

Potential of an ensemble Kalman smoother for stratospheric chemical-dynamical data assimilation

By THOMAS MILEWSKI* and MICHEL S. BOURQUI, *Department of Atmospheric and Oceanic Sciences, McGill University, Montréal, Canada*

(Manuscript received 14 April 2012; in final form 4 January 2013)

ABSTRACT

A new stratospheric ensemble Kalman smoother (EnKS) system is introduced, and the potential of assimilating posterior stratospheric observations to better constrain the whole model state at analysis time is investigated. A set of idealised perfect-model Observation System Simulation Experiments (OSSE) assimilating synthetic limb-sounding temperature or ozone retrievals are performed with a chemistry–climate model. The impact during the analysis step is characterised in terms of the root mean square error reduction between the forecast state and the analysis state. The performances of (1) a fixed-lag EnKS assimilating observations spread over 48 hours and (2) an ensemble Kalman Filter (EnKF) assimilating a denser network of observations are compared with a reference EnKF. The ozone assimilation with EnKS shows a significant additional reduction of analysis error of the order of 10% for dynamical and chemical variables in the extratropical upper troposphere lower stratosphere (UTLS) and Polar Vortex regions when compared to the reference EnKF. This reduction has similar magnitude to the one achieved by the denser-network EnKF assimilation. Similarly, the temperature assimilation with EnKS significantly decreases the error in the UTLS for the wind variables like the denser-network EnKF assimilation. However, the temperature assimilation with EnKS has little or no significant impact on the temperature and ozone analyses, whereas the denser-network EnKF shows improvement with respect to the reference EnKF. The different analysis impacts from the assimilation of current and posterior ozone observations indicate the capacity of time-lagged background-error covariances to represent temporal interactions up to 48 hours between variables during the ensemble data assimilation analysis step, and the possibility to use posterior observations whenever additional current observations are unavailable. The possible application of the EnKS for reanalyses is highlighted.

Keywords: ensemble data assimilation, Kalman smoother, stratospheric dynamics, stratospheric ozone

1. Introduction

Over the past decade, there has been an increasing recognition that stratospheric dynamics influence tropospheric medium- to long-range forecasts (Charlton et al., 2003; Gerber et al., 2012). Correspondingly, most operational centres have been raising the lid of their numerical weather prediction models. This extension of data assimilation to the stratosphere brings forth issues that are specific to the stratosphere, and its relative scarcity of observations. In particular, focus must be on model and observation biases, dynamical-chemical non-linearities and chemical or dynamical balances that require proper specification of background-error covariances (Polavarapu

et al., 2005). Effective assimilation strategies should thus be developed to address these outstanding issues.

The ensemble Kalman Filter (EnKF; Evensen, 1994) is now established as a promising data assimilation approach for the initialisation of operational weather forecasts (Szunyogh et al., 2008; Whitaker et al., 2008; Buehner et al., 2010). Its relatively simple implementation, and its reasonable performance even with non-Gaussian background states, (Evensen, 2003) makes the EnKF an excellent candidate for constraining the evolution of coupled non-linear systems, such as the stratosphere, with observations. It also provides an interesting framework for studying these complicated systems. An important issue of non-linearity in the stratosphere arises from the chemical-dynamical coupling. Chemical species are advected by the flow and ozone also feeds back on the radiation thus contributing to determining the flow. Hence, to adequately represent the stratosphere, interactions between the dynamics and

*Corresponding author.
email: thomas.milewski@mail.mcgill.ca

chemistry should be included in the modelling process, such as in chemistry–climate models (CCM, Eyring et al., 2006). The EnKF, particularly the perturbed observations type (Burgers et al., 1998), deals acceptably with a certain level of non-linearity (Lawson and Hansen, 2004) despite being optimal for linear dynamics. It is thus likely to be appropriate for application to a CCM to perform chemical-dynamical data assimilation.

The EnKF uses a Monte Carlo approach to estimate flow-dependent error covariances between variables from an ensemble of initially perturbed forecasts fields. These (forecast–, prior– or background–) error covariances have two main purposes. First, they quantify uncertainties in the model forecasts and determine the weights applied to the model with respect to the observations. Second, they provide estimated correlations between variables of the model state for the propagation of the weighted information from the observed variable to the correlated variables, including in particular the unobserved ones. The error covariances are particularly important in the stratosphere considering the sparseness of observations, both spatiotemporally and in terms of the variables covered. As a consequence, the analysis significantly relies on the background-error covariances to appropriately ‘fill the gaps’ and produce an accurate overall estimate of the model state.

Ensemble data assimilation may also be used to exploit temporal structures in observations for longer assimilation time windows, especially using observations posterior to the analysis time. Using posterior observations thus has the potential to impose an additional constraint on the model state by performing a *retrospective* analysis. This procedure is often referred to as the ‘smoothing problem’ (see Cosme et al., 2012, for a complete overview of the different algorithms). The smoothing problem may offer an additional potential to improve the stratospheric analysis, for two main reasons.

First, assimilation of observations posterior to analysis time allows the inclusion of additional data to constrain the analysis. In a linear dynamics regime context with a perfect model, the time lag between analysis and observations is irrelevant to the optimality of the analysis (as explained in Sakov et al., 2010), and, therefore, the assimilation of posterior data has the same potential constraint on the analysis as the assimilation of current data. Of course, the stratospheric dynamics are forced by non-linear phenomena such as momentum deposition from wave breaking. It is therefore important to investigate the differential impact of assimilating posterior versus current data. Model errors are also a significant source of analysis error, but their treatment is not the focus of this study.

Second, assimilation of stratospheric chemical constituent observations spread over time may help retrieve the

unobserved winds at analysis time. Indeed, temporal evolution of chemical constituents holds information on the winds through advection as long as it dominates over dissipation and chemistry within the assimilation time window, and the horizontal chemical tracer fields exhibit a gradient component parallel to the wind field (Daley, 1995; Riishøjgaard, 1996). This is likely to be the case in the lower stratosphere where ozone photochemical decay time-scales are long (McLinden et al., 2000) and within the Surf Zone where chaotic advection is dominant (Ngan and Shepherd, 1999). Furthermore, Milewski and Bourqui (2011), hereafter MB2011, have shown the ability of ensemble data assimilation of current ozone data to constrain unobserved winds in an idealised setting through the background-error covariances.

In ensemble data assimilation applied to the smoothing problem, error covariances are calculated between analysis and observation times using ensemble statistics. This allows the background-error covariances to include temporal information. This temporal information might be better estimated than in four-dimensional variational assimilation (4D-VAR) where the adjoint to the tangent linear model is used to propagate posterior information backwards in time. Ensemble data assimilation smoothers use the full non-linear model to estimate error covariances at various times, which transfer information from the posterior observations to the analysis. The smoothing problem in ensemble data assimilation was first approached with the ensemble smoother (EnS; van Leeuwen and Evensen, 1996), similar to the Asynchronous EnKF (AEnKF; Sakov et al., 2010), the four-dimensional local ensemble transform Kalman filter (4D-LETKF; Hunt et al., 2007) or the operational EnKF configuration at the Meteorological Service of Canada (Houtekamer et al., 2009). In these schemes, background-error covariances at all observation times within the assimilation window are obtained from a same ensemble of forecasts. These smoothers mostly allow for the assimilation of asynchronous observations at the time of the Filter analysis, useful for Numerical Weather Prediction (NWP) purposes. The ensemble Kalman smoother (EnKS; Evensen and van Leeuwen, 2000), on the contrary, assimilates posterior observations to perform a smoother pass on top of a previous Filter pass to further reduce analysis errors. Specifically, background-error covariances at a given observation time are calculated from an ensemble of forecasts initialised from an analysis obtained from the previous EnKF assimilation. The EnKS can perform retrospective analyses, for example, for reanalyses at little extra cost, if ensemble model forecast and observation states are stored from the previous EnKF pass (Evensen, 2003). Efficient smoother algorithms (see Ravela and McLaughlin, 2007) have also been developed to perform fixed-lag smoothing (Cohn et al., 1994), where

only the previous K analysis states are being retrospectively updated with the actual set of observations. In more realistic settings, the EnKS was applied to an atmospheric general circulation model by Khare et al. (2008) and proper spatial localisation and maximum smoother time-lag were estimated.

The objective of this study is to explore the ensemble assimilation of posterior stratospheric data with a CCM with the goal of obtaining a better analysis at the beginning of the assimilation time window. This is relevant especially for improving reanalyses or forecasts with lead time greater than the observation–analysis time lag (see for example Zhu et al., 2003). In both cases, posterior observations are readily available for assimilation. This study follows that of MB2011 and uses similar modelling and data assimilation system. In MB2011, ensemble data assimilation allowed a direct specification of flow-dependent multivariate background-error covariances, including chemical-dynamical covariances, which enabled the transfer of information from the observed variable to all state variables through repeated assimilation steps. Here, we focus on the specific role of the analysis step and its ability to reduce forecast errors. These corrections obtained from the assimilation step are represented using the analysis minus forecast (AmF) differences applied in calculating zonally-averaged root mean square errors (RMSE) for the variable of interest. This is first applied to an EnKF assimilation to provide a reference framework from which it will be easier to understand the added potential of the smoother. Comparisons with an EnKF assimilating a denser network of observations (to assimilate the same total amount of observations as the EnKS) are also reported to estimate the time-lag effect only.

A description of the EnKS is given in Section 2. The model, assimilation system, and experiments are described in Section 3. The assimilation impact with the resulting error structures are shown in Section 4. Conclusions are drawn in Section 5.

2. Ensemble Kalman smoother

In this study, the smoothing problem is approached with the EnKS as it yields better results than the EnS in non-linear contexts (Evensen and van Leeuwen, 2000), including for this experimental setup (not shown). The EnKS data assimilation technique solves an analysis equation very similar to that of the EnKF but with time lags between analysis and innovations (observations minus forecast). It also uses forecasts initialised from a previous EnKF data assimilation experiment as background fields.

The expression for the matrix $\mathbf{X}_{\text{EnKS}}^a \in \mathbb{R}^{n \times m}$ holding the ensemble of m analysis states vectors $\mathbf{X}_{\text{EnKS}}^a$ of size n at analysis time t_k is:

$$\mathbf{X}_{\text{EnKS}}^a(t_k) = \mathbf{X}_{\text{EnKF}}^a(t_k) + \sum_{k'=k+1}^K \mathbf{P}^{\text{af}}(t_k, t_{k'}) \mathbf{H}_{k'}^T [\mathbf{H}_{k'} \mathbf{P}^f(t_{k'}) \mathbf{H}_{k'}^T - \mathbf{R}(t_{k'})]^{-1} \mathbf{D}_{k'} \quad (1)$$

with

$$\mathbf{D}_{k'} = \mathbf{Y}_{k'} - \mathbf{H}_{k'} \mathbf{X}_{\text{EnKF}}^f(t_{k'}) \quad (2)$$

$$\mathbf{P}^{\text{af}}(t_k, t_{k'}) \mathbf{H}_{k'}^T = \frac{1}{m-1} \mathbf{X}'^a(t_k) [\mathbf{H}_{k'} \mathbf{X}_{\text{EnKF}}^f(t_{k'})]^T \quad (3)$$

$$\mathbf{H}_{k'} \mathbf{P}^f(t_{k'}) \mathbf{H}_{k'}^T = \frac{1}{m-1} [\mathbf{H}_{k'} \mathbf{X}_{\text{EnKF}}^f(t_{k'})] [\mathbf{H}_{k'} \mathbf{X}_{\text{EnKF}}^f(t_{k'})]^T \quad (4)$$

The EnKS analysis depends on the EnKF analysis $\mathbf{X}_{\text{EnKS}}^a$ at time t_k and the innovations $\mathbf{D}$ at observation time $t_{k'}$, where $k' = k+1, k+2, \dots, K$. The innovations are the difference between the ensemble of observations $\mathbf{Y}$ and the ensemble of forecasts from the EnKF data assimilation step $\mathbf{X}_{\text{EnKS}}^f$ mapped to observation space with the linear measurement operator $\mathbf{H}$. Here, we adapted the unified notation of Ide et al. (1997) to encompass the practical formulation of Evensen (2003). For example, the prime operator $(.)'$ refers to the departure from ensemble average and not to a linearisation.

The initial EnKF analysis at time t_k is repeatedly corrected with observations at each posterior time $t_{k'}$ up to t_K , to eventually obtain the final EnKS analysis at time t_k . The innovations are weighted based on the forecast and observation uncertainties at observation time $t_{k'}$, given by the forecast-error and observation-error covariance matrices $\mathbf{H}_{k'} \mathbf{P}^f(t_{k'}) \mathbf{H}_{k'}^T$ and $\mathbf{R}_{k'}$. In addition to weighting the forecast errors at time $t_{k'}$, the time-lagged background-error covariance matrix $\mathbf{P}^{\text{af}}(t_k, t_{k'}) \mathbf{H}_{k'}^T$ also transfers the information from the weighted innovations from posterior time to analysis time. The observation error covariances are prescribed from knowledge on instrument error as well as representativeness error linked to the observation operator $\mathbf{H}$. In addition to the forecast error weighting, the background-error covariance matrix $\mathbf{P}^{\text{af}}(t_k, t_{k'}) \mathbf{H}_{k'}^T$ also transfers the time-lagged innovation information from observation time $t_{k'}$ to analysis time t_k . Note that eq. (1) is iterative since the error covariance matrix $\mathbf{P}^{\text{af}}(t_k, t_{k'})$, defined in eq. (3), is computed using the latest analysis $\mathbf{X}'^a(t_k)$, which is updated every time a new posterior observation at time k' is introduced in the calculation of eq. (1).

We introduce the notation EnKS $[\Delta t_1, \Delta t_2, \dots, \Delta t_K]$ to include the time lag information $\Delta t_{k'-k} = t_{k'} - t_k$ (in hours) between observation time $t_{k'}$ and analysis time t_k , for better clarity when describing the different experiments in this study. Experiments use fixed-lag K EnKS with

maximum time lag Δt_k , with values of K ranging between 0 and 2, and maximum time lags between 0 and 48 hours.

The sequential treatment of observations by batches (Houtekamer and Mitchell, 2001) used in MB2011 in the context of the EnKF could not be used in this study with the EnKS because of instabilities that arose from the sequential assimilation of observations in batches at a given posterior observation time. The RMSE and SPREAD (root mean square ensemble deviation about the ensemble mean as described in MB2011) of the ensemble dramatically increase in the assimilated variables after each batch in some localised spots in the North and South polar upper troposphere lower stratosphere (UTLS) regions. The instabilities grow batch after batch and propagate to other variables and regions. Further investigation is needed to understand the causes of this instability. However, it disappears when assimilating all observations of a given posterior time at once using the compressed row storage technique (CRS, sometimes called ‘Compressed Sparse Row’ CSR; Dongarra, 2000) to retain computational efficiency (details in Appendix A).

3. Experimental setup

3.1. Reference EnKF experiment

This study, as in MB2011, is cast in a perfect-model Observation System Simulation Experiment (OSSE) context using a perturbed-observation Double-EnKF (Evensen, 1994; Burgers et al., 1998; Houtekamer and Mitchell, 2001) coupled with the chemistry–climate model IGCM-FASTOC (Bourqui et al., 2005; Taylor and Bourqui, 2005) run at T21 resolution with 26 vertical sigma levels reaching up to 0.1 hPa.

A realisation of the IGCM-FASTOC is taken to represent the true state of the atmosphere. Observations are simulated from the true state at 00:00 UTC every 24 hours by trimming the state vector to the appropriate observation vector using the measurement operator $\mathbf{H}$. The simulated observations represent the amount of retrieved atmospheric profiles from the Envisat Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) instrument over a 24-hour period. Observations are not assimilated in batches of up to 100 observations as in MB2011, but all at once. Each profile has data on 11 levels, ranging from 180 to 4 hPa (roughly 12 to 38 km altitudes), and the variables assimilated are either temperature (T) or odd oxygen mixing ratio (O_x , the chemical family composed of an ozone molecule O_3 and a single oxygen atom O). Note that O_x and O_3 will be used interchangeably in the following, as is usual in the stratosphere (Dessler, 2000). The observation vector is normally randomised, with typical MIPAS observation-error standard deviations equal to 10% for ozone and 2 K

for temperature. The ensemble of observations generated are assimilated at 00:00 UTC every 24 hours to the ensemble of forecasts to obtain the ensemble of analyses.

The advantage of OSSEs is that, unlike assimilation experiments with real observations, the truth is known. It is therefore possible to evaluate if the analysis converges towards the right solution. In ensemble data assimilation systems, this is typically done by calculating the RMSE of the ensemble mean with respect to the true state. The RMSE diagnostic expresses the accuracy of the EnKF solution and should be compared to the root mean square difference between the ensemble members and the ensemble mean (SPREAD) to estimate the reliability of the EnKF solution (Sacher and Bartello, 2009). The SPREAD is essentially the standard deviation of the ensemble, and equality between the SPREAD and the RMSE must be achieved for the EnKF solution to be consistent. If successful behaviour is obtained from the assimilation in a perfect-model OSSE, the next step is to perform imperfect-model (non-identical twin) experiments, where the true state is again known, but is now a simulation from a different version of the model (fraternal-twin OSSE) or preferably a different model. This introduces model errors in the assimilation system that need to be accounted for. The ultimate step is the assimilation of real observations. Note, however, that the objective of this study is not to evaluate the assimilation impact of a forthcoming observation source as is often the goal in OSSEs. The infrastructure in such OSSEs is more complex as conventional meteorological observations are also assimilated within already well calibrated data assimilation systems (see for example the OSSE evaluating the potential of assimilating Doppler Wind Lidar data in the NCEP system of Masutani et al., 2010). The intent of the OSSE in this study is rather to study the properties of an EnKS for stratospheric data assimilation.

The assimilation setup in this study retains the basic properties of the $m = 128$ member Double-EnKF assimilation of MIPAS-like profile retrievals of MB2011. The background state vector for assimilation is composed of the control variables: zonal wind U , meridional wind V , temperature T , surface pressure P_s , specific humidity q , odd oxygen mixing ratio O_x , odd nitrogen mixing ratio NO_x , dinitrogen pentoxide mixing ratio N_2O_5 and nitric acid mixing ratio HNO_3 . For reference, the NO_x mixing ratio is the sum of the individual NO , NO_2 , NO_3 and HONO mixing ratios. A small change in the localisation parameters for the temperature assimilation has been applied. Localisation length parameters for all temperature–chemical tracer error covariances have been reduced to $[C_h, C_v] = [5600 \text{ km}, 4\Delta \ln(P)]$. In MB2011, only the T– O_x error covariances were localised with this set of shorter-length parameters, whereas temperature error covariances with other chemical tracers (NO_x , N_2O_5 , and HNO_3) were treated with

the longer-length parameters [14000 km, $10\Delta\ln(P)$]. This change slightly improves results on the odd-nitrogen family $\text{NO}_y = \text{NO}_x + \text{HNO}_3 + 2 \times \text{N}_2\text{O}_5$ (not shown).

The reference EnKF experiment is referred to as ‘EnKF MIPAS’. It is an assimilation of current MIPAS temperature or ozone observations. Note that this is equivalent to doing an EnKS[0] update. Figure 1 shows the RMSE and SPREAD for the EnKF MIPAS Ox and T assimilation experiments (red and blue lines) and for the free-running ensemble (no assimilation performed, black lines). All statistics and other experiments are started from day 31 onwards, as both simulations present a stationary evolution with the desired saw-teeth pattern of forecast error growth and analysis error decay. Note that the SPREAD (dashed lines) is higher than the RMSE (solid lines) in both experiments without the use of any artificial covariance inflation. This is likely due to the limited amount of observations assimilated, as reported by Gottwald et al. (2011) and witnessed in Whitaker et al. (2009). There is a slight difference in magnitude between both experiments in the RMSE (solid lines) and SPREAD (dashed lines)

quantities, with the EnKF MIPAS T displaying the lowest values. This likely owes to the fact that the error level in temperature observations provides a stronger impact on the dynamical analysis than ozone observations do.

The spatial distribution of the time-averaged, zonally-averaged forecast RMSE fields for zonal wind, temperature and ozone concentrations are showed in Figure 2 for the EnKF MIPAS T and EnKF MIPAS Ox experiments. For each of the two experiments (left and right columns of Fig. 2), the RMSE fields are averaged over the 28 d of February. The forecast RMSE structures are quite similar in both experiments, except for the notable dynamical (temperature and zonal wind) errors in the Northern Hemisphere extratropical upper stratosphere only present in the EnKF MIPAS Ox experiment. With this exception, the differences are only in the magnitudes of the forecast RMSE. Smaller errors are witnessed in the ozone maximum for the EnKF MIPAS Ox experiment and in the tropical upper stratosphere temperatures for the EnKF MIPAS T experiment. Overall, the differences in the forecast RMSE from both experiments are generally small.

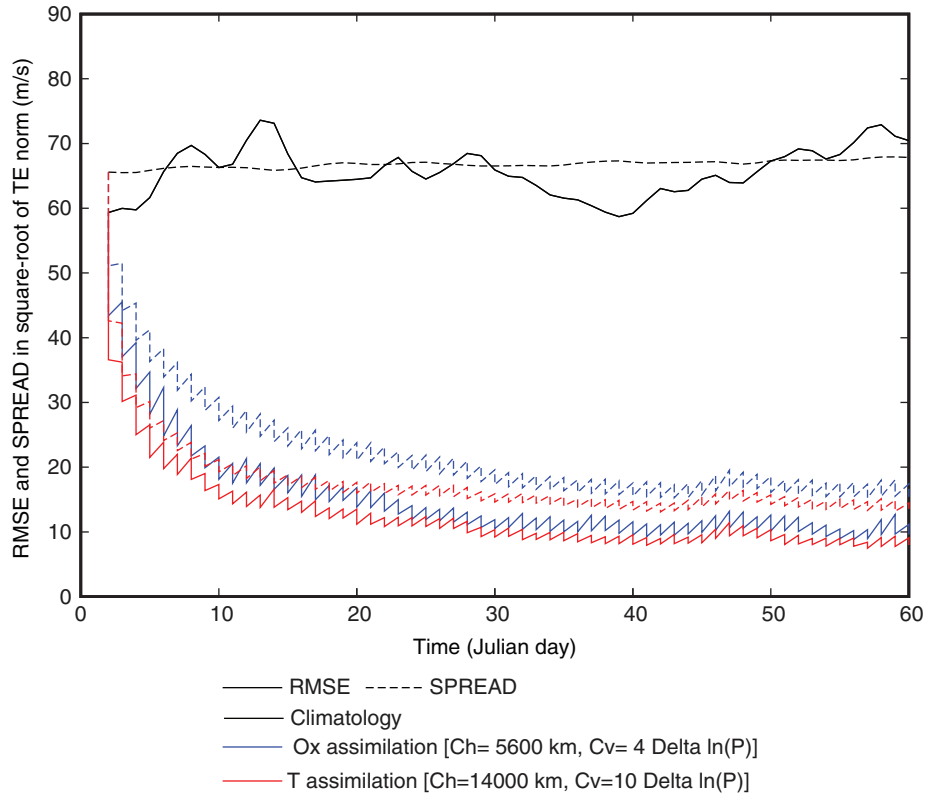


Fig. 1. Time evolution of the root mean square error (RMSE, solid line) and ensemble spread (SPREAD, dashed line) of the square root of the total energy norm $\text{TE} = \text{U}^2 + \text{V}^2 + \frac{R_s T_{\text{ref}}}{P_{\text{ref}}} \text{P}_s^2 + \frac{C_p}{T_{\text{ref}}} \text{T}^2$ (in m.s^{-1}) with $P_{\text{ref}} = 1000$ hPa, $T_{\text{ref}} = 300$ K, C_p and R_s being the specific heat at constant pressure and the gas constant of dry air. Black lines: free-running (no assimilation performed) model ensemble forecasts; blue lines: ozone assimilation cycle; red lines: temperature assimilation cycle. Horizontal and vertical localisation length parameters (C_h and C_v) are given in the legend.

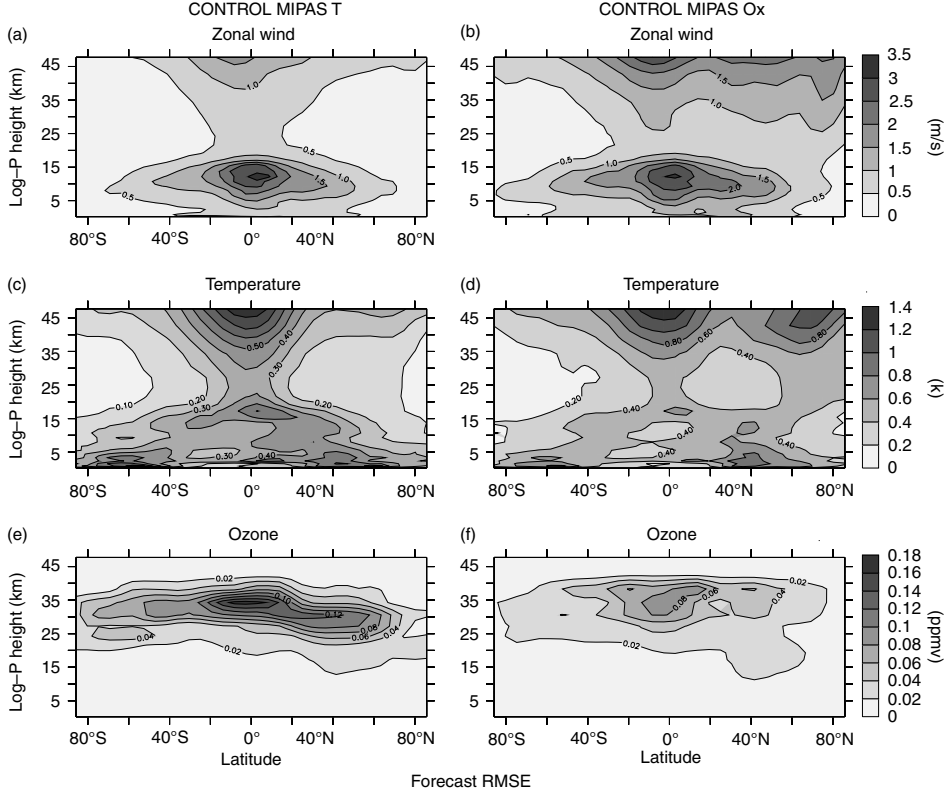


Fig. 2. 28-d time-averaged zonal-mean forecast root mean square error (RMSE) for zonal wind (top row, m.s^{-1}), temperature (middle row, K) and ozone (bottom row, ppmv). Left and right columns show forecast RMSE from the EnKF MIPAS temperature and ozone assimilation experiments, respectively.

3.2. EnKS experiments

A first set of experiments is designed to determine the accuracy of the analyses obtained when observations are assimilated at a single time posterior to the analysis time. Two ozone (temperature) EnKS experiments are performed using the EnKF MIPAS Ox (EnKF MIPAS T) forecasts over the whole month of February (Julian days 31 to 58) as basis for the EnKS assimilation [i.e. $\mathbf{X}_{\text{EnKF}}^f$ in eq. (2–4)]. These ozone (temperature) assimilation experiments perform an EnKS analysis [eq. (1) of Section 2] with a single fixed observation time lag ($K=1$) but with different time intervals: $\Delta t_1=24$ hours and $\Delta t_1=48$ hours between observation and analysis, respectively. They are denoted EnKS[24] and EnKS[48] hereafter, based on the notation introduced in Section 2. These experiments allow to determine the contribution to the analysis error reduction of posterior temperature and ozone observations as a function of their time lag with respect to the analysis.

A second set of experiments is designed to study the additive impact of using observations at several times, this for both temperature and ozone assimilation. The corresponding experiments are described below.

EnKS MIPAS: MIPAS temperature or ozone observations are assimilated for two posterior days. This is equivalent to a fixed-lag $K=2$ smoothing procedure, for a maximum time lag of $\Delta t_K=48$ hours. The update procedure for the EnKS MIPAS is an EnKS[24, 48]. Note that, during each analysis step, the EnKS MIPAS experiment assimilates three times the amount of observations that the EnKF MIPAS experiment does (a single daily MIPAS network during the Filter pass plus two daily MIPAS networks during the smoother pass for the EnKS versus a single daily MIPAS network during the Filter pass for the EnKF). The difference in the volume of data assimilated may lead to analysis improvements due to larger amounts of data assimilated, making ambiguous the benefits gained from the assimilation of posterior observations. The following experiment is designed to partially alleviate this ambiguity.

EnKF 3 $\times$ MIPAS: Assimilation of current observations corresponding to a three times denser MIPAS network. The EnKF 3 $\times$ MIPAS is performed by repeating the EnKF pass three times, with horizontally-offset MIPAS observations each time. In this experiment, observations are filling the entire horizontal model grid for the 11 stratospheric levels of the IGCM-FASTOC.

4. Assimilation impact on analyses

In this section, the impact on the analysis of assimilating either temperature or ozone, at current or posterior time, is investigated. The impact is quantified in terms of time-averaged zonal mean RMSE change between the forecast and the analysis (hereafter, Δ RMSE, expressed in percentage relative to the forecast RMSE) for the variable of interest. This diagnostic was chosen because it isolates to some degree the effect of the analysis step from the subsequent balancing by the model forecasting step. It therefore allows to estimate typical error corrections accomplished by the EnKF or EnKS assimilations alone.

4.1. Assimilation impact of current observations

Before focusing on the assimilation impact of posterior observations, it is instructive to view which regions are affected in a typical EnKF assimilation of current temperature or ozone observations. The error correction structures from the EnKF MIPAS are shown in Fig. 3 for the zonal wind (top row), temperature (middle row) and ozone (bottom row) variables.

The EnKF MIPAS assimilation of temperature observations (left column) and ozone observations (right column) yields Δ RMSE values that are globally negative (blue to purple colours) implying beneficial assimilation impacts everywhere, except for a few small regions in the upper stratosphere with neutral or slightly detrimental assimilation impact (pink colour). The EnKF MIPAS assimilation produces different dynamical Δ RMSE with assimilation of temperature (left column) than with assimilation of ozone (right column). Temperature assimilation beneficially impacts the Polar Vortex winds and temperatures (Figs. 3a and 3c) while ozone assimilation achieves only limited improvements there (Figs. 3b and 3d). The reason for such a difference is likely due to the isolated nature of the Polar Vortex, in addition to the fact that chemical-dynamical covariances are generally weaker than purely dynamical covariances. The vortex nature reduces the error covariance between its interior and exterior and thereby the impact of observations outside the vortex on the interior dynamical state. Note that the smaller impact of ozone observations on the dynamical variables explains in part the larger forecast errors in the Northern Hemisphere in the EnKF MIPAS Ox experiment (see Fig. 2). The most significant

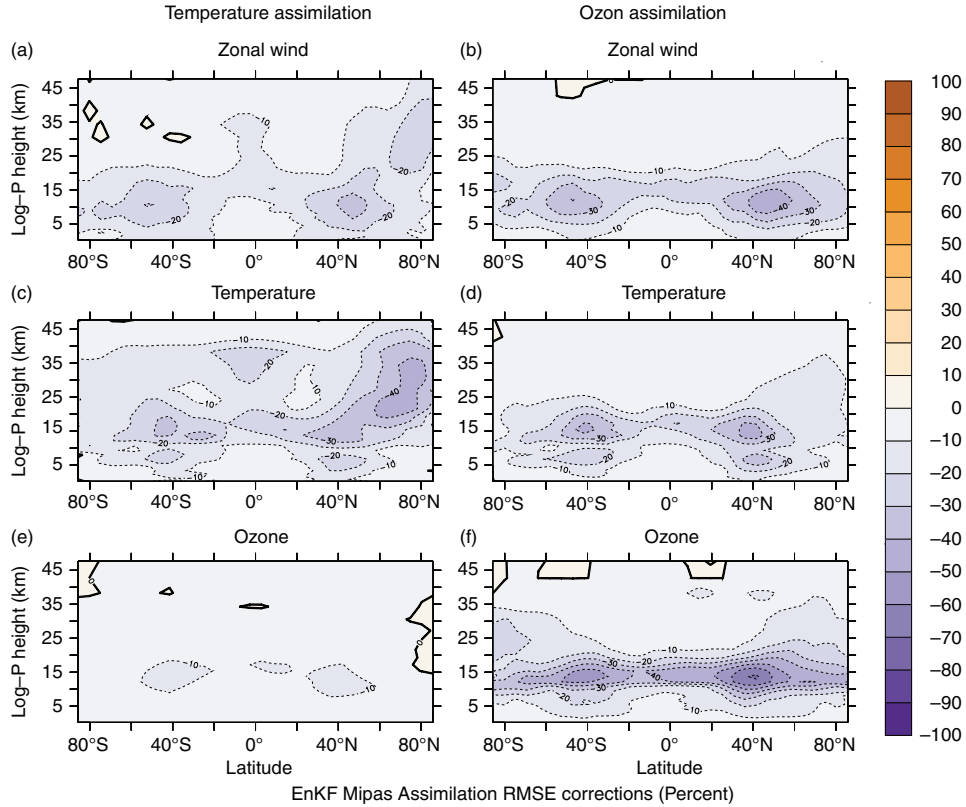


Fig. 3. 28-d average analysis minus forecast RMSE changes (in %) for EnKF MIPAS temperature assimilation (left column) and EnKF MIPAS ozone assimilation (right column). Top row: zonal mean zonal wind; middle row: zonal mean temperature; bottom row: zonal mean ozone. Zero line of RMSE changes contoured with solid black line. Purple (orange) colours indicate a reduction (increase) in RMSE.

Δ RMSE achieved with EnKF MIPAS temperature assimilation are in Northern Hemisphere Polar Vortex and tropospheric jet regions, with 40% error decrease per assimilation step in vortex temperature and 30% error decrease in zonal jet speeds. Strongest error corrections for ozone assimilation occur in the UTLS region, with maximal RMSE reductions of up to 40% per analysis step in the tropospheric jets (Fig. 3b). The Δ RMSE are also different in the equatorial region where assimilation of temperature observations can better constrain the mid-stratosphere thermal and dynamical fields. In the UTLS, assimilation of ozone and temperature observations constrain the winds and temperature (Figs. 3a–d) quite similarly. The improvements in the equatorial UTLS winds are an important achievement, considering the difficulties for models to correctly represent the dynamics in this region (Pawson et al., 2000; Eyring et al., 2006). The EnKF statistically captures some aspects of equatorial dynamical balances in its error covariances, which might otherwise be difficult to prescribe (Zagar et al., 2004).

For each temperature and ozone assimilation (left and right columns of Fig. 3, respectively), there is a strong similarity in the temperature and zonal wind Δ RMSE (Fig. 3a versus Fig. 3c and Fig. 3b versus Fig. 3d), but the impact on the ozone field is different. In fact, temperature assimilation corrects the global ozone errors only slightly. The only distinct signal is seen near the tropospheric jets. The ability of temperature assimilation to correct errors in the middle to upper stratosphere, as for the dynamical variables, is not witnessed in the ozone variable. Conversely, the ozone assimilation can only reduce the dynamical variable errors in the UTLS region. Its impact on the ozone state is not only strong near the tropopause but also extends in the mid-stratosphere, particularly in both polar regions.

Apart from the UTLS and tropospheric jets that are constrained in all variables by chemical and dynamical observations, there is a general complementary impact from assimilation of temperature and ozone in the mid-stratosphere. Temperature assimilation can constrain dynamical variables in that region, something that ozone assimilation cannot achieve. In turn, ozone assimilation can constrain ozone, this being something that temperature assimilation cannot achieve.

4.2. Assimilation impact of posterior observations

Observations are rarely exactly measured at the time of analysis. In fact, ‘asynoptic’ observations like satellite or radar data are spread over time. A purely sequential assimilation of such data, where a new assimilation step is started at every discrete time when observations are available would be extremely impractical. In the case of linear dynamics with a perfect model, the analysis will not

be different if the observations are assimilated at or posterior to the analysis time (Sakov et al., 2010). But in realistic atmospheric simulations, the growing non-linearities associated with such time lags (and the model errors) are expected to reduce the EnKS data assimilation performance compared to the EnKF. The assimilation impact as a function of time lag is examined here.

The different assimilation impacts of the EnKF and the two fixed-lag ($K=1$) EnKS experiments EnKS[24] and EnKS[48] are investigated in terms of Δ RMSE, averaged here over the 28 daily analyses performed during a typical month of February. The time-averaged zonal mean zonal wind Δ RMSE is presented in Fig. 4 for temperature assimilation (left column) and ozone assimilation (right column) experiments. Note that the zonal mean zonal wind Δ RMSE for the EnKF MIPAS temperature and ozone assimilation experiments in Fig. 3a and 3b and Fig. 4a and 4b are the same.

The middle and bottom rows of Fig. 4 display the assimilation impact from the EnKS[24] and EnKS[48], respectively. For a 24-hour time lag between analysis and observations (Fig. 4c and 4d), the temperature and ozone assimilation Δ RMSE structures remain relatively similar to the ones seen in the EnKF but the magnitude of the correction are globally reduced by a factor of about two. As the time lag is increased to 48 hours (Fig. 4e and 4f), assimilation corrections are further reduced to improvements generally lower than 10% per analysis step. The assimilation impact remains globally beneficial in the ozone assimilation experiment for a 2-d time lag, but becomes increasingly deleterious for the temperature assimilation case in the subtropical and tropical upper stratosphere. This decrease in Δ RMSE with time lag is also witnessed for the other state variables O_x and T (not shown).

The background-error covariances are therefore able to convey information backwards in time, but with a loss in signal with increasing time lag. The overall decrease in magnitude of error corrections with time lag does not allow an arbitrary choice of analysis time with respect to observations as expected for a linear model, following Sakov et al. (2010). Instead, a maximum time lag needs to be imposed. With model error-free stratospheric ensemble assimilation, 48 hours seems like the appropriate choice. This is somewhat equivalent to applying temporal localisation, following the logic of Cosme et al. (2012). In our non-linear context, the assimilation of current observations has a better impact on the analysis than the assimilation of posterior observations. Yet, the presence of a beneficial correction on the analysis based upon posterior observations suggests that assimilating posterior observations in addition to current ones has the potential of further improving the analysis.

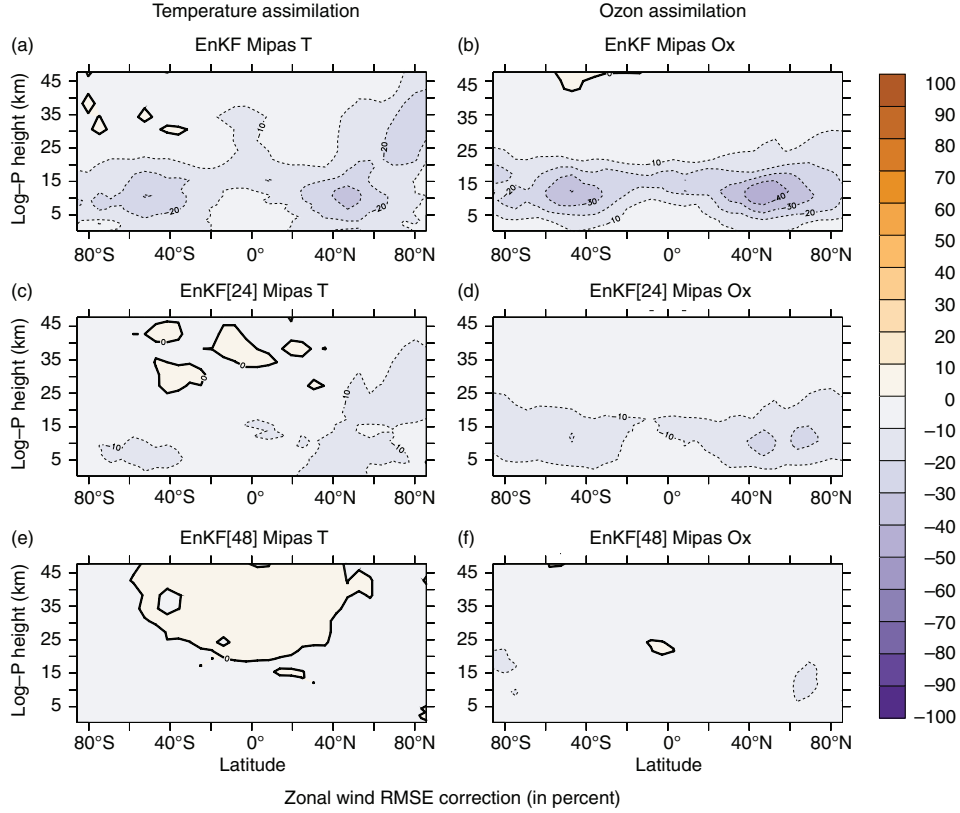


Fig. 4. 28 d average of zonal mean zonal wind analysis minus forecast RMSE changes (in %) for MIPAS ozone assimilation with (a) an EnKF, (b) an EnKS with 24-hour lag between observation and analysis time, (c) an EnKS with 48-hour lag. Purple (orange) colours indicate a reduction (increase) in RMSE.

4.3. EnKS assimilation impact

The EnKF MIPAS assimilation experiment provides a reference to determine the relative impact of assimilating additional temporal data with the EnKS MIPAS or spatial data with the EnKF $3 \times$ MIPAS. These impacts are calculated by taking the difference between the EnKS MIPAS analysis RMSE and the EnKF MIPAS analysis RMSE, and the difference between the EnKF $3 \times$ MIPAS analysis RMSE and the EnKF MIPAS analysis RMSE, respectively. These Δ RMSE values are also expressed in percentage relative to the EnKF MIPAS forecast RMSE, so that they can be compared to the Δ RMSE values of the EnKF assimilation case of Section 4.1.

The results for the EnKS MIPAS Ox and EnKF $3 \times$ MIPAS Ox ozone assimilation experiments are displayed in the left and right columns of Fig. 5, respectively. The difference in zonal mean zonal wind Δ RMSE is shown in the top row. Black lines enclose regions where differences are significant at the 95% confidence level, as determined from a bilateral (two-sided) student's t test. Looking first at the impact of assimilating additional ozone observations posterior to the analysis time with the EnKS MIPAS

(Fig. 5a), the largest regions displaying statistically significant differences are in the extratropical troposphere as well as the UTLS. These improvements brought on by the EnKS are happening in regions already constrained by the EnKF (top row of Fig. 3, beware of scale differences) and add an average 10% further analysis error reduction. The Polar Vortex is one particularly notable region, with maximum RMSE corrections above 20%, as it was only very slightly affected by the EnKF MIPAS ozone assimilation (Fig. 3b).

However, the additional wind correction in the Polar Vortex is not exclusive to the EnKS MIPAS as the EnKF $3 \times$ MIPAS ozone assimilation also obtains a similar significant correction (Fig. 5b). The EnKF $3 \times$ MIPAS experiment serves as a comparison for the EnKS MIPAS to see the difference between assimilating additional spatial data and temporal data. The additional error correction in Polar Vortex winds from both experiments (in conjunction with the small constraint from the EnKF MIPAS Ox) indicates that a good observational coverage is necessary for a strong constraint on the Polar Vortex wind analysis, regardless of how the observations are distributed over the observation time window. However, some regions display

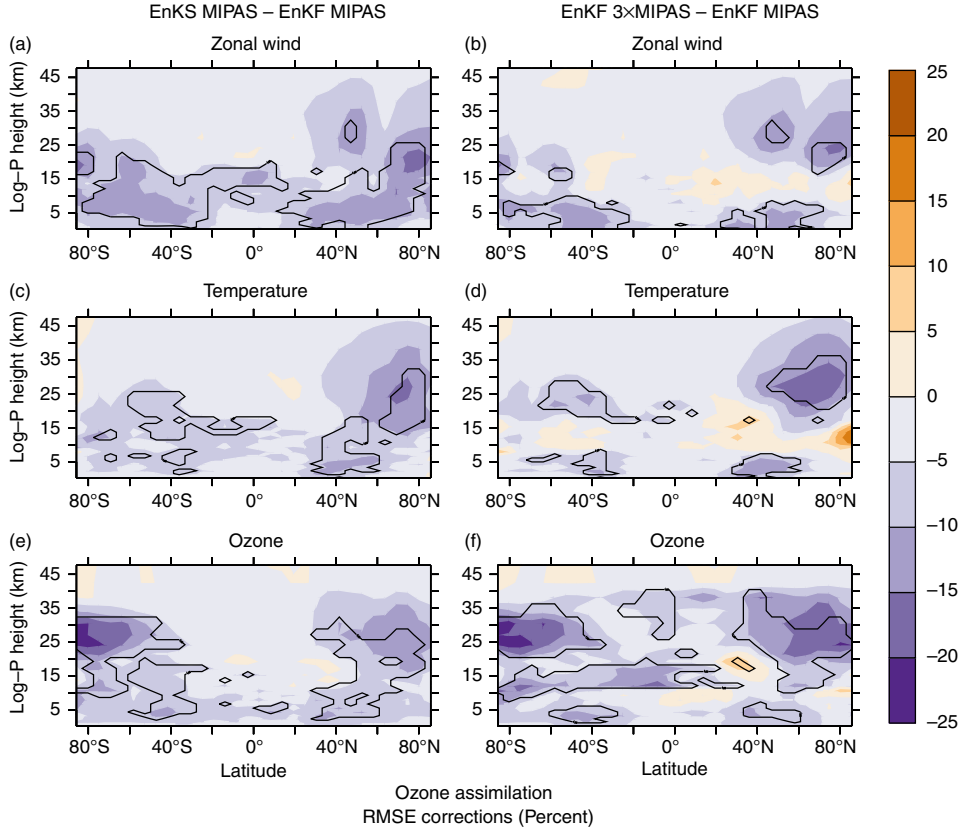


Fig. 5. Zonal mean zonal wind (upper row), temperature (middle row) and ozone (bottom row) Δ RMSE differences (in %) between the EnKS MIPAS and the EnKF MIPAS (left column) and between the EnKF 3 $\times$ MIPAS and the EnKF MIPAS (right column) for ozone assimilation. Purple (orange) colours indicate improvement (degradation) of the EnKS MIPAS or EnKF 3 $\times$ MIPAS over the EnKF MIPAS. Statistically significant differences at the 95% confidence level are enclosed by the black lines.

different correction patterns between the EnKS MIPAS Ox and the EnKF 3 $\times$ MIPAS Ox, namely the tropical and northern extratropical UTLS. Unlike what is expected from linear theory, the additional assimilation of current observations produces near zero corrections there, whereas the additional assimilation of posterior observations produces significant further constraints. It is therefore safe to say that the smoothing approach represents a signification additional improvement with respect to the EnKF stratospheric ozone assimilation to constrain the winds, in the context of a perfect-model experiment.

The situation is very similar for the impact of assimilating additional ozone observations on the temperature analysis (middle row of Fig. 5). The significant temperature analysis improvements with the EnKS MIPAS Ox are located in the same regions than for the winds, although they are less spatially extensive and of smaller magnitude (Fig. 5c versus Fig. 5a). The EnKS MIPAS offers a generally similar constraint on the UTLS than the EnKF 3 $\times$ MIPAS, except on the Polar Vortex temperatures where the denser-network Filter is very effective. Overall, this tells us that there is

valuable and available additional information in future ozone observations that the EnKS is able to use to further constrain the EnKF MIPAS dynamical (zonal wind and temperature) analyses. The total number of observations being equal, the smoother does so at least as well as the Filter.

For the impact on the ozone analyses, the improvement from adding current ozone observations (EnKF 3 $\times$ MIPAS Ox of Fig. 5f) has a pronounced beneficial effect in the middle stratosphere and equatorial UTLS. These regions are not well constrained by the reference EnKF MIPAS experiment (see Fig. 3f). The additional error corrections from the EnKS MIPAS are also significant and of similar magnitudes in the extratropical middle stratosphere compared to the EnKF 3 $\times$ MIPAS ones (Fig. 5e versus Fig. 5f). However, ozone improvements are not as notable from the EnKS MIPAS ozone assimilation in the equatorial stratosphere. This is surprising considering that the EnKS MIPAS Ox performed better on the dynamical variables in the equatorial regions than the EnKF 3 $\times$ MIPAS Ox.

The impact of assimilating additional temperature observations can be evaluated with Fig. 6. The EnKS MIPAS temperature assimilation (Fig. 6a) produces significant additional zonal wind error corrections in the extratropical troposphere and equatorial UTLS. However, there is a lack of additional correction on the Polar Vortex winds. This can be explained by the already strong correction that the EnKF MIPAS temperature assimilation produced there (see Fig. 3a), as opposed to the weak correction from the EnKF MIPAS ozone assimilation. Again, the EnKF $3 \times$ MIPAS corrects the same regions as the EnKS MIPAS but to a weaker extent. This shows again the potential of the EnKS for correcting further the wind analyses.

However, the temperature assimilation experiments EnKS MIPAS T and EnKF $3 \times$ MIPAS T exhibit very different error-correction patterns on the temperature and ozone analyses (middle and bottom rows of Fig. 6), contrary to the general similarity of EnKS and EnKF $3 \times$ MIPAS ozone assimilation impacts. The EnKF $3 \times$ MIPAS temperature assimilation case displays significant temperature improvements in the tropical upper and lower

stratosphere (Fig. 6d). This is likely from the finer horizontal resolution in the observation network that helps determining the smaller-scale dynamical features proper to the tropics. On the contrary, the EnKS MIPAS temperature assimilation (Fig. 6c) can only achieve marginal significant temperature improvements limited to small regions near the tropospheric jets, and almost nothing in the tropical stratosphere. As for the impact on the ozone analysis, the EnKF $3 \times$ MIPAS temperature assimilation is effective in further reducing the error in the UTLS belt from 60°S to 60°N (Fig. 6f). The effect of the EnKS MIPAS temperature experiment (Fig. 6e) is simply not statistically significant anywhere. The quick decay in the effectiveness of temperature and time-lagged error covariances, as witnessed in Fig. 4, is certainly the cause of the poor performance of the temperature smoother, compared to the ozone smoother.

5. Conclusions

The goal of this study was to investigate the potential benefit of a smoothing approach where posterior observations are assimilated to further improve the model state

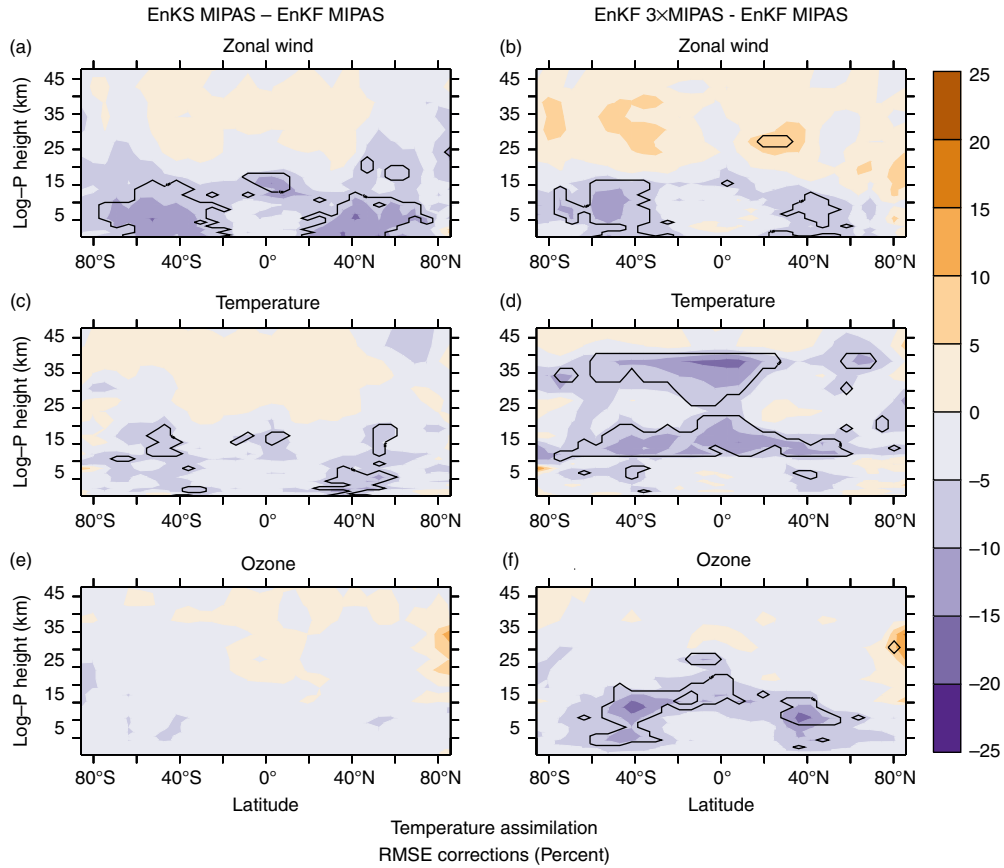


Fig. 6. Same as Fig. 5 but for temperature assimilation.

analysis. This was done using dynamical and chemical data assimilation in an idealised perfect-model observation system simulation experiment setup. An EnKS based on Evensen and van Leeuwen (2000) was developed and applied to the IGCM-FASTOC chemistry–climate model. The system is identical to the one in MB2011, except for a few changes in localisation parameters and the fact that all available observations from a typical daily MIPAS network are assimilated simultaneously, instead of in successive observation batches following Houtekamer and Mitchell (2001). A sparse-matrix treatment of the background-error covariances was implemented for computational efficiency.

The smoothing problem was investigated through the properties and skill of the EnKS update. To separate the effects of the analysis and the forecast in the data assimilation cycle, relative difference of RMSE (Δ RMSE) between analyses and forecasts, normalized by the forecast RMSE, were produced and examined.

Understanding the impact of assimilating posterior observations required a prior knowledge of the assimilation impact obtained by the reference EnKF. As expected from MB2011, stratospheric ozone assimilation with an EnKF was found to constrain the dynamical state globally with larger benefits for the UTLS and tropospheric jets regions. The benefit of an EnKF ozone assimilation on the ozone state is greater than on the dynamical state and extends to the mid-stratosphere. Stratospheric EnKF temperature assimilation constrains the UTLS more than the ozone assimilation. It is particularly effective on dynamical variables in the Polar Vortex and tropospheric jets. However, it provides a limited effect on the ozone error, restricted to the tropospheric jet region.

A first set of EnKS temperature or ozone assimilation experiments is performed to characterise the constraint on the forecast errors that the assimilation of posterior observations provides as a function of time lag between observation and analysis time. The EnKS assimilations display an increasing loss of impact with time lag, with temperature assimilation losing skill faster than ozone assimilation. It was found that EnKS assimilation with a 48-hour time lag started to produce minimally deleterious analysis increments, particularly for the assimilation of posterior temperature observations.

The second set of experiments was performed with a fixed-lag ($K=2$ with maximum time lag of 48 hours) EnKS and with an EnKF using a denser observation network composed of two additional MIPAS networks. The denser-network EnKF assimilates the same amount of observations on the same grid points as the EnKS, the only difference between the two experiments being the temporal distribution of the observations. This provides a stronger comparison for evaluating the EnKS. It was found that the

EnKS ozone assimilation provides significant further error reductions of the order of 10 to 15% on the reference EnKF wind, temperature and ozone analyses in parts of the lower stratosphere and troposphere. It also impacts significantly the Polar Vortex, which is not well constrained by the reference EnKF ozone assimilation. The EnKS temperature assimilation only improves significantly and on large scales the wind analyses, having only neutral results on the temperature and ozone analyses. The comparison of the EnKS temperature and ozone assimilation impact to the denser-network temperature EnKF assimilation impact revealed that the gains made by the ozone smoother on all analysed variables and by the temperature smoother on the winds were very valuable as they were similar, and in specific regions slightly better, to the gains made by the denser-network EnKF. The smoother does provide a viable alternative to the Filter in some contexts if additional current observations are not available.

This study was motivated in part by the need to improve long-term reanalyses representing stratospheric dynamics and chemistry. Considering the sparseness of observations in the stratosphere before and even during the recent satellite era, the proper assimilation of readily available observations posterior to the (re)analysis time could be precious. We believe that these preliminary results motivate further research on using (ensemble Kalman) smoothing methods for the improvement of reanalyses, for both traditional dynamical data assimilation as well as chemical (ozone) data assimilation. In addition, the application of the EnKS in the context of reanalyses can be done at marginal extra time costs as explained in Evensen (2003). Also, the explicit calculation of error covariances in EnKS allows to perform ‘variable localisation’ approaches to shorten the computational time. For example, as the EnKS temperature assimilation is ineffective in reducing the ozone errors, the time-lagged temperature-ozone error cross-covariances may be artificially nullified (with the appropriate verification that state balances are not compromised) as in MB2011.

The idealised environment in which this study is cast is highlighted. Although this study shows clear potential for EnKS, it needs to be confirmed in the presence of model errors, model biases and instrument biases. The absolute magnitude of error corrections as well as the optimal maximum time lag K will undoubtedly be affected if the model is imperfect. However, it is unclear if the relative error corrections (i.e. EnKS versus EnKF) would be significantly altered. The ensemble of background states from an imperfect model is also likely to yield error covariances that are unrealistic, which will in turn affect the analysis increments during the smoother analysis step. Further experiments in a more realistic experiment setting should thus be attempted.

6. Acknowledgements

The authors would like to thank the members of the SPARC-DA community, in particular Saroja Polavarapu and Richard Ménard, who have, through discussion, helped shape this study. We thank William Lahoz, Yves Rochon, Yi Huang, Emmanuel Cosme and an anonymous reviewer for their reviews. This study was funded by the Canadian Foundation for Climate and Atmospheric Sciences (C-SPARC project), the Natural Sciences and Engineering Research Council of Canada and the Canadian Space Agency.

7. Appendix A

The sequential treatment of observations reduces the size of the observation-space matrices and therefore speeds up the analysis computation. Another approach is used to reduce the computational cost without any underlying approximation. The localisation of the error covariance matrix $\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T$ by Schur product nullifies a certain proportion of its elements, thereby increasing its sparseness. This motivates the utilisation of a sparse-matrix numerical treatment. The background-error covariance matrix must simply be sparse enough to benefit from this method.

We use the method named ‘compressed row storage’ (CRS, sometimes called ‘compressed sparse row’ CSR, Dongarra, 2000) applied on the large matrix:

$$\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T \in \mathbb{R}^{n \times p}$$

where n is the total number of model state vector elements and p is the number of observations. The CRS technique reduces this $\mathbb{R}^{n \times p}$ into three vectors, which lengths are given by $\tilde{n}$, $\tilde{n}$ and $n+1$, respectively, where $\tilde{n}$ is the number of non-zero values in the matrix. We define the sparseness in $\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T$ as $S = 1 - \tilde{n}/(n \times p)$.

The main vector lists the successive non-zero row entries of $\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T$.

The second vector holds the column indices of each non-zero values. The third vector gives for each row of $\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T$ the index of the value of the main vector corresponding to the first non-zero value on this row. The third vector takes a length of $n+1$ as by convention we add the value $\tilde{n}+1$ after the n -th element. The total number of elements needed to represent $\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T$ decreases from $n \times p$ in the matrix treatment to $n+1+2 \times (1-S) \times n \times p$ with the CRS method. This technique becomes beneficial in terms of computer memory requirements for a $\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T$ matrix more than half-sparse ($S > 0.5$).

Besides the reduced storage need and smaller amount of arithmetic operations, the CRS method allows efficient parallel matrix-vector multiplications. In particular,

it allows to efficiently parallelise the product between $\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T$ and the scaled innovations $\tilde{\mathbf{d}}_{k'} = (\mathbf{H}_{k'}\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T + \mathbf{R}_{k'})^{-1}\mathbf{d}_{k'}$ in eq. (1). In our case, utilisation of the CRS method reduces the computational cost by a factor 7 and 8 for typical temperature and ozone assimilation, respectively, when compared to a parallelised matrix product method. Our sparseness is $S=0.527$ and $S=0.827$, respectively, for the matrix $\mathbf{P}(t_k, t_{k'})\mathbf{H}_{k'}^T$. Nevertheless, the sequential assimilation of batches of 100 observations is still faster than the assimilation of all observations at once with the CRS technique by about 40% for temperature assimilation and 10% for ozone assimilation.

References

- Bourqui, M. S., Taylor, C. P. and Shine, K. P. 2005. A new fast stratospheric ozone chemistry scheme in an intermediate general-circulation model. II: Applications to effects of future increases in greenhouse gases. *Q. J. Roy. Meteorol. Soc.* **131**(610, Part B), 2243–2261. DOI: 10.1256/qj.04.19.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L. and He, B. 2010. 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. DOI: 10.1175/2009MWR3158.1.
- Burgers, G., Van Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the Ensemble Kalman Filter. *Mon. Weather Rev.* **126**, 1719–1724. DOI: 10.1175/1520-0493(1998)126 <1719:ASITEK >2.0.CO;2.
- Charlton, A. J., O’Neill, A., Stephenson, D. B., Lahoz, W. A. and Baldwin, M. P. 2003. Can knowledge of the state of the stratosphere be used to improve statistical forecasts of the troposphere? *Q. J. Roy. Meteorol. Soc.* **129**(595, Part B), 3205–3224. DOI: 10.1256/qj.02.232.
- Cohn, S. E., Sivakumaran, N. S. and Todling, R. 1994. A Fixed-Lag Kalman Smoother for retrospective data assimilation. *Mon. Weather Rev.* **122**, 2838–2867.
- Cosme, E., Verron, J., Brasseur, P., Blum, J. and Auroux, D. 2012. Smoothing problems in a Bayesian framework and their linear Gaussian solutions. *Mon. Weather Rev.* **140**, 683–695. DOI: 10.1175/MWR-D-10-05025.1.
- Daley, R. 1995. Estimating the wind field from chemical constituent observations: experiments with a one-dimensional extended Kalman Filter. *Mon. Weather Rev.* **123**, 181–198. DOI: 10.1175/1520-0493(1995)123 <0181:ETWFFC >2.0.CO;2.
- Dessler, A. E. 2000. *The Chemistry and Physics of Stratospheric Ozone*. San Diego, CA: Academic Press.
- Dongarra, J. 2000. *Sparse Matrix Storage Formats, Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. Philadelphia, PA: SIAM, pp. 315–336.
- Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10143–10162. DOI: 10.1029/94JC00572.

- Evensen, G. 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynam.* **53**(4), 343–367. DOI: 10.1007/s10236-003-0036-9.
- Evensen, G. and van Leeuwen, P. J. 2000. An Ensemble Kalman Smoother for nonlinear dynamics. *Mon. Weather Rev.* **128**, 1852–1867. DOI: 10.1175/1520-0493(2000)128 <1852:AEKSFN >2.0.CO;2.
- Eyring, V., Butchart, N., Waugh, D. W., Akiyoshi, H., Austin, J. and co-authors. 2006. Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past. *J. Geophys. Res.* **111**, D22308. DOI: 10.1029/2006JD007327.
- Gerber, E. P., Butler, A., Calvo, N., Charlton-Perez, A., Giorgetta, M. and co-authors. 2012. Assessing and understanding the impact of stratospheric dynamics and variability on the earth system. *Bull. Am. Meteorol. Soc.* **93**, 845–859. DOI: 10.1175/BAMS-D-11-00145.1.
- Gottwald, G. A., Mitchell, L. and Reich, S. 2011. Controlling overestimation of error covariance in ensemble Kalman filters with sparse observations: a variance limiting Kalman filter. *Mon. Weather Rev.* **139**, 2650–2667. DOI: 10.1175/2011MWR3557.1.
- Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential Ensemble Kalman Filter for atmospheric data assimilation. *Mon. Weather Rev.* **129**, 123–137. DOI: 10.1175/1520-0493(2001)129 <0123:ASEKFF >2.0.CO;2.
- Houtekamer, P. L., Mitchell, H. L. and Deng, X. 2009. Model error representation in an operational Ensemble Kalman Filter. *Mon. Weather Rev.* **137**, 2126–2143. DOI: 10.1175/2008MWR2737.1.
- Hunt, B. R., Kostelich, E. J. and Szunyogh, I. 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. *Physica. D.* **230**(1–2), 112–126. DOI: 10.1016/j.physd.2006.11.008.
- Ide, K., Courtier, P., Ghil, M. and Lorenc, A. 1997. Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Jpn.* **75**(1B), 181–189.
- Khare, S. P., Anderson, J. L., Hoar, T. J. and Nychka, D. 2008. An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems. *Tellus A.* **60**, 97–112. DOI: 10.1111/j.1600-0870.2007.00281.x.
- Lawson, W. G. and Hansen, J. A. 2004. Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Weather Rev.* **132**, 1966–1981. DOI: 10.1175/1520-0493(2004)132 <1966:IOSADF >2.0.CO;2.
- Masutani, M., Woollen, J. S., Lord, S. J., Emmitt, G. D., Kleespies, T. J. and co-authors. 2010. Observing system simulation experiments at the National Centers for Environmental Prediction. *J. Geophys. Res.* **115**(D07101). DOI: 10.1029/2009JD012528.
- McLinden, C. A., Olsen, S. C., Hannegan, B., Wild, O., Prather, M. J. and co-authors. 2000. Stratospheric ozone in 3-D models: a simple chemistry and the cross-tropopause flux. *J. Geophys. Res.* **105**(D11), 14653–14665. DOI: 10.1029/2000JD900124.
- Milewski, T. and Bourqui, M. S. 2011. Assimilation of stratospheric temperature and ozone with an Ensemble Kalman Filter in a chemistry-climate model. *Mon. Weather Rev.* **139**, 3389–3404. DOI: 10.1175/2011MWR3540.1.
- Ngan, K. and Shepherd, T. G. 1999. A closer look at chaotic advection in the stratosphere. Part I: geometric structure. *J. Atmos. Sci.* **56**, 4134–4152. DOI: 10.1175/1520-0469(1999)056 <4134:ACLACA >2.0.CO;2.
- Pawson, S., Kodera, K., Hamilton, K., Shepherd, T. G., Beagley, S. R. and co-authors. 2000. The GCM-Reality Intercomparison Project for SPARC (GRIPS): Scientific issues and initial results. *Bull. Am. Meteorol. Soc.* **81**(4), 781–796. DOI: 10.1175/1520-0477(2000)081 <0781:TGIPFS >2.3.CO;2.
- Polavarapu, S., Shepherd, T. G., Rochon, Y. and Ren, S. 2005. Some challenges of middle atmosphere data assimilation. *Q. J. Roy. Meteorol. Soc.* **131**(613, Part C), 3513–3527. DOI: 10.1256/qj.05.87.
- Ravela, S. and McLaughlin, D. 2007. Fast ensemble smoothing. *Ocean Dynam.* **57**, 123–134. DOI: 10.1007/s10236-006-0098-6.
- Riishøjgaard, L. P. 1996. On four-dimensional variational assimilation of ozone data in weather-prediction models. *Q. J. Roy. Meteorol. Soc.* **122**(535, Part A), 1545–1571. DOI: 10.1002/qj.49712253505.
- Sacher, W. and Bartello, P. 2009. Sampling errors in Ensemble Kalman Filtering. Part II: Application to a Barotropic Model. *Mon. Weather Rev.* **137**, 1640–1654. DOI: 10.1175/2008MWR2685.1.
- Sakov, P., Evensen, G. and Bertino, L. 2010. Asynchronous data assimilation with the EnKF. *Tellus A.* **62**(1), 24–29. DOI: 10.1111/j.1600-0870.2009.00417.x.
- Szunyogh, I., Kostelich, E. J., Gyarmati, G., Kalnay, E., Hunt, B. R. and co-authors. 2008. A local ensemble Kalman filter data assimilation system for the NCEP global model. *Tellus A.* **60**(1), 113–130. DOI: 10.1111/j.1600-0870.2007.00274.x.
- Taylor, C. P. and Bourqui, M. S. 2005. A new fast stratospheric ozone chemistry scheme in an intermediate general-circulation model. I: description and evaluation. *Q. J. Roy. Meteorol. Soc.* **131**(610, Part B), 2225–2242. DOI: 10.1256/qj.04.18.
- van Leeuwen, P. J. and Evensen, G. 1996. Data assimilation and inverse methods in terms of a probabilistic formulation. *Mon. Weather Rev.* **124**, 2898–2913. DOI: 10.1175/1520-0493(1996)124 <2898:DAAIMI >2.0.CO;2.
- Whitaker, J. S., Compo, G. P. and Thépaut, J.-N. 2009. A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations. *Mon. Weather Rev.* **137**, 1991–1999. DOI: 10.1175/2008MWR2781.1.
- Whitaker, J. S., Hamill, T. M., Wei, X., Song, Y. and Toth, Z. 2008. Ensemble data assimilation with the NCEP global forecasting system. *Mon. Weather Rev.* **136**, 463–482. DOI: 10.1175/2007MWR2018.1.
- Zagar, N., Gustafsson, N. and Källén, E. 2004. Variational data assimilation in the tropics: the impact of a background-error constraint. *Q. J. Roy. Meteorol. Soc.* **130**(596, Part A), 103–125. DOI: 10.1256/qj.03.13.
- Zhu, Y., Todling, R., Guo, J., Cohn, S. E., Navon, I. M. and co-authors. 2003. The GEOS-3 retrospective data assimilation system: the 6-Hour Lag Case. *Mon. Weather Rev.* **131**, 2129. DOI: 10.1175/1520-0493(2003)131 <2129:TGRDAS >2.0.CO;2.