 20 ozone depletion is not complete yet),

ozone depletion started during the first days of September i.e. exactly during the sudden growth of χ^2 . If this growth is really due to a missing process in our model (in this case the ozone polar chemistry), then χ^2 may be used as a tool to monitor the model error. Note that although the time series of the standard deviation in the OmF also increases during the formation of the ozone hole (see Fig. 6), the growth in the standard deviation is smoother than displayed by the χ^2 and thus provides a less clear signal. Future work will extend the comparison to the full BASCOE CTM including the ozone polar chemistry, and if our explanation is correct this sharp increase should disappear from the χ^2 time series.

5 Conclusions

The first aim of this paper was to present the implementation of the EnKF method in the BASCOE system. This system was originally based on 4D-Var, and our motivation was to bypass the development and maintenance of an adjoint model. The new EnKF version of BASCOE was developed accounting for two adjustable parameters: the parameter α controlling the model error term of the EnKF and the parameter r controlling the observational error. These two parameters have been adjusted based on the monitoring of a χ^2 -test measuring the misfit between the control variable and the observations. In this study, we have turned off the chemistry in the CTM of BASCOE and have considered only ozone transport. This configuration allowed considerably faster execution of both systems and a large number of assimilation experiments.

The second aim of this paper was to properly compare the EnKF and 4D-Var despite the fundamental differences between the two methods. To this end, we have used the same numerical model, an identical set of observations, an identical observation operator \mathbf{H} and the same observation error covariance matrix \mathbf{R} . Furthermore, the background error covariances have been carefully calibrated in both systems. First, the background error covariance matrix \mathbf{B}_0 of the 4D-Var system has been calibrated using the NMC method. Two components of \mathbf{B}_0 , i.e. the background error spatial correlation matrix \mathbf{A} and the background error standard deviation matrix $\mathbf{\Sigma}$, have been transferred to the EnKF system to generate the initial ensemble and the model error term. The background error statistics were then carefully designed to use the same correlation models and equivalent length-scales in both systems. The EnKF parameters α and r have been calibrated using the χ^2 test. The value of r has been passed to the 4D-Var system to ensure that both systems use identical observation error covariances. Despite a straightforward implementation of the EnKF numerical algorithm, the resulting EnKF version of BASCOE was shown to be extremely sensitive to the parameter values. Thus, an accurate adjusting procedure of these parameters is of great importance to the performance of the system.

We note that the EnKF has a model error term whereas the 4D-Var is considered as a strong constraint problem. It may thus be argued that their comparison cannot be totally fair. Note however that the EnKF is not able to work properly without a model error and the implementation of the model error term was relatively fast. On the other hand, the effort required to implement a weak constraint in the 4D-Var seems quite large. We thus have considered both methods in their original form. Consequently, we have not attempted to introduce the background error covariance localization within 4D-Var. Such localization would have resulted in a comparison with a form of 4D-Var that is never used in practice.

The two systems were compared through the assimilation of Aura-MLS ozone observations for the period May-December 2008, thus comprising the polar night and ozone hole periods. We focused our attention on the lower and middle stratosphere where the ozone lifetime is much larger than the observational update frequency.

We have assessed the performance primarily in terms of observation minus 24h forecast statistics and found that the analyses provided by the two systems are significantly similar for confidence interval of 95%, with biases smaller than 5 % and standard deviation errors smaller than 10% in most of the stratosphere. In September-October, the two systems display increase in their OmF above South Pole due to (1) the vortex breakup in the upper stratosphere and (2) the ozone hole formation in the lower stratosphere. The degradation of the OmF in these cases are attributed to (1) inaccuracies in the modelling of the dynamics and (2) the missing of ozone hole chemistry in the model.

Since the biases are markedly similar, they have most probably the same causes: these can be deficiencies in the model and in the observation dataset. The remarkably similar performance also shows that in the context of stratospheric transport, the choice of the assimilation method can be based on application-dependent factors, such as CPU cost or the ability to generate an ensemble of forecasts.

The BASCOE 4D-Var system can provide analyses taking explicitly into account stratospheric chemistry, in the forward as well as the adjoint model. The EnKF system presented here accounts only for stratospheric transport, not chemistry. The application of the EnKF method to the full-chemistry model may require a careful tuning procedure for each chemical species, a task that can be time consuming. Hence an adaptive calibration procedure of the error covariances (similar to Anderson (2009) or Li et al. (2009)) should be implemented. The implementation of EnKF with chemistry is ongoing and will be reported in future studies.

Appendix A Statistical tests to compare the OmF errors

A1 The Student Test

We use the two-sample significance Student test (T-test; see e.g. Snedecor and Cochran, 1989) to compare two mean OmF residuals $\bar{\mathbf{d}}(l)$ computed by the EnKF and the 4D-Var, where the bar denotes the time-averaged OmF residual value at level l . This test is used in our case under the assumption that the two samples have the same size and variance. The t-statistics is computed at each level l as follows:

$$t(l) = \frac{\bar{\mathbf{d}}_1(l) - \bar{\mathbf{d}}_2(l)}{S_{\mathbf{d}_1\mathbf{d}_2}(l)} \cdot \sqrt{\frac{2}{n(l)}}, \quad (\text{A1})$$

where $n(l)$ is the size of the samples, $S_{\mathbf{d}_1\mathbf{d}_2}(l) = \sqrt{\frac{1}{2}(\sigma_1^2 + \sigma_2^2)}$ is the grand or pooled standard deviation and $\sigma_1(l)$ and $\sigma_2(l)$ are the OmF standard deviations for the EnKF and the 4D-Var at level l , respectively.

Then, for a given value of significance level α (typically set to 5%), the hypothesis that two means are statistically equal is rejected if:

$$|t(l)| > T_{1-\alpha/2, n(l)}, \quad (\text{A2})$$

where $T_{\alpha, n(l)}$ is a critical value of $t(l)$ computed as the inverse of the Student's T cumulative distribution function (cdf) for a given α and $n(l)$. The well known t-tables provide the values of $T_{\alpha, n(l)}$ only for small sample sizes $n(l)$. In practice, the Student's T-cdf is easily computed by many statistical packages like SciPy, the Matlab statistical toolbox etc.

A2 The Fisher Test

460 The Fisher or F-test (Snedecor and Cochran, 1989) is used to compare two OmF standard deviations and determine if they are significantly different. For a two-tailed significance test it is supposed that $\sigma_1^2 \neq \sigma_2^2$. The F-test statistics is computed as at each level l as follows

$$f(l) = \frac{\sigma_1^2(l)}{\sigma_2^2(l)}. \quad (\text{A3})$$

The more this ratio deviates from 1, the stronger the evidence for unequal OmF standard deviations. The hypothesis that the two OmF standard deviations are equal is rejected if

$$465 \quad f(l) > F_{\alpha/2, n(l)}, \quad (\text{A4})$$

or

$$f(l) < F_{1-\alpha/2, n(l)}, \quad (\text{A5})$$

where $F_{\alpha, n(l)}$ is the critical value of F-distribution computed as F cumulative distribution function with $n(l)$ degrees of freedom and a significance level of α . The F-cdf can be computed using the same statistical packages as for the T-cdf.

470 *Acknowledgements.* This research was financially supported at BIRA-IASB by the Belspo/ESA/PRODEX programme (PRODEX project BACCHUS). Sergey Skachko wishes to thank Laurent Bertino for his help in the numerical implementation of the inversion of the innovation matrix within the EnKF Kalman gain, as well as Mark Buehner, Takemasa Miyoshi, Emil Constantinescu and Saroja Polavarapu for helpful discussion on EnKF issues. We are grateful to NASA JPL for providing EOS Aura-MLS data and to ECMWF for providing the ERA-Interim reanalysis.

475 References

- Anderson, J. L.: Spatially and temporally varying adaptive covariance inflation for ensemble filters, *Tellus*, 61A, 72–83, 2009.
- Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances, *Q. J. R. Meteorol. Soc.*, 134, 1951–1970, 2008.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of Variational Data Assimilation and the
480 Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments, *Mon. Weather Rev.*, 138, 1550–1566, 2010a.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations, *Mon. Weather Rev.*, 138, 1567–1586, 2010b.
- 485 Constantinescu, E. M., Sandu, A., Chai, T., and Carmichael, G. R.: Ensemble-based chemical data assimilation. I: General approach, *Q. J. R. Meteorol. Soc.*, 133, 1229–1243, 2007a.
- Constantinescu, E. M., Sandu, A., Chai, T., and Carmichael, G. R.: Ensemble-based chemical data assimilation. II: Covariance localization, *Q. J. R. Meteorol. Soc.*, 133, 1245–1256, 2007b.
- Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., and Fisher, M.:
490 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, *Q. J. R. Meteorol. Soc.*, 124, 1783–1807, 1998.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P.,
495 Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137, 553–597, 2011.
- Elbern, H., Schwinger, J., and Botchorishvili, R.: Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: System configuration, *J. Geophys. Res.*, 115, doi:10.1029/2009JD011953, 2010.
- 500 Errera, Q. and Ménard, R.: Technical Note: Spectral representation of spatial correlations in variational assimilation with grid point models and application to the Belgian Assimilation System for Chemical Observations (BASCOE), *Atmos. Chem. Phys.*, 12, 10 015–10 031, 2012.
- Errera, Q., Daerden, F., Chabrilat, S., Lambert, J. C., Lahoz, W. A., Viscardy, S., Bonjean, S., and Fonteyn, D.: 4D-Var assimilation of MIPAS chemical observations: ozone and nitrogen dioxide analyses, *Atmos. Chem. Phys.*, 8, 6169–6187, 2008.
- 505 Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10 143–10 162, 1994.
- Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343–367, 2003.
- Fertig, E. J., Hunt, B. R., Ott, E., and Szunyogh, I.: Assimilating non-local observations with a local ensemble Kalman filter, *Tellus A*, 59, 719–730, 2007.
- 510 Flemming, J., Inness, A., Flentje, H., Huijnen, V., Moinat, P., Schultz, M. G., and Stein, O.: Coupling global chemistry transport models to ECMWF’s integrated forecast system, *Geoscientific Model Development*, 2, 253–265, <http://www.geosci-model-dev.net/2/253/2009/>, 2009.
- Froidevaux, L., Jiang, Y. B., Lambert, A., Livesey, N. J., Read, W. G., Waters, J. W., Browell, E. V., Hair, J. W., Avery, M. A., McGee, T. J., Twigg, L. W., Sumnicht, G. K., Jucks, K. W., Margitan, J. J., Sen, B., Stachnik, R. A., Toon, G. C., Bernath, P. F., Boone,

- 515 C. D., Walker, K. A., Filipiak, M. J., Harwood, R. S., Fuller, R. A., Manney, G. L., Schwartz, M. J., Daffer, W. H., Drouin, B. J., Cofield, R. E., Cuddy, D. T., Jarnot, R. F., Knosp, B. W., Perun, V. S., Snyder, W. V., Stek, P. C., Thurstans, R. P., and Wagner, P. A.: Validation of Aura Microwave Limb Sounder stratospheric ozone measurements, *Journal of Geophysical Research: Atmospheres*, 113, doi:10.1029/2007JD008771, <http://dx.doi.org/10.1029/2007JD008771>, 2008.
- Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, 125, 723–757, 520 1999.
- Hamill, T. M.: Ensemble-based atmospheric data assimilation, in: *Predictability of Weather and Climate*, edited by Palmer, T. and Hagedorn, T., pp. 124–156, Cambridge University Press, <http://dx.doi.org/10.1017/CBO9780511617652.007>, 2006.
- Hamill, T. M., Whitaker, J., and Snyder, C.: Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter, *Mon. Weather Rev.*, 129, 2776–2790, 2001.
- 525 Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, *Mon. Weather Rev.*, 126, 796–811, 1998.
- Houtekamer, P. L. and Mitchell, H. L.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, *Mon. Weather Rev.*, 129, 123–137, 2001.
- Kalman, R.: A new approach to linear filtering and prediction problems, *Journal of Basic Engineering*, 82D, 35–45, 1960.
- 530 Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J.: 4-D-Var or ensemble Kalman filter?, *Tellus A*, 59, 758–773, 2007.
- Khattatov, B. V., Gille, J. C., Lyjak, L. V., Brasseur, G. P., Dvortsov, V. L., Roche, A. E., and Waters, J. W.: Assimilation of photochemically active species and a case analysis of UARS data, *J. Geophys. Res.*, 104, 18 715–18 737, 1999.
- Khattatov, B. V., Lamarque, J.-F., Lyjak, L. V., Menard, R., Levelt, P., Tie, X., Brasseur, G. P., and Gille, J. C.: Assimilation of satellite observations of long-lived chemical species in global chemistry transport models, *J. Geophys. Res.*, 105, 29 135–29 144, 2000.
- 535 Lahoz, W. and Errera, Q.: Constituent Assimilation, in: *Data Assimilation: Making sense of observations*, edited by Lahoz, W., Khattatov, B., and Ménard, R., pp. 449–490, Springer, 2010.
- Li, H., Kalnay, E., Miyoshi, T., and Danforth, C.: Accounting for model errors in ensemble data assimilation, *Mon. Weather Rev.*, 137, 3407–3419, 2009.
- Lin, S.-J. and Rood, R. B.: Multidimensional Flux-Form Semi-Lagrangian Transport Schemes, *Mon. Weather Rev.*, 124, 2046–2070, 540 1996.
- Liu, J., Fung, I., Kalnay, E., Kang, J.-S., Olsen, E. T., and Chen, L.: Simultaneous assimilation of AIRS Xco₂ and meteorological observations in a carbon climate model with an ensemble Kalman filter, *J. Geophys. Res.*, 117, D05309, doi:10.1029/2011JD016642, 2012.
- Lorenc, A. C.: The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var, *Q. J. R. Meteorol. Soc.*, 129, 545 3183–3203, 2003.
- Ménard, R. and Chang, L.-P.: 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, 2000.
- Ménard, R. and Daley, R.: The application of Kalman smoother theory to the estimation of 4DVAR error statistics, *Tellus A*, 48, 221–237, 1996.
- 550 Ménard, R., Cohn, S. E., Chang, L.-P., and Lyster, P. M.: Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part I: Formulation, *Mon. Weather Rev.*, 128, 2654–2671, 2000.
- Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., van Weele, M., and Boersma, K. F.: Simultaneous assimilation of satellite NO₂, O₃, CO, and HNO₃ data for the analysis of tropospheric chemical composition and emissions, *Atmos. Chem. Phys.*, 12, 9545–9579, 2012.

- 555 Miyoshi, T., Sato, Y., and Kadowaki, T.: Ensemble Kalman Filter and 4D-Var Intercomparison with the Japanese Operational Global Analysis and Prediction System, *Mon. Weather Rev.*, 138, 2846–2866, 2010.
- Nakamura, T., Akiyoshi, H., Deushi, M., Miyazaki, K., Kobayashi, C., Shibata, K., and Iwasaki, T.: A multi-model comparison of stratospheric ozone data assimilation based on an ensemble Kalman filter approach, *J. Geophys. Res.*, 118, 3848–3868, doi:10.1002/jgrd.50338., 2013.
- 560 Parrish, D. and Derber, J. C.: The National Meteorological Center’s spectral statistical interpolation analysis system, *Mon. Weather Rev.*, 120, 1747–1763, 1992.
- Rabier, F., McNally, A., Andersson, E., Courtier, P., Undén, P., Eyre, J., Hollingsworth, A., and Bouttier, F.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions, *Quarterly Journal of the Royal Meteorological Society*, 124, 1809–1829, doi:10.1002/qj.49712455003, 1998.
- 565 Raspollini, P., Carli, B., Carlotti, M., Ceccherini, S., Dehn, A., Dinelli, B. M., Dudhia, A., Flaud, J.-M., López-Puertas, M., Niro, F., Remedios, J. J., Ridolfi, M., Sembhi, H., Sgheri, L., and von Clarmann, T.: Ten years of MIPAS measurements with ESA Level 2 processor V6 – Part 1: Retrieval algorithm and diagnostics of the products, *Atmospheric Measurement Techniques*, 6, 2419–2439, doi:10.5194/amt-6-2419-2013, <http://www.atmos-meas-tech.net/6/2419/2013/>, 2013.
- Sakov, P., Evensen, G., and Bertino, L.: Asynchronous data assimilation with the EnKF, *Tellus A*, 62, 24–29, 2010.
- 570 Sandu, A. and Chai, T.: Chemical Data Assimilation - An Overview, *Atmosphere*, 2(3), 426–463, 2011.
- Sekiyama, T. T., Deushi, M., and Miyoshi, T.: Operation-Oriented Ensemble Data Assimilation of Total Column Ozone, *SOLA*, 7, 41–44, 2011.
- Snedecor, G. and Cochran, W.: *Statistical Methods*, Iowa State University Press, 8th edn., 1989.
- Talagrand, O.: Evaluation of Assimilation Algorithms, in: *Data Assimilation: Making sense of observations*, edited by Lahoz, W., Khattatov, B., and Ménard, R., pp. 217–240, Springer, 2010.
- 575 Talagrand, O. and Courtier, P.: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, 113, 1311–1328, 1987.
- Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecasts, *J. Geophys. Res.*, 113, D20310, doi:10.1029/2008JD009991, 2008.

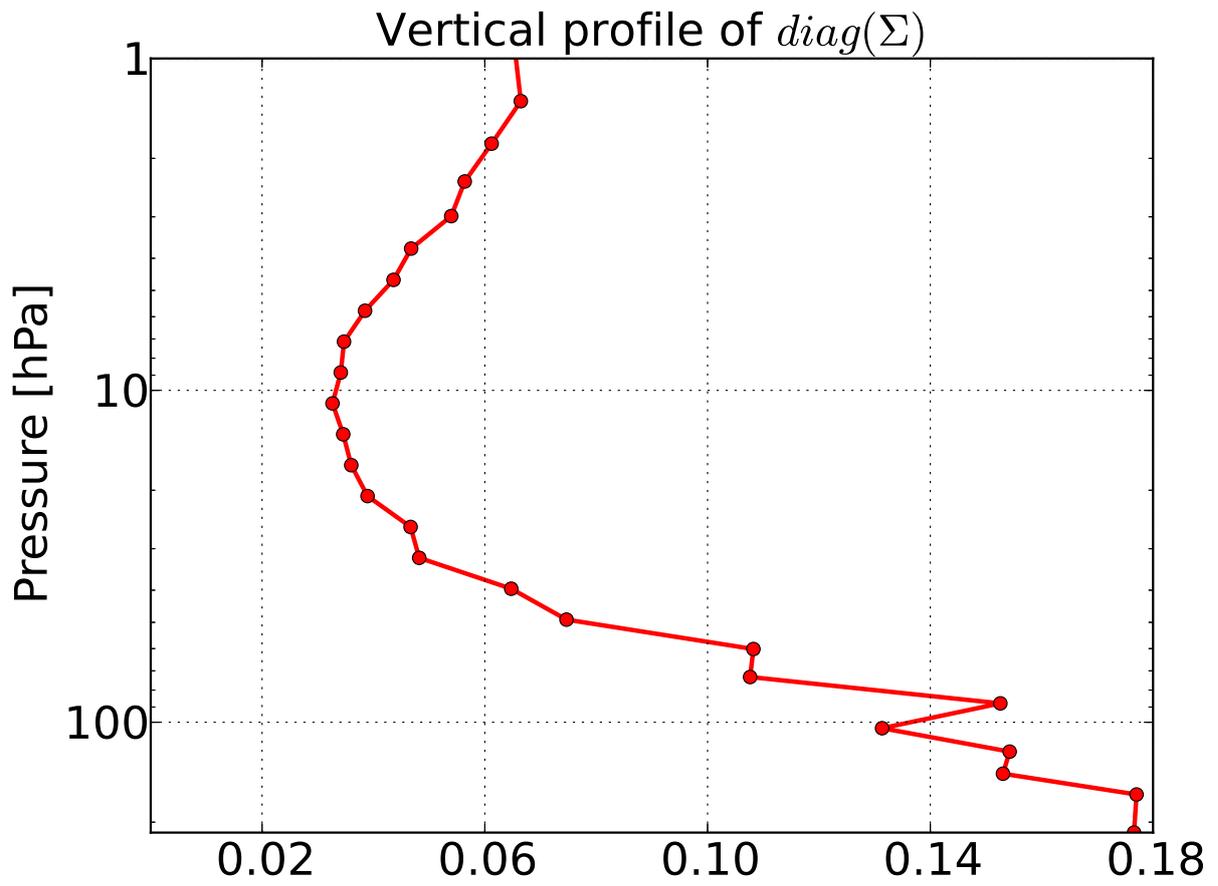


Fig. 1: Vertical profile of the background error as used for the standard deviation matrix Σ .

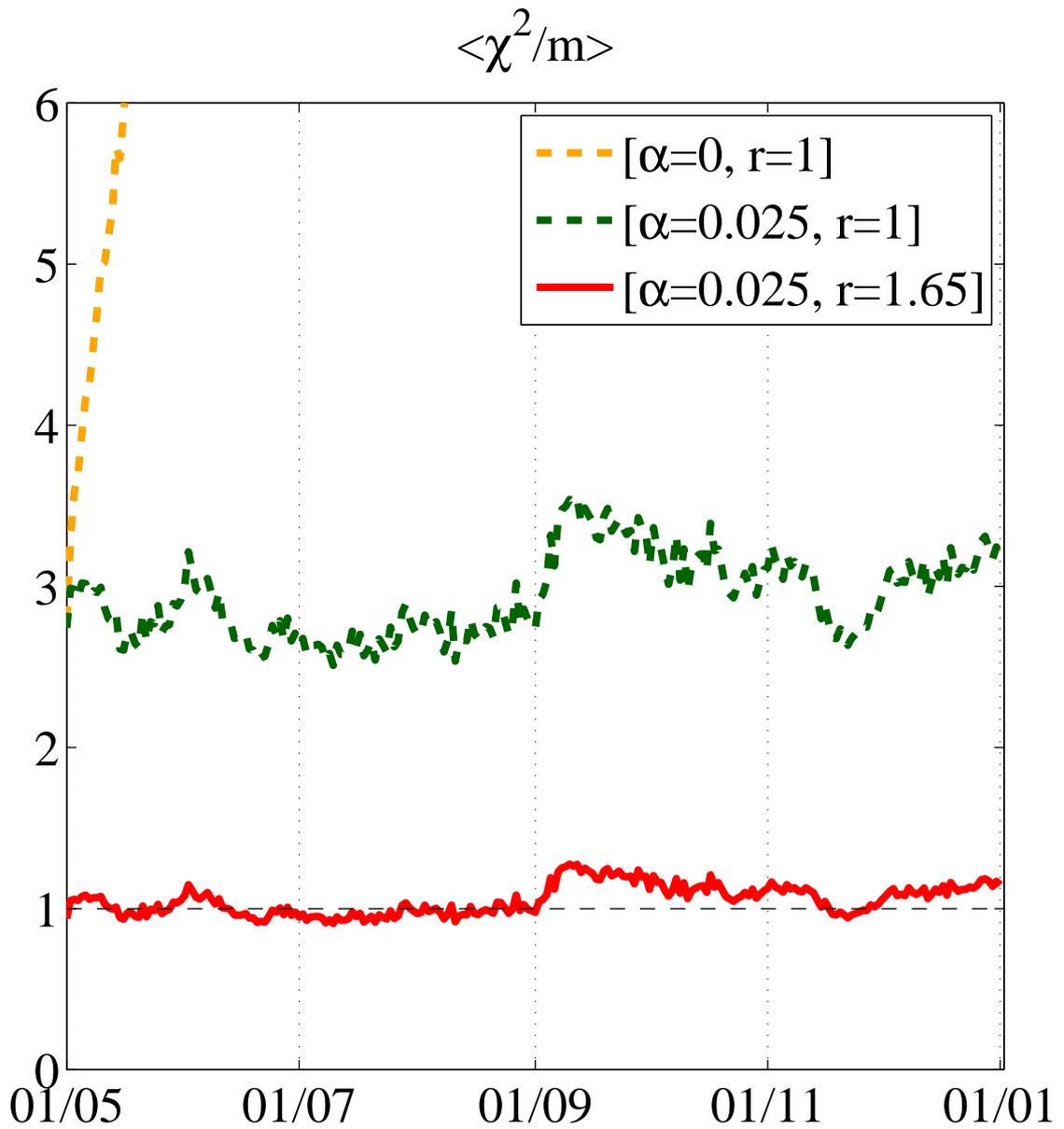


Fig. 2: $\langle \chi^2/m \rangle$ evolution for the EnKF experiments using $[\alpha = 0, r = 1]$ (orange dashed line), $[\alpha = 0.04, r = 1]$ (green dashed), and $[\alpha = 0.04, r = 1.6]$ (red solid). for the period from 00 UTC on 1 May 2008 to 00 UTC on 1 January 2009.

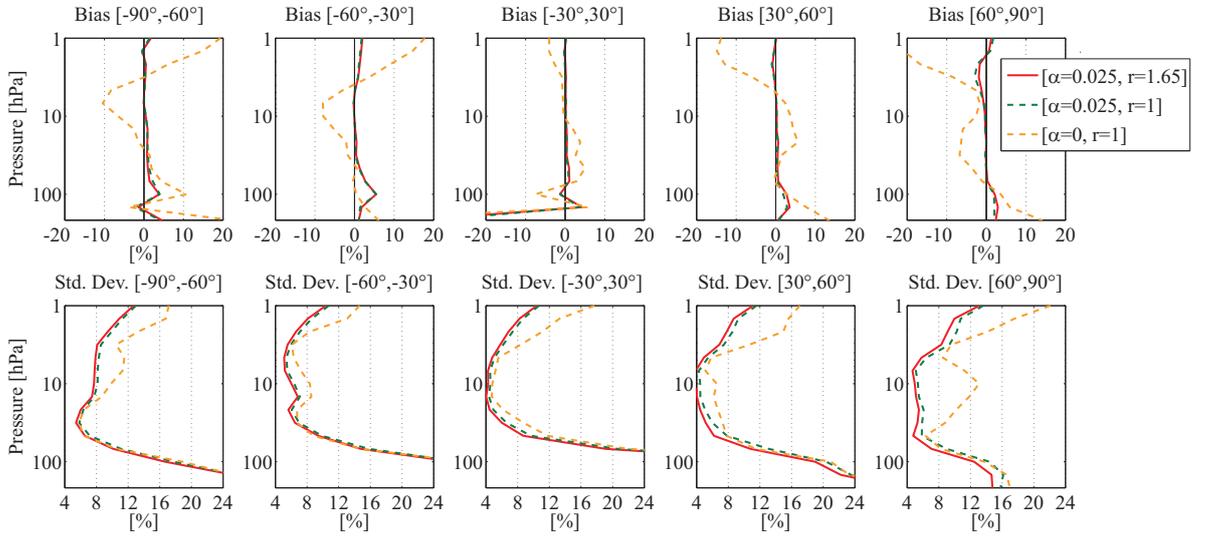


Fig. 3: OmF statistics, bias and standard deviation, with respect to the MLS data of the EnKF experiments using $[\alpha = 0, r = 1]$ (orange dashed line), $[\alpha = 0.025, r = 1]$ (green dashed), and $[\alpha = 0.025, r = 1.65]$ (red solid). for the period from 1 May to 31 June 2008.

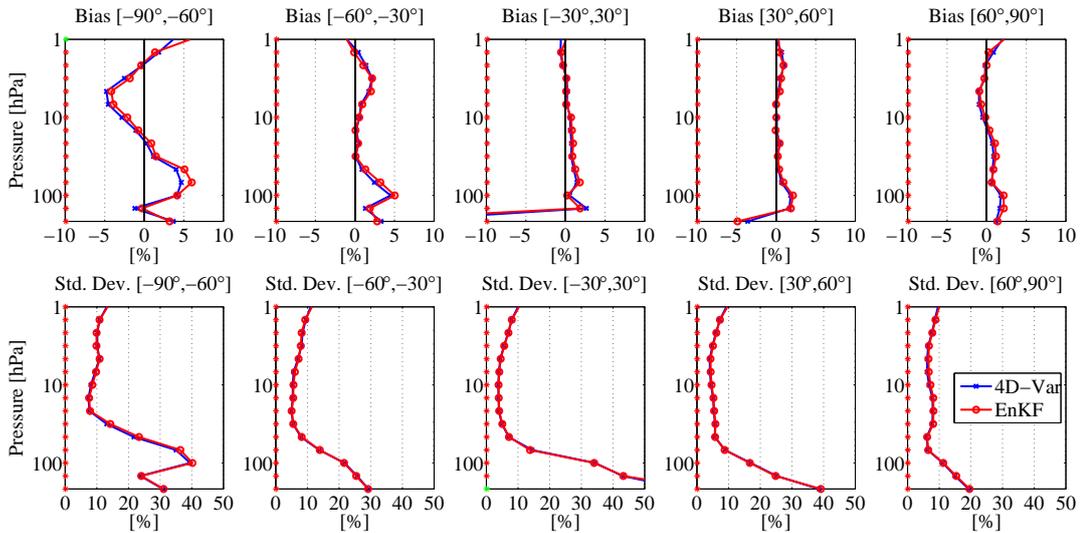


Fig. 4: OmF statistics for the EnKF (red lines) and the 4D-Var (blue lines) with respect to the assimilated EOS Aura-MLS data for the period September-October 2008. Bias (top row) and standard deviation (bottom row) for 5 different latitudinal bands. The green or red stars show the result of the Student- and Fisher-tests of significance on the 95% level (see text).

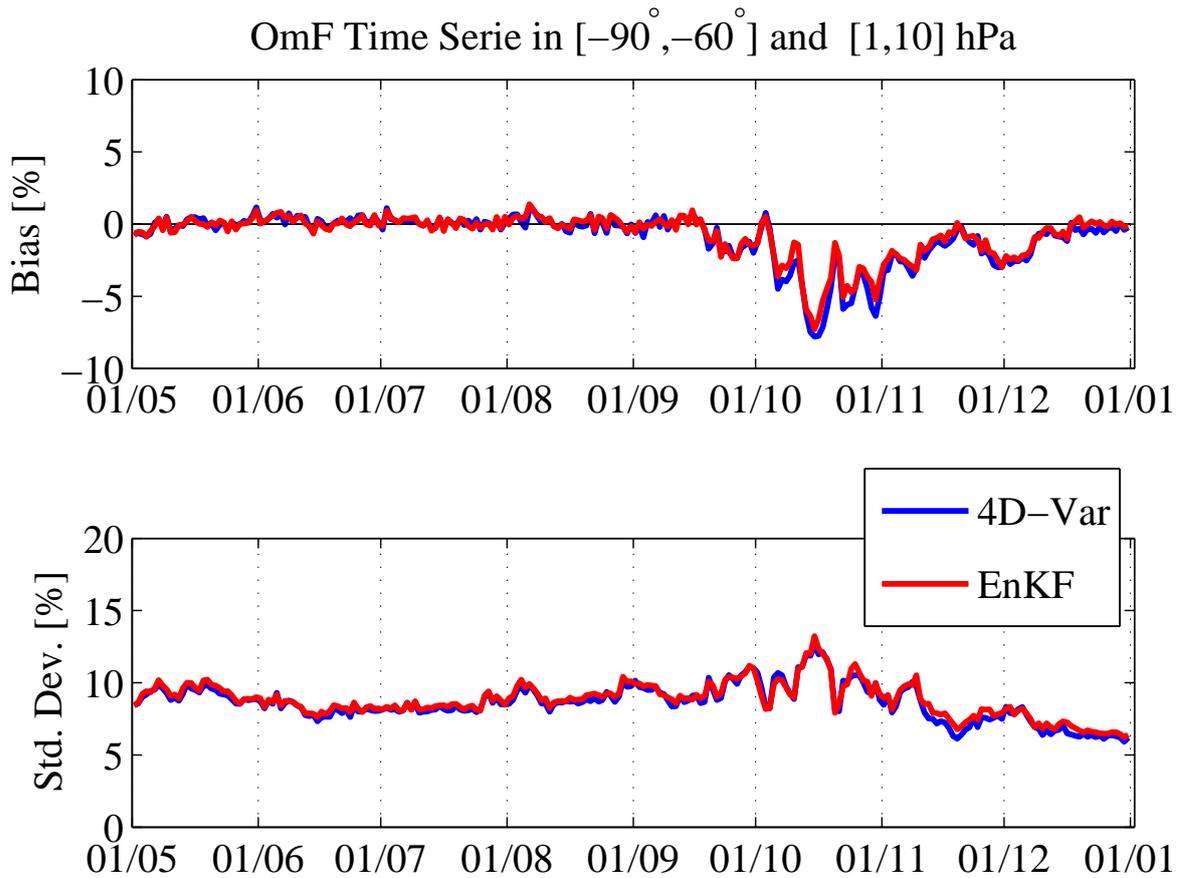


Fig. 5: OmF time series, bias (top row) and standard deviation (bottom row), for the EnKF (red lines) and the 4D-Var (blue lines) with respect to the assimilated EOS Aura-MLS data between 1 and 10 hPa and between -90° and -60° for the period May-December 2008.

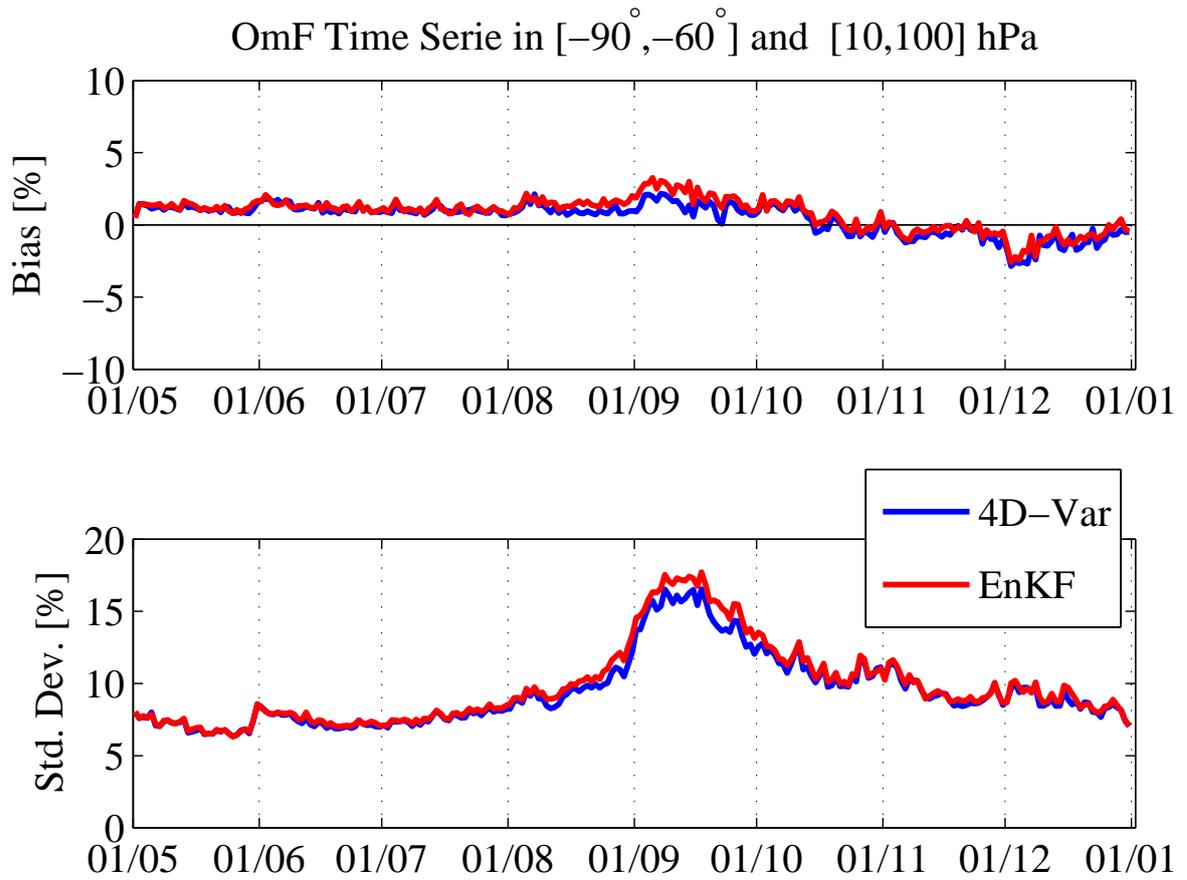


Fig. 6: As Fig. 5 but between 10 and 100 hPa.

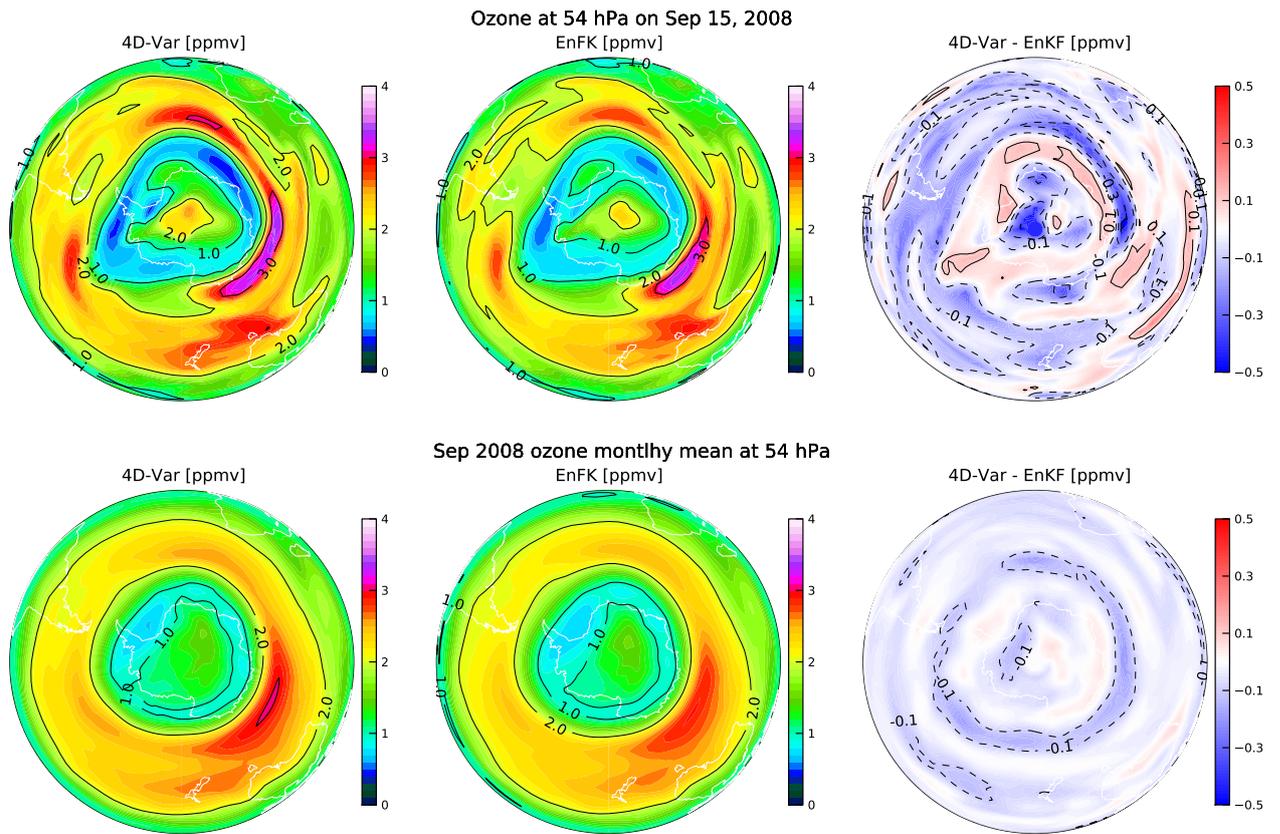


Fig. 7: O_3 at 54.6 hPa (model level 22) from 4D-Var (left), EnKF (middle) and their absolute differences (right). The upper row corresponds to a snapshot on September 15, 2008 while the lower row corresponds to a monthly mean for the month of September 2008.

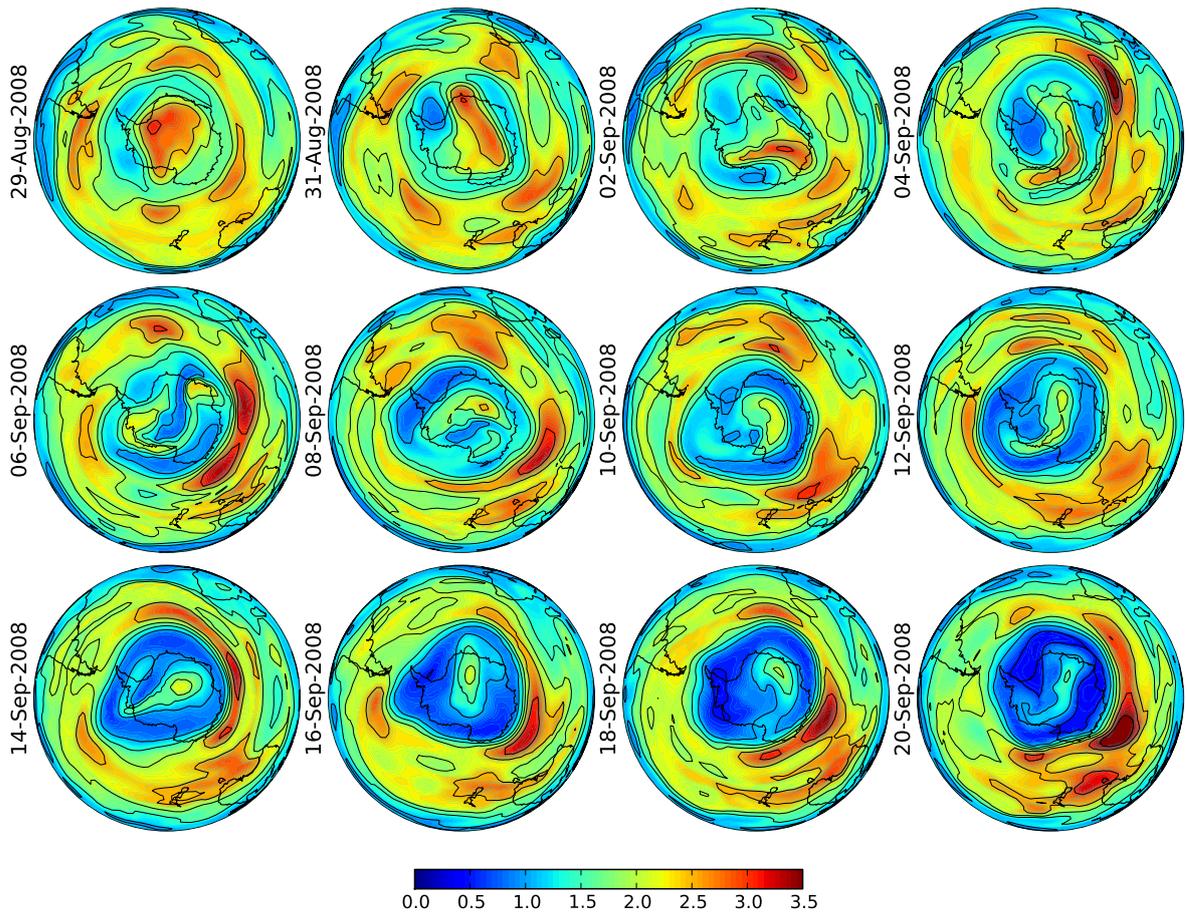


Fig. 8: Ozone distributions at 54.6 hPa (model level 22) for the Antarctic region. The snapshots are taken for the period between August 29 and September 20 every two days. Contours are plotted for every 0.25 ppmv.