Simplified Kalman smoother and ensemble Kalman smoother for improving reanalyses

Bo Dong[1,2], Ross Bannister[1,2], Yumeng Chen[1,2], Alison Fowler[1,2], and Keith Haines[1,2]

[1]Department of Meteorology, University of Reading, Reading, UK
[2]National Centre for Earth Observation, University of Reading, Reading, UK

Correspondence: Bo Dong (bo.dong@reading.ac.uk) and Keith Haines (k.haines@reading.ac.uk)

Abstract. The paper presents a simplification to the Kalman smoother that can be run as a post processing step using only minimal stored information from a Kalman filter analysis, which is intended for use with large model products such as reanalyses of Earth system variability. A simple decay assumption is applied to cross time error covariances, and we show how the resulting equations relate formally to the fixed-lag Kalman smoother, and how they can be solved to give a smoother analysis along with an uncertainty estimate. The method is demonstrated in the Lorenz 1963 idealised system, being applied with both an extended Kalman smoother and an ensemble Kalman smoother. In each case the root mean square errors (RMSE) against truth, for both assimilated and unassimilated (independent) data, of the new smoother analyses are substantially smaller than for the original filter analyses, while being larger than for the full smoother solution. Typically 60% of the full smoother error reduction with respect to the filter, is achieved. The uncertainties derived for the new smoother also agree remarkably well with the actual RMSE values throughout the assimilation period. The ability to run this smoother very efficiently as a post processor should allow it to be useful for real large model reanalysis products, especially ensemble products that are already being developed by various operational centres.

1 Introduction

Data assimilation is widely used for making atmosphere and ocean predictions, providing a best estimate of the current state of the system, by combining the information from model forecasts with new observations available up to the current time. These state estimates are used for two purposes. First, they are used to initialise new model forecasts (from minutes to seasons ahead). Second, the state estimates can provide reanalysis products, representing a best estimate of past environmental conditions. This involves assimilating historical observational data using the newest models and assimilation methods available to us today, eg. Uppala et al. (2005); Balmaseda et al. (2013). However, assimilation systems suitable to initialise forecasts may be less than optimal when used for reanalysis production.

The main distinction we will draw is between sequential assimilation methods, which use only past data, as appropriate for forecasting, and temporal smoothing methods which can use past and future data to obtain a better state estimation, which may be more useful for reanalysis. Although 4DVar is used in operational meteorology and provides some temporal smoothing, it is only used to smooth within a short past data window when applied to initialise forecasts.

The archetypal sequential data assimilation approach, originally for linear systems, is the Kalman filter (KF), see e.g. Chapter 6 of Evensen et al. (2022)). While the basic KF is inefficient to use in applications with large state spaces (due to the difficulty of propagating very large error covariance matrices from one time to the next), the ensemble Kalman filter (EnKF, Evensen (1994)) is a popular and tractable approximation, which also allows non-linear systems to be treated. The EnKF exists in many flavours (e.g. in 'stochastic' (Burgers et al., 1998; Houtekamer and Mitchell, 1998) and 'square-root' forms (Bishop et al., 2001; Whitaker and Hamill, 2002)), which, like the standard KF, are all based on Bayes' theorem, and assume that errors in observed, prior, and posterior quantities are Gaussian distributed. Under the EnKF, the prior distribution is described by an ensemble of model forecast states, and the posterior distribution by an ensemble of posterior states found by assimilating current observational information. This makes the EnKF suitable for model-based forecasting systems.

Ensemble Kalman filters, applied either on their own, or hybridised with variational approaches, have shown success in numerous geophysical applications: for example in meteorological applications with the Canadian forecasting system (Houtekamer et al., 2005), with the NCEP global (Hamill et al., 2011; Wang et al., 2013) and regional (Pan et al., 2014) models, and the WRF model (Zhang and Zhang, 2012); in ocean analysis (van Velzen et al., 2016); in ocean and sea ice analysis (Sakov et al., 2012); in atmospheric chemical analysis (Skachko et al., 2016); and in surface trace gas analysis (Feng et al., 2009).

However all *filtering* problems, as noted above, include only past and present observational data, but this can be extended to a *smoothing* problem, which also uses observations within a future time window, usually referred to as the lag (e.g., Todling and Cohn, 1996). Kalman smoothers (KS) are made possible by the construction of cross-time error covariance matrices that link the observations at future times with the current analysis, often up to some maximum lag time. A smoother analysis will therefore use more observational data than a filter analysis and should therefore provide a more accurate state estimate. This would seem particularly relevant for reanalysis applications when full time series of past and future observations are available for constructing system states. Various smoothers have been proposed for use in the geosciences (e.g. Evensen and van Leeuwen (2000); Ravela and McLaughlin (2007); Bocquet and Sakov (2014)). These have been proposed for both reanalyses, e.g. Zhu et al. (2003), and parameter estimation, e.g. Evensen (2009). Just like the EnKF, the ensemble Kalman smoother (EnKS) uses an ensemble of model realisations to estimate the error distribution of the model forecasts, which can be very efficient.

The KS has been shown to be effective in various applications. For example, Zhu et al. (2003) designed a meteorological reanalysis system using a fixed-lag KS, and Khare et al. (2008) with longer lags; Cosme et al. (2010) developed an EnKS for ocean data assimilation; and Pinnington et al. (2020) used KS techniques for land surface analysis. These applications all rely on calculating the cross-time covariance matrix (either explicitly or implicitly) for the smoothing.

For large operational forecasting and reanalysis systems, especially for high resolution global ocean, climate or Earth system models, which contain substantially long timescale processes of up to weeks or months, running a smoother with a reasonably long lag could be very expensive in computation and thus impractical. Even for the relatively cost effective EnKS, the ensemble anomaly matrix for each time-step could consist of billions of elements, which takes large chunks of computer memory space.

In addition, it would not be easy to retrofit smoothing code into an operational data assimilation system that has been developed over decades primarily for initialising forecasts. For reanalysis products developed in this way a simpler post-processing approach to smoothing could be very valuable.

Dong et al. (2021) recently proposed a new smoother designed to be used offline through post processing of a filter analysis. It was based on the simplifying physical assumption of decaying error covariances across time, resulting a formulation similar to an autoregressive model. This smoother uses only the filtering increments, without needing to seek other information. The method was shown capable of improving the Met Office GloSea5 ocean reanalysis (MacLachlan et al., 2015), reducing RMSE against both assimilated and independent data, and producing more realistically smooth temporal variability for important quantities such as ocean heat content.

In this study we further explore the characteristics of the Dong et al. (2021) smoother as an approximation to the Kalman smoother framework. We demonstrate that with proper assumptions, this method can be reproduced within an extended Kalman smoother and an Ensemble Kalman smoother, in the latter case retaining the benefit of the ensemble's flow dependent covariances. We also extend the theory to show how the uncertainty estimates of the smoothed analyses can be obtained from post processed filter information. The full and approximate smoother approaches are implemented in the Lorenz 1963 model and the results compared. We show that the Dong et al. (2021) post processing method produces intermediate error results between the filter and the full Kalman smoother, without costing significant computer time or adapting the filter codes.

Section 2 derives the Dong et al. (2021) smoother from the full extended KS equations, and the theory is extended to include the simplified smoothing of uncertainty estimates. Section 3 presents the implementation of the extended filter/smoother, in both the full and approximated forms, in the Lorenz 1963 system. Section 4 adapts the methods presented earlier for application for the EnKF/S and presents both full and approximated results for these methods, also in the Lorenz 1963 system. Section 5 is a discussion of the applicability of these approximations in larger models, were the simplifications should allow for post processor smoothing of operational reanalysis products. Section 6 presents conclusions and recommendations for stored variables that would allow post processed smoothing in larger systems. The appendix reviews the conventional KF and fixed-lag KS equations, and shows more formally where approximations are applied which lead to our simplified smoothing algorithms.

2 Methods

2.1 The simple smoother method

In Dong et al. (2021), a simple smoother method (hereafter referred to as DHM) was presented for application in operational ocean reanalysis products, where the original analysis had been performed with a purely time sequential approach, as in forecasting situations when future data are never available. This simple approach was designed to use the archive of increments to create a post hoc smoothing of the original reanalysis. Dong et al. (2021) showed the positive impact of this smoothing on both a full ocean reanalysis and also on the low-dimensional Lorenz 1963 system. The algorithm is as follows. Let A_t

be the forward sequential (filtering) analysis at time t, and I_t be the analysis increment field used to produce A_t. The smoother solution at time t is denoted S_t. The smoother algorithm is then written as follows, firstly for S_0:

$$S_0 = A_0 + \gamma_a I_1 + \gamma_a^2 I_2 + \gamma_a^3 I_3 + \gamma_a^4 I_4 + \ldots, \tag{1}$$

where $0 < \gamma_a < 1$ is the increment decay rate per analysis time window, so that analysis increments from future analysis times decay in their influence on S_0. Similarly for S_1:

$$S_1 = A_1 + \gamma_a I_2 + \gamma_a^2 I_3 + \gamma_a^3 I_4 + \gamma_a^4 I_5 + \ldots. \tag{2}$$

By rearrangement we get

$$S_0 = A_0 + \gamma_a(S_1 - A_1 + I_1) = A_0 + SI_0, \tag{3}$$

with

$$SI_0 = \gamma_a(SI_1 + I_1), \tag{4}$$

where $SI_t = S_t - A_t$ defines the 'smoother increment'. These recursive relationships allow the smoother to be applied as a post processing algorithm, ~~which is~~ run backwards in time starting with the final sequentially analysed time window, using the stored archive of ~~analysis~~ filter increments. It will be convenient later to define the increment decay per model timestep which we will write as just γ where $\gamma^N = \gamma_a$ and N is the number of model timesteps between filter analyses. Later we will assume each analysis window consists of one timestep ~~, therefore~~ with $N = 1$ and $\gamma = \gamma_a$. The decay timescale τ associated with the smoothing is ~~then~~ given, in ~~model~~ timesteps δt, by

$$\tau = -\delta t / (\ln \gamma), \tag{5}$$

which is effectively a measure of the smoother lag ~~which is not given an explicit maximum cutoff.~~. This equation can also be re-arranged to be $\gamma = e^{-\frac{\delta t}{\tau}}$ implying an exponential increase of the forecast error. These smoother Eqs. 1 and 2 do not have a fixed lag cutoff, unlike the fixed-lag smoothers discussed below.

~~Below we~~ We now discuss how this simple smoother is related to the conventional KS approach (a more formal proof of ~~the~~ equivalence is given in the appendix).

2.2 Extended Kalman filter and extended Kalman smoother

We start from the classical extended Kalman filter (ExtKF) and fixed-lag extended Kalman smoother (ExtKS) ~~formulation~~ formulations, in which a tangent linear model is used for error covariance propagation when the model is nonlinear. ~~We will use superscripts~~ Superscripts f, a, s ~~to~~ describe *filter forecasts*, *filter analyses*, and *smoother analyses* respectively. The analysis of the Kalman filter at time k is given by

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k^a(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}_k^f)), \tag{6}$$

where the subscript represents the time step, $\mathbf{x} \in \mathbb{R}^n$ is the n-dimensional state vector, $\mathbf{y} \in \mathbb{R}^m$ is the observations, $\mathcal{H}_k$ is the observation operator. In the ExtKF, the observation operator and the model can both be nonlinear, where the state vector evolves with a model $\mathbf{x}_k = \mathcal{M}(\mathbf{x}_{k-1})$. ~~The nonlinear operators have to be replaced by their tangent linear approximations in the forecast and analysis steps, leading to the linear transformations $\mathbf{M}_k \in \mathbb{R}^{n \times n}$ and $\mathbf{H}_k \in \mathbb{R}^{m \times n}$ for the tangent linear model and tangent linear observations matrix respectively.~~

$\mathbf{K}_k^a \in \mathbb{R}^{n \times m}$ is the Kalman gain for the analysis, which is given by;

$$\mathbf{K}_k^a = \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} (\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}, \tag{7}$$

with $\mathbf{P}_k^f \in \mathbb{R}^{n \times n}$ being the forecast error covariance matrix, $\mathbf{R}_k \in \mathbb{R}^{m \times m}$ being the observation error covariance, ~~all at the current~~ and $^{\mathrm{T}}$ being the transpose operator, all at timestep k. ~~The~~ Here, instead of the non-linear observation operator, the tangent linear approximation, $\mathbf{H}_k \in \mathbb{R}^{m \times n}$, is used, and the tangent linear model, $\mathbf{M}_k \in \mathbb{R}^{n \times n}$, of the nonlinear model is used for the propagation of the forecast error covariance matrix, $\mathbf{P}_k^f = \mathbf{M}_k \mathbf{P}_{k-1}^f \mathbf{M}_k^T$. Finally the analysis error covariance can be derived from the forecast error covariance as follows:

$$\mathbf{P}_k^a = \mathbf{P}_k^f - \mathbf{K}_k^a \mathbf{H}_k \mathbf{P}_k^f. \tag{8}$$

~~which we will return to later.~~

For the fixed-lag ExtKS, Todling and Cohn (1996) (hereafter TC96) derive *backward looking* equations for the smoother, which run interleaved with every filter timestep, however here we will present *forward looking* equations aimed at expressing the fully smoothed state, including contributions from multiple future filter steps, as presented for the simple smoother in Eqs. (1)-(5). The full equivalence between TC96's and our notation is demonstrated in the appendix.

The contributions from observations at ~~time step~~ timestep $k + \ell$ to the smoother solution at timestep k can be written in the same Kalman gain notation as;

$$\mathbf{K}_{k,k+\ell}^s (\mathbf{y}_{k+\ell} - \mathcal{H}_{k+\ell}(\mathbf{x}_{k+\ell}^f)). \tag{9}$$

We note that if filter states are only stored at some assimilation frequency, eg. once per day, the index ℓ ~~could be defined as future analysis timesteps rather than model timesteps if data are only introduced at regular analysis intervals.~~ can be defined at the same frequency as these filter increments, as smoother states are only required at the same frequency as the stored filter states.

The full smoother solution for timestep k ~~looking forward,~~ is then obtained by the summation of smoother increments ~~for~~ from all future timesteps (here assumed truncated to maximum lag L) as;

$$\mathbf{x}_k^s = \mathbf{x}_k^a + \sum_{\ell=1}^L \mathbf{K}_{k,k+\ell}^s (\mathbf{y}_{k+\ell} - \mathcal{H}_{k+\ell}(\mathbf{x}_{k+\ell}^f)) \tag{10}$$

(see ~~the appendixfor the derivation~~appendix). The cross time smoother gain matrix is simply a modified version of the standard ExtKF gain, and can be written~~as~~;

$$\mathbf{K}_{k,k+\ell}^s = \mathbf{P}_{k,k+\ell} \mathbf{H}_{k+\ell}^{\mathrm{T}} (\mathbf{H}_{k+\ell} \mathbf{P}_{k+\ell}^f \mathbf{H}_{k+\ell}^{\mathrm{T}} + \mathbf{R}_{k+\ell})^{-1}. \tag{11}$$

There is a subtlety here because in the ~~full TC96 smoother~~ appendix we will see that the cross time error covariance $\mathbf{P}_{k,k+\ell}$ is not independent of $\mathbf{P}_{k,k+\ell-1}$. However, this will not be ~~relevant~~ the case in the simple smoother approximation as applied below.

155   ~~At this point we~~ To introduce the key simple smoother approximation ~~when~~ we re-write the cross time error covariance as a ~~simple decay rate~~ decay rate, and consequently also neglect any inter-dependence of smoother contributions from different times:

$$\mathbf{P}_{k,k+\ell} \approx \gamma^\ell \mathbf{P}^f_{k+\ell}, \tag{12}$$

which is equivalent to assuming

160   $$\mathbf{K}^s_{k,k+\ell} \approx \gamma^\ell \mathbf{K}^a_{k+\ell}. \tag{13}$$

This Eq. (13), when substituted into Eq. (10), clearly expresses the approximation being made to recover the simple smoother solution from the ExtKS equations:

$$\mathbf{x}^s_k \approx \mathbf{x}^a_k + \sum_{\ell=1}^{L} \gamma^\ell \mathbf{K}^a_{k+\ell}(\mathbf{y}_{k+\ell} - \mathcal{H}_{k+\ell}(\mathbf{x}^f_{k+\ell})), \tag{14}$$

~~which when~~ When using Eq. (6) ~~then~~ this gives

165   $$\mathbf{x}^s_k \approx \mathbf{x}^a_k + \sum_{\ell=1}^{L} \gamma^\ell \mathbf{I}_{k+\ell}, \tag{15}$$

where $\mathbf{I}_{k+\ell} = \mathbf{x}^a_{k+\ell} - \mathbf{x}^f_{k+\ell}$, reproducing the simple smoother under the additional assumption that $L >> \tau$ in Eq. (5). The smoother is now defined entirely in terms of the sequential analysis increments which allow post processing from an archive of increments from the sequential filter run. Another way to interpret this approximation is to say that the spatial and temporal error covariances in the KS are assumed separable, with the spatial (and cross-variable) error covariances being determined by

170   the KF equations, but the temporal covariances (from times $k + \ell$ to k) being approximated by a simple decay. We will return to this description later when we seek to extend the approximations to the ~~ensemble KS~~ EnKS case.

It is also possible to make the equivalent approximations to the smoothed uncertainties. For each smoother increment introduced, Eq. (9), there will be a corresponding reduction in the smoother error covariance given by;

$$-\mathbf{K}^s_{k,k+\ell}\mathbf{H}_{k+\ell}\mathbf{P}_{k+\ell,k}, \tag{16}$$

175   so that the fully smoothed error covariance can be written as

$$\mathbf{P}^s_k = \mathbf{P}^a_k - \sum_{\ell=1}^{L} \mathbf{K}^s_{k,k+\ell}\mathbf{H}_{k+\ell}\mathbf{P}_{k+\ell,k}, \tag{17}$$

(see ~~the~~ appendix). Then, making the simple smoother approximation, Eqs. (12) and (13) ~~, here gives~~ give

$$\mathbf{P}^s_k \approx \mathbf{P}^a_k - \sum_{\ell=1}^{L} \gamma^{2\ell} \mathbf{K}^a_{k+\ell}\mathbf{H}_{k+\ell}\mathbf{P}^f_{k+\ell}. \tag{18}$$

Now returning to use Eq. (8) we finally obtain

$$180 \quad \mathbf{P}_k^s \approx \mathbf{P}_k^a - \sum_{\ell=1}^{L} \gamma^{2\ell} \mathbf{IP}_{k+\ell}, \tag{19}$$

where $\mathbf{IP}_{k+\ell} = \mathbf{P}_{k+\ell}^f - \mathbf{P}_{k+\ell}^a$ are the filter error covariance increments, mirroring Eq. (15), the simple smoother equations for the increments. The smoothing equations (15) and (19) could clearly both be written in recursive format like Eq. (3), for ease of post processing. In the following sections we investigate how well these approximations work through comparisons in the Lorenz 1963 system.

3 Extended Kalman Smoother experiments in the Lorenz 1963 system

3.1 Assimilation setup

A twin experiment using the Lorenz 1963 model (Lorenz, 1963), hereafter L63, was carried out to evaluate the smoother. The L63 uses the classical setup with model equations ~~of~~:

$$190 \quad \frac{dx}{dt} = \sigma(y - x) \tag{20}$$

$$\frac{dy}{dt} = \rho x - y - xz \tag{21}$$

$$\frac{dz}{dt} = xy - \beta z, \tag{22}$$

where the standard model parameters are chosen as $\sigma = 10; \rho = 28; \beta = \frac{8}{3}$. All experiments are run for 20 time units, with each time unit consisting 100 timesteps of 0.01. We performed a 'truth' run first with x, y and z values of 5 as the initial condition. Observations are assigned for x, y with frequency of ~~every~~ 5 and 20 timesteps respectively ~~and error standard deviation~~, and Gaussian errors added with standard deviations of 2. No observations are taken for z.

Dong et al. (2021) used 3DVar for assimilation into L63 and they used a fixed background error covariance ~~from a climatological~~ based on variability from another L63 run. Here we ran the extended Kalman filter (ExtKF) and extended Kalman smoother (ExtKS) with 100 different initial conditions, ~~but~~ where the background error covariance is modelled. However with this assimilation frequency we found that to avoid filter divergence we ~~use~~ needed a hybrid forecast error covariance, retaining a 5% weighting of the L63 climatological covariances in the background error. The ~~fixed lag smoother uses~~ smoother uses fixed lag $L = 40$ timesteps, as we found that ~~errors of this lag are smaller than other lags in our L63~~ this lag gives smallest errors in our experiments.

The simple smoother (DHM), Eqs. (1)-(5), was executed with $\gamma = 0.9$ when the smoothing results have smallest error ~~as~~ compared to other γ values (experiments not shown). ~~The reasons for this appear later.~~ We also ran a modified Kalman smoother (MKS) using the approximated cross time Kalman gain as in Eq. (13). This is implemented by directly substituting the approximation into the full KS equations described in the appendix, ~~in order~~ and is used to demonstrate DHM equivalent results. The uncertainty ~~estimation~~ estimates for the MKS smoother is also obtained in the same way.
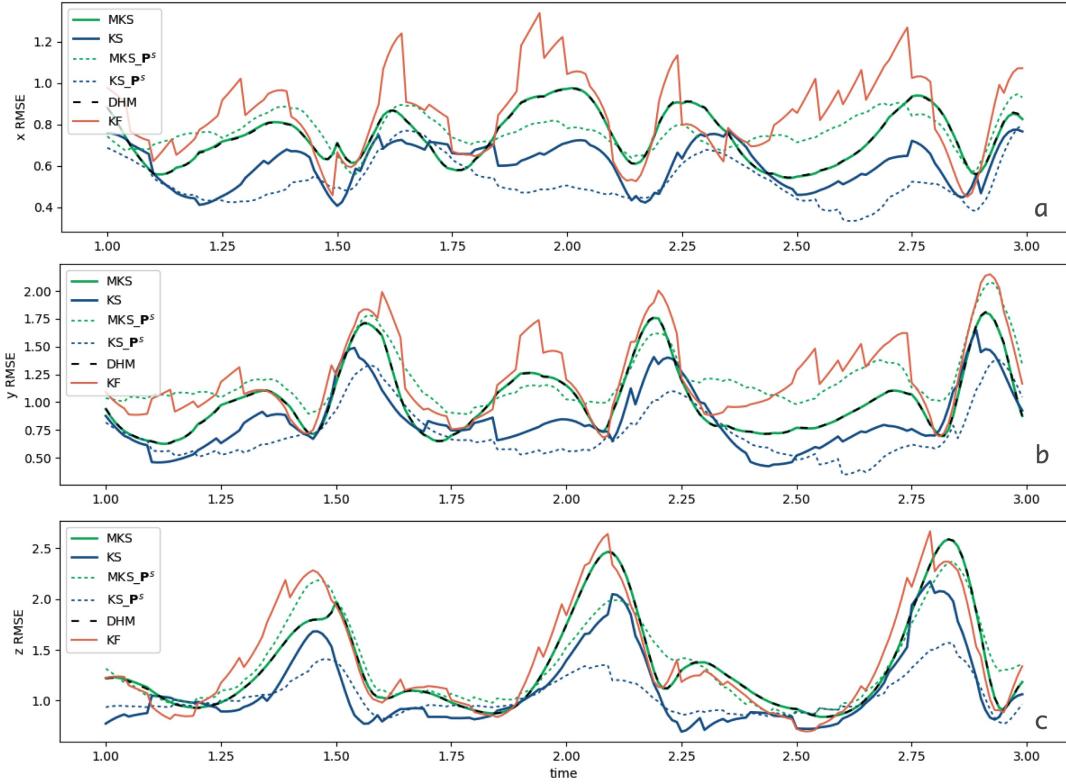
Figure 1. The RMSE timeseries for the ExtKF and ExtKS along with the modified MKS and the simplified DHM smoother in the L63 system for (a) x, (b) y and (c) z. The RMSE are averaged across 100 independent assimilation runs starting from different initial conditions, assimilating observations from the same 20 time unit truth run. Only 2 time units of the run are shown to allow the behaviour around analysis times to be clearly seen. Dotted green lines and blue lines are the posterior uncertainty STD estimates for the MKS and ExtKS respectively.

3.2 ExtKS Assimilation Results

Across the 100 ~~member ensemble~~assimilation runs, we calculated the root mean square error (RMSE) timeseries against the truth for each smoothing method. Figure 1(a,b,c) show a portion of the (x, y, z) RMSE timeseries respectively, for the filter and the different smoothing methods in the ~~thicker lines. Without any smoothing, for~~ solid lines. For most timesteps the KF errors are larger than the smoother errors. The full ExtKS has smallest errors, however the DHM and MKS are almost identical and lie in between those for the KF and KS. Also on Figure 1 are dashed ~~blue and green~~ lines representing the ~~ensemble mean average of the~~ smoothed standard deviation (STD) uncertainty estimates for the ExtKS ~~(~~runs (blue, Eq. (17)) and MKS ~~(~~runs (green, Eq. (19))~~respectively.~~, respectively. It should be emphasised that these uncertainty estimates are calculated entirely independently of the actual truth values themselves, which would not be known in a real assimilation problem. The level of agreement between these uncertainty estimates and the ~~true RMSE~~ RMSE against truth is remarkable.
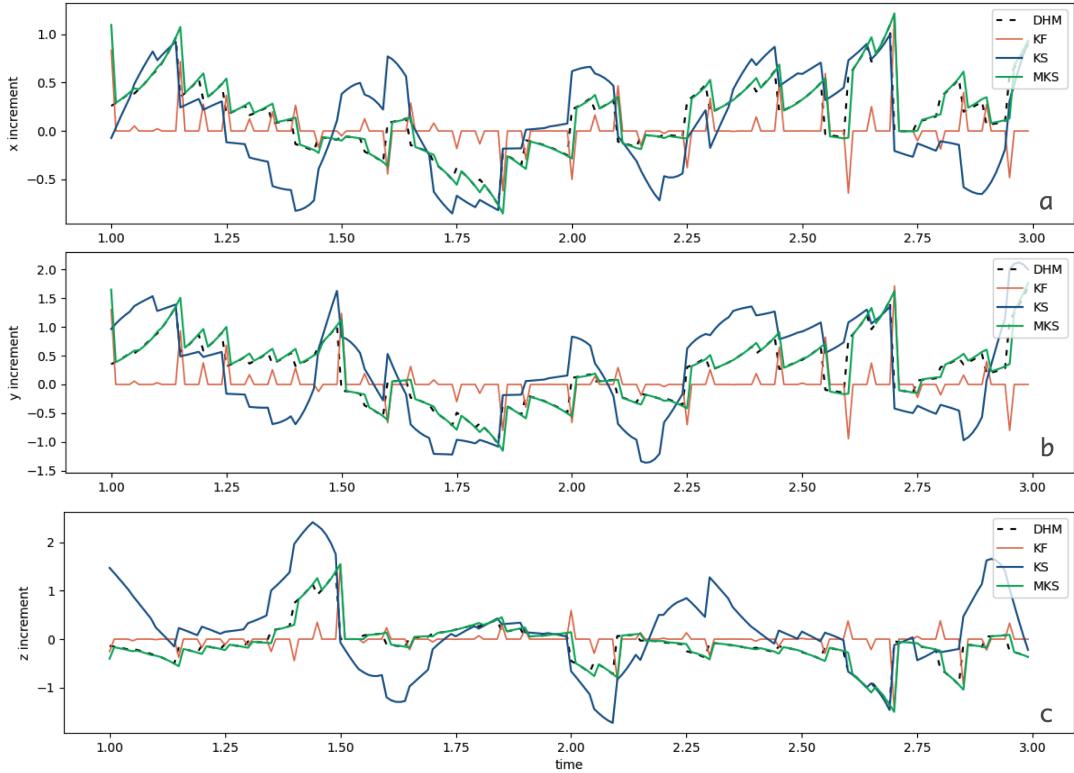
Figure 2. As in Fig. 1 but the filter and smoother increments are shown, where the smoother increments where applied are additive to the filter increments.

Time mean RMSE for x, y, z are summarised in Table 1, along with the uncertainty STD, where calculated, over the entire 20 time units of the runs. Both DHM and MKS provide an improvement on the ExtKF results by ~60-70% relative to the
220   ExtKS improvements for x and y, although the RMSE for z is not reduced in DHM and MKS. This is perhaps because the instantaneous error covariances between z and the assimilated x, y variables, as used in DHM, are insufficient to improve z, whereas the full ExtKS allows some history of x, y evolution to be used in deriving z smoothing increments. The RMSE numbers in parentheses are evaluated only at filter update timesteps where observation data are assimilated. These errors are smaller than the ~~all time~~ all-timesteps RMSE by $\sim 5\%$, as a result of the data assimilation at these timesteps. This is consistent
225   with the RMSE time series in Fig. 1~~where the red line~~, where the Kalman filter (red line) usually declines sharply where data are available. ~~For the~~ The smoother solutions, however, ~~the RMSEs are not only reduced at these times but as smoothers, by design, also yield~~ also yield much improved analyses in between observation timesteps.

Figure 2(a,b,c) show the actual x, y, z increments respectively, being introduced by the ExtKF and the different smoothers, through the analysis times. In each case the smoother increments are additional to the filter increments, which appear clearly
230   as red spikes every 5 timesteps where data are available. Smoother increments in between analysis times can be seen, with
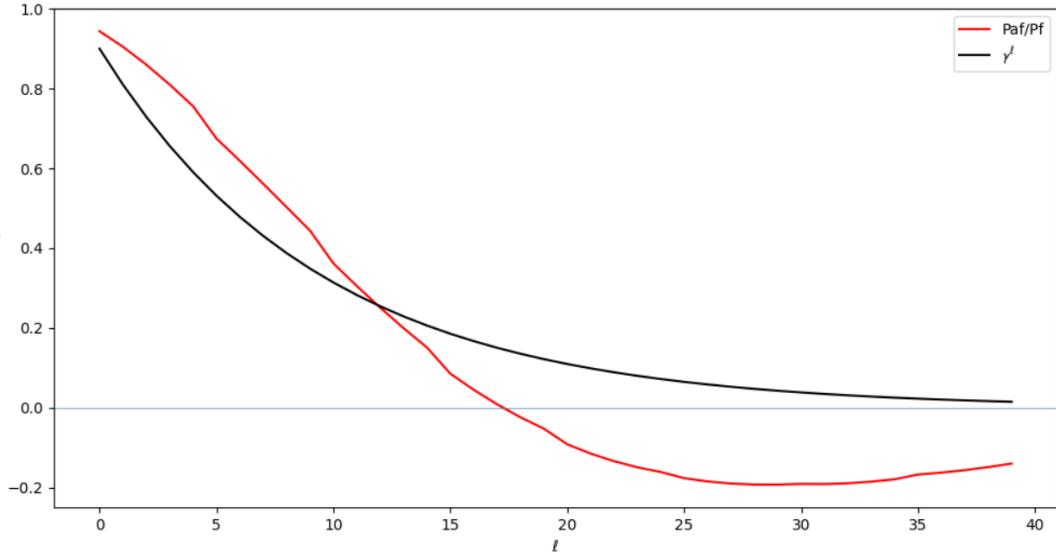
Figure 3. Cross time error covariance decay rates for the ExtKS and the simple smoother. For the ExtKS the y axis is the smoother cross time error covariance divided by the filter forecast error covariance, averaged for all times over the 20 time unit L63 run. This is compared to the decay of γ^ℓ used in the simple smoother.

DHM and MKS increments being virtually identical, ~~and~~ decaying backwards in time from each filter increment. The ExtKS increments are more complex, sometimes being similar to ~~the~~ DHM, but sometimes they can be considerably larger.

If we look at the mean ratio of cross time error covariances relative to the filter forecast error covariances, in comparison to the simplified γ decay representation across time in Fig. 3, we can understand something of the performance of the smoothers. We do not expect these to be identical because the full cross time smoother covariances for larger lags, ℓ, take account of increments from intermediate times. For small ~~lags~~ lag values the average cross time error decay rates are fairly similar, however for larger lag values the model derived cross time error covariance ~~can take~~ on average takes the opposite sign. This happens on a similar timescale to the short oscillation period of x, y in L63, and is associated with the growing amplitude of these oscillations, reaching larger and smaller x, y values before phase lobe transitions. This is a very model specific behaviour and the simple ~~smoothers~~ smoother's γ decay error covariances cannot represent this. This also explains why larger γ values make the simple smoother worse in L63, because ~~larger~~ positive cross time error covariances ~~would then be used~~ are used for larger lags when negative cross time error covariances ~~should be used for larger lags~~ are more appropriate.

The key point is that the simplified smoother DHM provides substantial improvement over the ExtKF while incurring very little computational cost (no tangent linear model (TLM) runs and no storage of cross time error covariances) compared to the ExtKS. The DHM smoother can therefore be applied entirely through post processing of the filter results. While this was demonstrated in Dong et al. (2021), here we show more clearly how the equivalent MKS approximation is derived from

the ExtKS equations and we also show how the smoothed uncertainties can be cheaply post processed and still give useful information.

In the next section we extend the decay assumption for cross time error covariances to apply to the ensemble Kalman filter/smoother equations which are much more relevant to large nonlinear models where direct modelling of error covariances across time is in any case infeasible.

Method	RMSE(x)	STD(x)	RMSE(y)	STD(y)	RMSE(z)	STD(z)
KF	1.13 (1.09)	0.93	1.79 (1.73)	1.50	1.64 (1.62)	1.41
KS	0.75 (0.75)	0.50	1.10 (1.09)	0.74	1.36 (1.35)	0.98
MKS	0.87 (0.87)	0.76	1.29 (1.29)	1.25	1.64 (1.64)	1.28
DHM	0.87 (0.87)	0.76	1.29 (1.29)	1.25	1.64 (1.64)	1.28

Table 1.

Time mean RMSE against truth for the ExtKF and ExtKS, along with the modified MKS and the simplified DHM smoother for each variable in the L63 system (see also legend for Fig. 1). Numbers in parentheses are mean RMSE at observation timesteps only (not independent data). The time mean of the standard deviations calculated for the uncertainties are also shown as STD. Time averaging is now over the entire 20 time units of each assimilation run.

4 Ensemble Kalman Smoother experiments in the Lorenz 1963 system

4.1 Approximating ensemble error covariances

In the ExtKF/S, a TLM propagates the flow-dependent error statistics which are then used to calculate increments. However, the TLM ~~is not always reliable~~ reliability declines sharply with propagation time, for a system as non-linear as the L63 model. The ensemble Kalman filter (EnKF) ~~gives~~ can then give better results by estimating the error statistics with a finite ensemble of state realisations propagated by the full nonlinear model, rather than by a TLM. This ~~improves~~ can then improve the quality of the forecast error covariance matrix. However, the update gains for the EnKF and ensemble Kalman smoother (EnKS) are defined identically to Eqs. (7) and (9) respectively, although the error covariances, $\mathbf{P}_k^{\mathrm{f}}$ (the error covariance at timestep k) and $\mathbf{P}_{k,k+\ell}$ (the error covariance between timesteps k and $k+\ell$) are calculated differently, being emulated from the limited ensemble of state vectors whose variability represents the uncertainty of the system. While ensemble filter methods ~~are starting to be~~ have been adopted for larger environmental models, these have not generally added smoother steps because the cost to store, update and apply posterior ensemble covariances still makes ensemble smoother methods generally infeasible.

However, these constraints can again be overcome by retaining the EnKF flow dependent ensemble spreads to represent current errors, while making a simple decay approximation for the time shift error covariances, similar to our modified ExtKS in Sect. 2.2. For comparison purposes, we demonstrate this method by starting with the forecast error covariance estimate in general ~~form of EnKS~~ EnKS form in state space, keeping the same notation as in ~~the~~ MKS, although the actual computation is

performed in ~~the~~ ensemble space as in the ensemble-transform Kalman filter (ETKF; e.g., Bishop et al., 2001; Zeng and Janjić, 2016).

$$\mathbf{P}_k^{\mathrm{f}} = \mathbf{X}_k^{\mathrm{f}} \mathbf{X}_k^{\mathrm{f}\ \mathrm{T}} \tag{23}$$

where

$$[\mathbf{X}_k^{\mathrm{f}}]_i = \frac{(\mathbf{x}_k^{\mathrm{f}})_i - \overline{\mathbf{x}}_k^{\mathrm{f}}}{\sqrt{N_{\mathrm{e}} - 1}} \tag{24}$$

is the normalised anomaly ~~around the mean~~ of the ith ensemble member from the mean, in an N_{e} member ensemble of forecasts.

The first step in the ensemble filter is to update the ensemble mean:

$$\overline{\mathbf{x}}_k^{\mathrm{a}} = \overline{\mathbf{x}}_k^{\mathrm{f}} + \mathbf{K}_k^{\mathrm{a}}(\mathbf{y}_k - \mathcal{H}(\overline{\mathbf{x}}_k^{\mathrm{f}})), \tag{25}$$

and the second step is to update the uncertainty, using Eq. (8). Then in order to regenerate the ensemble of analysis perturbations the ETKF uses the transformation:

$$\mathbf{X}_k^{\mathrm{a}} = \mathbf{X}_k^{\mathrm{f}} \mathbf{T}, \tag{26}$$

where $\mathbf{T}$ is chosen to ensure Eq. (8) is satisfied.

~~The EnKS can be viewed with EnKF as the basis, but with additional needs for the cross time error covariances for smoothing updates. Similar to~~ Then following the EnKF, the cross time covariances in the EnKS can be expressed in state space as

$$\mathbf{P}_{k,k+\ell} = \mathbf{X}_k \mathbf{X}_{k+\ell}^{\mathrm{f}\ \mathrm{T}}. \tag{27}$$

As explained in the appendix, ~~in the full smoother these~~ the full cross time error covariances are calculated between the filter forecast ($\mathbf{X}_{k+\ell}^{\mathrm{f}}$ in the EnKS) and previous partially smoothed states ($\mathbf{X}_k$ again in the EnKS), which require both past ensemble means and error covariances to be repeatedly smoothed. However, this is not necessary for the modified (simple) ensemble smoother, which we ~~will~~ here call MEnKS. The ensemble mean smoothing using Eqs. (13) and (10) can be written:

$$\overline{\mathbf{x}}_k^{\mathrm{s}} \approx \overline{\mathbf{x}}_k^{\mathrm{a}} + \sum_{\ell=1}^{L} \gamma^{\ell} \mathbf{K}_{k+\ell}^{\mathrm{a}}(\mathbf{y}_{k+\ell} - \mathcal{H}_{k+\ell}(\overline{\mathbf{x}}_{k+\ell}^{\mathrm{f}})), \tag{28}$$

and then Eq. (18) can be used to obtain the smoothed uncertainties.

This is a great simplification because to perform full ensemble smoothing would require the whole past ensemble to be stored at all times and re-processed. Equation (19) suggests that the error covariance increments must also be stored during the EnKF filter phase which ~~would~~ could still be a large storage requirement for a big model~~but~~, however in fact only the diagonal elements of $\mathbf{P}^s$ are likely to be of interest, i.e. the uncertainty variance of the state fields, or even just a subset of these, so only a smaller set of uncertainty increments may ~~need to be stored~~ be of interest to store for post processing through Eq. (19). In the next subsection we show results from applying these approximations in L63.

4.2 EnKS Assimilation Results

Using the same L63 twin experiment as in Section 3 we solve both a full smoother (EnKS), and using the modified (MEnKS) algorithm Eq. (28). To be consistent with the extended KS configuration, we use an ensemble size of 100 and a fixed lag of 40 timesteps for smoothing. Figures 4(a,b,c) show the x, y, z RMSE and STD uncertainties, respectively, from these Ensemble Filter and Smoother runs in the same format as Figures 1 for the Extended Kalman filter.

These ensemble results are seen to produce lower RMSE than the equivalent ExtKF/KS results, c.f. Fig. 1, demonstrating the superiority of the ensemble assimilation method in dealing with the nonlinearity of the L63 model. Again the EnKS substantially reduces the RMSE compared to the filter, and again the approximated simple ensemble filter MEnKS gives intermediate RMSE results with much less computational effort than the full EnKS. Although not as optimal as the EnKS, the simplified MEnKS shows much smoother temporal evolution of the RMSE ~~, which will~~ than the filter, which would be a significant improvement if eg. applied to an ocean reanalysis. The post processed uncertainty estimates also reproduce a reasonable estimate of the true RMSEs of the smoothed ensemble mean analyses.

Figures 5(a,b,c) show the mean x, y, z increment timeseries, respectively, for the ensemble filter and the 2 smoothers. The increments are smaller than those from the ExtKF/KS ~~.~~ in Fig. 2, reflecting the improved assimilation approaches. Table 2 summarises the average RMSE and STD uncertainty results over the full 20 time units of the ~~Ensemble~~ ensemble runs.

Method	RMSE(x)	STD(x)	RMSE(y)	STD(y)	RMSE(z)	STD(z)
EnKF	0.82 (0.79)	0.66	1.26 (1.22)	1.03	1.23 (1.21)	0.99
EnKS	0.50 (0.48)	0.39	0.69 (0.69)	0.56	0.90 (0.89)	0.74
MEnKS	0.66 (0.66)	0.57	1.02 (0.96)	0.87	1.15 (1.15)	0.90

Table 2.

Time mean RMSE and STD uncertainties for each variable in EnKF, EnKS and MEnKS in the L63 model, averaged over time unit 1-20; numbers in parentheses are RMSE calculated at timesteps with observations only (no independent data comparison). Lag=40, γ=0.9

5   Discussion

The aim of this paper is clearly not to present an improved data assimilation approach for simple models but to explore traceable simplifications to current assimilation approaches which could be applied to high-dimensional models. In particular ocean and earth system models are starting to be used for reanalysis of past climate states using essentially the same codes that have been developed for operational forecasting, especially of the atmosphere i.e. sequential "filter" codes. Even when 4D Variational approaches are being used, eg. at ECMWF, the effective temporal smoothing window timescales are generally short reflecting ~~atmospheric timescales~~ the validity of adjoint modelling for the atmosphere. In these cases Kalman smoothing approaches
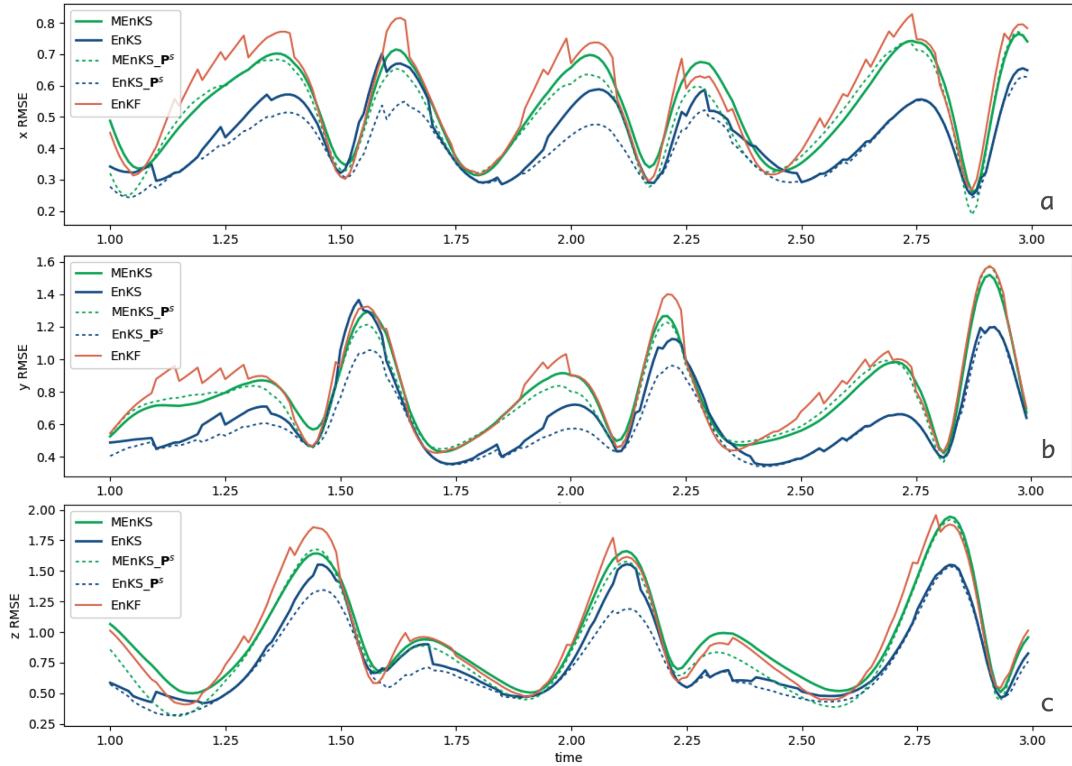
Figure 4. As in Fig. 1, but for EnKF, EnKS and MEnKS.

could still yield tangible benefits especially for long timescale process variables associated with the earth system, and when reanalysing using sparse historical observing systems.

However there are still further challenges to applying ~~smoothing~~ smoother algorithms in real large systems. In Dong et al.
320   (2021), the simple smoother was applied to an ocean reanalysis and it was found that the smoothed analysis gave reduced errors compared to the filter, against independent, unassimilated, data. However it was also noted that problems can occur when observations or model are biased. Biased increments can be detected when the same increment gets repeatedly ~~assimilated~~ produced by the filter, which is a signal that the model is unable to retain the information. While this may not invalidate the filter analysis it could have a very detrimental impact on smoothing when multiple versions of the same increment may be
325   added without the model being re-run. While bias can be allowed for if it is identified prior to smoothing, any real application of smoothing needs to consider this carefully. This is perhaps another reason to prefer smoothing as a post processing step when bias assessments can be made beforehand, rather than as an integrated part of a sequential forward analysis as it is usually presented in the literature (Todling and Cohn, 1996; Evensen and van Leeuwen, 2000; Bocquet and Sakov, 2014).

Another option not explored here~~because~~, ~~because~~ L63 is too simple, is the ability to tune the γ decay timescale for different
330   state variables. In Dong et al. (2021), this was suggested as allowing subsurface ocean increments to decay more slowly than
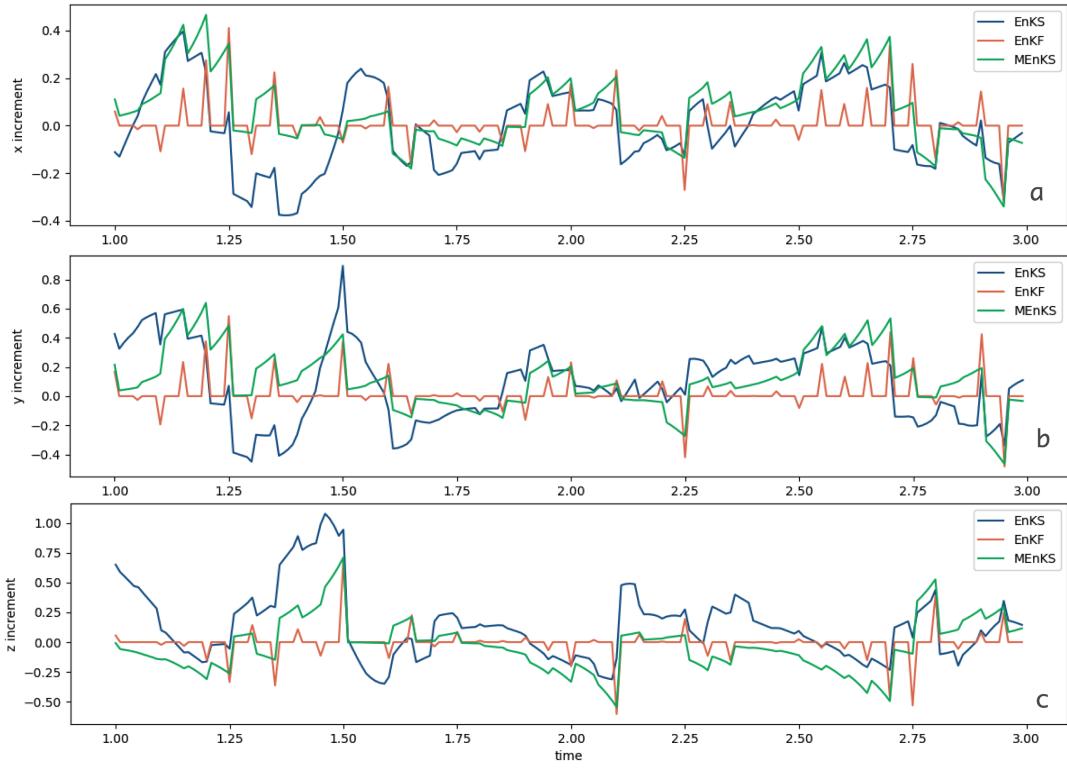
Figure 5. As in Fig. 2 but for EnKF, EnKS and MEnKS.

surface increments, for example. In the notation used here γ would then become a diagonal decay matrix multiplying the forecast error covariance to convert to a cross time error covariance.

A key benefit to smoothing in real systems would be to bring influence from observations made in the near future where none have been available in the near past, ~~eg~~ for instance, after deployment of new observing platforms. A key difference between our approach and using a full smoother is that in a full smoother the cross time error covariances depend upon observations

335 previously assimilated within the smoothing lag time window (see Appendix). Thus a full smoother will reduce analysed error covariances due to the influence of short lag future data first, and ~~in doing so~~ therefore will reduce the cross time error covariances to be applied for longer lag future data, ensuring that the most important near future data has the biggest smoothing influence. This ~~shielding~~ reduction of longer lag influences if shorter lag data are available is missing in the simple smoother

340 as presented, and could cause the application of the smoother to give ~~poor~~ poorer results when very frequent observations are available. Further simple modifications to take this into account ~~might~~ could be envisioned, for example, allowing γ to reflect upgrades in the observing network during the period of the reanalysis. Alternatively, uncertainty reduction information for each future increment, as estimated through Eq. 19, could be used to truncate, or reduce γ, for the smoothing of longer lag increments through Eq. 15.

345   Although we have proposed how these ideas could be used in ensemble systems we have not explored other challenges of using ensembles in large model products. In particular localisation is often required to remove unrealistic error covariances arising from limited ensemble sizes (eg. Petrie and Dance (2010)~~, and when~~). When extended to ensemble smoothing that localisation may need to vary with lag for the cross time error covariances (eg. Desroziers et al. (2016)). Faced with such challenges the simple smoothing method is at least explicit in its assumption that the spatial structure of the error covariances

350   are static while guaranteeing that cross time error covariances will always decay away with time.

    We have included the smoothing of uncertainty estimates in the analysis here despite the fact that these have rarely been attempted for previous large model reanalysis products even when only forward filter steps are involved. However, with the recent trend towards ensemble analysis products, for both operational ~~work and for~~ and reanalysis systems, it ~~seems sensible~~ makes sense to ask how well uncertainty estimates do correspond to the errors in an idealised system where this can be evaluated

355   eg. against independent data. At the same time ~~we have therefore~~, we have demonstrated the ability to ~~also~~ evaluate smoother uncertainty estimates, and we have found these ~~first~~ results very encouraging.

6   Conclusions

We have demonstrated that both the extended Kalman smoother and the ensemble Kalman smoother can be simplified to use only a relatively small amount of information stored during a forward filter analysis. This permits the simple smoothing

360   approach to be applied through post processing. The essential novelty is to treat cross time error covariance information as decaying exponentially on some tuneable timescale, rather than seeking to ~~model it directly~~ calculate these covariances with the system model. This allows stored state increments to be down-weighted and added to previous filter analyses. We also show how the smoother uncertainty information can be post processed, provided the increments (changes) in the error covariances between the forecast and analysis for each filter assimilation window are ~~also~~ stored. And we note that the error variance of

365   state fields alone could be smoothed, meaning that only 1 additional state field needs to be stored from each filter analysis.

    The method has been demonstrated using assimilation runs of the ~~Lorentz~~ Lorenz 1963 model, using the same idealised assimilated data over a 20 time unit truth run, when starting the model from different initial conditions. Observational, but no model errors, are being simulated. In both the extended and ensemble ~~smoother cases~~ Kalman smoother cases, using the full smoother ~~approaches~~ equations give the best RMSE results against the truth. However in each case the simple smoother method

370   still gives substantially ~~improved RMSE results over the filter~~ reduced RMSE values compared with the respective Kalman filters, typically giving around ~~50-60~~ 60-70% of the ~~improvements obtained from the full smoothing~~ error reductions obtained using the full smoother equations. We also include the RMSE evaluated only at filter analysis times, when the truth comparison data ~~is not independent~~, are not independent (observations from these times, albeit with added errors, have been assimilated) and still find that the smoother results still provide substantial improvements over the filter. We include these comparisons

375   because operational systems do not usually hold back independent data for assessment. The ensemble filter/smoother results are substantially better than the extended filter/smoother results, as would be expected for such a nonlinear system as L63. The

simple smoother retains this benefit as the flow dependent ensemble filter error covariances are ~~used in filter analyses and it is these that are~~ retained, although down-weighted~~again~~, for the smoother's cross-time errors covariances.

We also demonstrate the smoothing of the uncertainty estimates in both the ExtKS and EnKS systems. Remarkably the uncertainty estimates, presented as the STD of the smoother state variances, are in very good agreement with the RMSE errors ~~actually~~ being calculated against the truth. The uncertainties rise and fall over time similarly to the RMSEs as the model moves through more stable and unstable regions of phase space. Uncertainty estimates are usually a little lower than the calculated RMSE values. The simple smoothing approach gives higher uncertainties than the full smoother estimates but is in excellent agreement with the simple smoother RMSE values.

We believe this approach offers a feasible offline post processing approach for improving reanalyses in real large Earth system models. An initial paper with first results on smoothing the Met Office ocean reanalysis using stored increments was presented in Dong et al. (2021). This paper demonstrates the traceable origin of the approach from Kalman filtering roots and puts the methods in a wider context, including showing how it can be used in ensemble systems that are just starting to be used operationally in order to get better representations of uncertainty.

To summarise the post processing requirements that would allow smoothing of large model datasets;

1. If increments from the sequential filter analysis are stored this should be sufficient to allow post processing of a smoother solution.

2. If an ensemble product is being generated only the ensemble mean fields and ensemble mean increments would be needed to obtain a smoothed ensemble mean solution.

3. If an uncertainty estimate is also needed for the smoother solution the minimum additional requirement would be to store the increments ~~(change from filter forecast to analysis)~~ of those components of the error covariance matrix of interest. This may consist of the error variances of ~~all state fields~~ each state field, or only a subset of state fields, eg. only surface fields from an ocean model.

4. If uncertainty information from an ensemble product is required the minimum additional storage requirement would still only be the filter increments in the error covariance components of interest. The whole past ensemble analyses would not be needed.

Appendix A: Formal derivation of the simple smoother system from the Kalman filter and Kalman smoother equations

In order to show formally how our simple smoother system in Sect. 2.1 is related to the classical Kalman smoother, we start with a brief summary of the Kalman filter (KF) and fixed-lag Kalman smoother (KS) formulations. We base our system of equations on Todling and Cohn (1996) (hereafter TC96), which we use as our reference for the classical smoother and we therefore adopt a notation similar to theirs. This is a more complex notation ~~from~~ than that of the main part of this paper, but is ~~necessary to complete our proof~~ useful for complete traceability.

A1 Background to the Kalman filter and smoother

The analysis of the KF at time k, and its error covariance are given by

$$\mathbf{x}^{\mathrm{a}}_{k|k} = \mathbf{x}^{\mathrm{f}}_{k|k-1} + \mathbf{K}_{k|k}\left(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}^{\mathrm{f}}_{k|k-1})\right), \tag{A1}$$

$$\mathbf{P}^{\mathrm{a}}_{k|k} = \mathbf{P}^{\mathrm{f}}_{k|k-1} - \mathbf{K}_{k|k}\mathbf{H}_k\mathbf{P}^{\mathrm{f}}_{k|k-1}, \tag{A2}$$

$$\text{where } \mathbf{K}_{k|k} = \mathbf{P}^{\mathrm{f}}_{k|k-1}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\mathbf{P}^{\mathrm{f}}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)^{-1}. \tag{A3}$$

Here the subscript $k|k-1$ indicates that the object is valid at timestep k, and has been formed from ~~information~~ observations up to and including timestep $k-1$. States $\mathbf{x}^{\mathrm{f}}_{k|k-1}$ and $\mathbf{x}^{\mathrm{a}}_{k|k}$ are the forecast state and filter analysis respectively at validity time k, where $\mathbf{x}^{\mathrm{f}}_{k|k-1}$ has been evolved by the model $\mathcal{M}$, eg. $\mathbf{x}^{\mathrm{f}}_{k|k-1} = \mathcal{M}(\mathbf{x}^{\mathrm{a}}_{k-1|k-1})$. The forecast state error covariance $\mathbf{P}^{\mathrm{f}}_{k|k-1}$ may be evolved by the model (as in the extended KF) or obtained from an ensemble of model state forecasts (ensemble KF), but either way the analysis error covariance Eq. (A2) for $\mathbf{P}^{\mathrm{a}}_{k|k}$ is obtained. The vector $\mathbf{y}_k$ represents the observations at k, whose model counterparts are found using the observation operator $\mathcal{H}_k$ via $\mathcal{H}_k(\mathbf{x}^{\mathrm{f}}_{k|k-1})$, and $\mathbf{H}_k$ is the tangent linear operator of $\mathcal{H}_k$. $\mathbf{R}_k$ is the observation error covariance matrix, and $\mathbf{K}_{k|k}$ is the Kalman gain. Equations (A1), (A2), and (A3) are the same equations as (6), (8), and (7) respectively in the main paper, but using the TC96 notation.

For the classical fixed-lag KS of maximum lag L, an interval of $L+1$ timesteps are updated together after every filter timestep. These $L+1$ states are valid for timesteps $k, k-1, \ldots, k-L$, and are to be updated by observations at timestep k. This is the *backward-looking* scheme of TC96 which runs interleaved with the filter (below we use j to represent backward-looking intervals). Prior to this update – and using a similar notation to TC96 – these states are $\mathbf{x}^{\mathrm{f}}_{k|k-1}, \mathbf{x}^{\mathrm{a}}_{k-1|k-1}, \ldots, \mathbf{x}^{\mathrm{a}}_{k-L|k-1}$. These are shown as the black states in Fig. A1(a). At this point, observations only up to $k-1$ have been assimilated, which is reflected in the notation, and so when $2 \le j \le L$, superscripts 'a' refer to partially smoothed analyses generated only using observations up to time $k-1$. The state $\mathbf{x}^{\mathrm{a}}_{k-1|k-1}$ is the pure $k-1$ filter analysis and $\mathbf{x}^{\mathrm{f}}_{k|k-1}$ is the filter forecast for k derived from it. The covariances of ~~and between~~, and between, these states are the black block matrices in Fig. A1(b), which are used to form the gain matrices in the current update (below).

The fixed-lag KS ~~states~~ determines how observations at k update the states at the ~~above~~ previous time levels to give $\mathbf{x}^{\mathrm{a}}_{k|k}$, $\mathbf{x}^{\mathrm{a}}_{k-1|k}, \ldots, \mathbf{x}^{\mathrm{a}}_{k-L|k}$ (the red states in Fig. A1(a)), and their covariances (the red matrices in (b)). The first state, $\mathbf{x}^{\mathrm{a}}_{k|k}$, and its covariance, $\mathbf{P}^{\mathrm{a}}_{k|k}$, are updated using the KF equations, but the remaining states, $\mathbf{x}^{\mathrm{a}}_{k-j|k}$, and their covariances, $\mathbf{P}^{\mathrm{a}}_{k-j|k}$, $(1 \le j \le L)$ are updated by the KS equations:

$$\mathbf{x}^{\mathrm{a}}_{k-j|k} = \mathbf{x}^{\mathrm{a}}_{k-j|k-1} + \mathbf{K}_{k-j|k}\left(\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}^{\mathrm{f}}_{k|k-1})\right), \tag{A4}$$

$$\mathbf{P}^{\mathrm{a}}_{k-j|k} = \mathbf{P}^{\mathrm{a}}_{k-j|k-1} - \mathbf{K}_{k-j|k}\mathbf{H}_k\mathbf{P}^{\mathrm{fa}}_{k,k-j|k-1}, \tag{A5}$$

$$\text{where } \mathbf{K}_{k-j|k} = \mathbf{P}^{\mathrm{af}}_{k-j,k|k-1}\mathbf{H}_k^{\mathrm{T}}\left(\mathbf{H}_k\mathbf{P}^{\mathrm{f}}_{k|k-1}\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k\right)^{-1}. \tag{A6}$$

The new objects are: $\mathbf{x}^{\mathrm{a}}_{k-j|k}$ ~~is~~ the updated smoothed state at $k-j$ due to observations at k, and $\mathbf{P}^{\mathrm{a}}_{k-j|k}$ ~~is~~ the corresponding updated covariance. Both objects are obtained using $\mathbf{K}_{k-j|k}$, which is the gain for the smoother state at $k-j$ due to observations at k.

(a) Update and evolution of the smoother states.

(b) Update and evolution of the smoother error covariances.
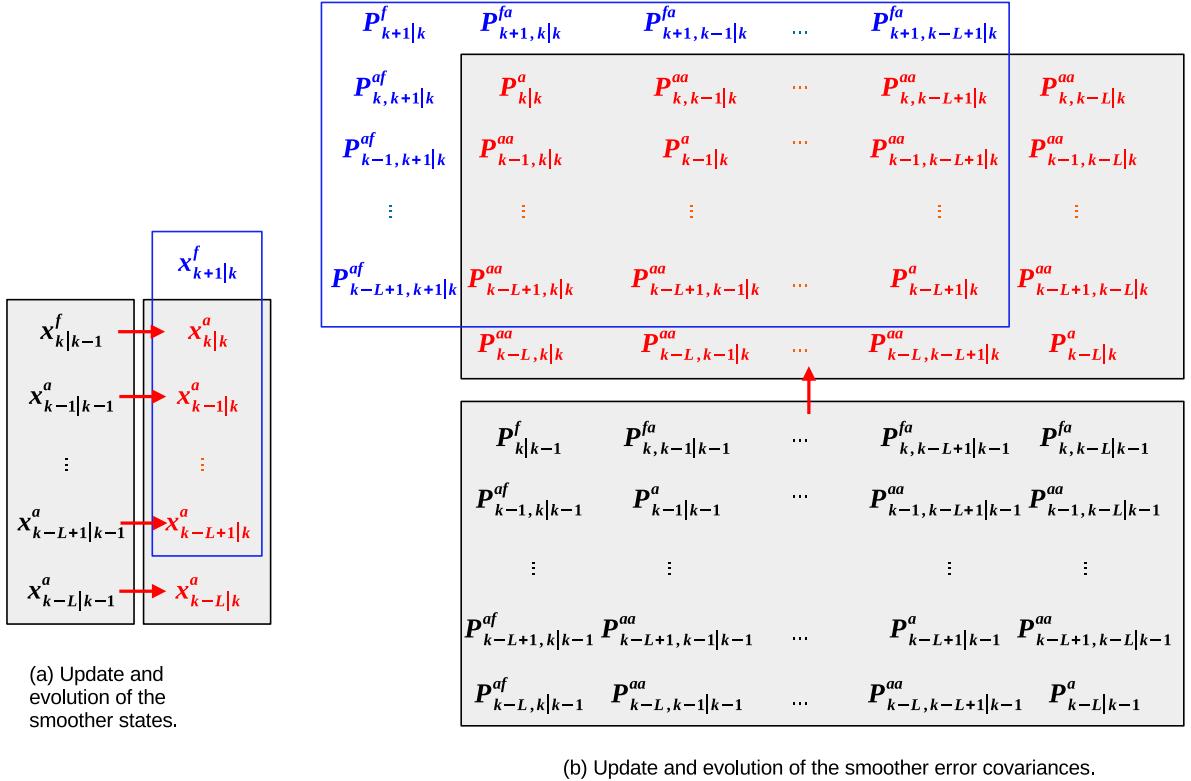
Figure A1. Schema of the fixed lag KS of TC96. Panel (a) shows the update and evolution of the set of states within the fixed lag interval of $L+1$ timesteps. The smoother update starts with the set of states in black, which are updated (or smoothed) by observations at k to the set of states in red using Eq. (A1) for the most recent state (at the top), and Eq. (A4) for the remainder. The subscripts have the form $k|p$, where k is the state's validity time, and p is the timestep of the latest observations that have contributed to estimating that state. The Kalman gains (Eqs. (A3) and (A6)) rely on knowledge of the covariances in panel (b) (first column of the black matrix). The lag interval then progresses by one timestep (blue box of panel (a)), where the forecast (blue state) is evolved from the latest analysis using the model. Panel (b) shows the update and evolution of the error covariances within the fixed lag interval. The black matrix blocks are the error covariances of the set of black states in (a). The diagonal blocks have a subscript of the same form as the states, but the off-diagonal blocks have subscripts of the form $k, k'|p$, which correspond to cross covariances between timesteps k and k'. The smoother updates the black covariances to the red covariances using equations not fully shown in this paper (see Eqs. (37), (39), and (41) of TC96, which use information from the black matrices). The lag interval then progresses, as it does for the states, by one timestep (blue box), where the extra covariances (blue) are propagated from the red covariances using Eqs. (46) and (47) of TC96 and the tangent linear model.

To make these updates requires a new kind of covariance for errors between different times. These have a subscript of the form $k, k'|k-1$, which indicates that the covariance is between timesteps k and k', and has been formed from information up to and including timestep $k-1$. In the above $\mathbf{P}^{\mathrm{af}}_{k-j,k|k-1} = \left(\mathbf{P}^{\mathrm{fa}}_{k,k-j|k-1}\right)^{\mathrm{T}}$ are the covariances between errors in $\mathbf{x}^{\mathrm{a}}_{k-j|k-1}$ and $\mathbf{x}^{\mathrm{f}}_{k|k-1}$. Each $\mathbf{P}^{\mathrm{af}}_{k-j,k|k-1}$ is obtained from $\mathbf{P}^{\mathrm{aa}}_{k-j,k-1|k-1}$ in Fig. A1(b) as separate covariance propagations (or from an ensemble of forecasts in the EnKS) for each $k-j$. If the EnKS is being used, these error covariances are not derived directly, but require all the partially smoothed ensemble members within the lag L to be retained. These are the same covariances as those expressed in Eq. (27) in the ~~trucated main text~~ main text truncated notation.

Equation (A4) is a version of TC96 Eq (26)~~given in TC96~~, (A5) is their (39), and (A6) is their (35). Notice that the same error covariance matrix, $\mathbf{P}^{\mathrm{f}}_{k|k-1}$, appears in the brackets in Eqs. (A3) and (A6). The incremental part of Eq. (A5) is the same as (16) in the main paper, Eq. (A6) is the same as (11), but here using the TC96 notation. There is no equivalent of (A4) given in the main paper.

The KS system is advanced one timestep by propagating $\mathbf{x}^{\mathrm{a}}_{k|k}$ using the model, to $\mathbf{x}^{\mathrm{f}}_{k+1|k}$ (the blue state in Fig. A1(a)), giving a shifted interval of states (blue box in (a)). The covariances are propagated by the tangent linear model (the blue block matrices in Fig. A1(b)), or by propagating the ensemble of new analyses, giving a shifted interval of covariances (blue box in (b)).

A2 ~~Explicit equations~~ Equations for maximally smoothed states and covariances, and equivalence to the simple smoother

Given the maximum lag, L, the sequence of states $\mathbf{x}^{\mathrm{a}}_{0|L}, \mathbf{x}^{\mathrm{a}}_{1|L+1}, \mathbf{x}^{\mathrm{a}}_{2|L+2}, \ldots$ (and their error covariances $\mathbf{P}^{\mathrm{a}}_{0|L}, \mathbf{P}^{\mathrm{a}}_{1|L+1}, \mathbf{P}^{\mathrm{a}}_{2|L+2}, \ldots$) exploit the maximum amount of observational information as they have been updated with all $L+1$ sets of ~~future and present~~ present and future observations. These are the states that are analogous to $S_0, S_1, S_2 \ldots$ in Sect. 2.1. ~~Other states in this appendix, such as $\mathbf{x}^{\mathrm{a}}_{1|L}$, are only partially smoothed, but are still needed as part of the classical algorithm.~~ For a general k, the fully smoothed states here are $\mathbf{x}^{\mathrm{a}}_{k|k+L}$, which can be found from cyclic application of the KS equations. By recursively applying Eq. (A4) over the lag window, it is straightforward to find the following explicit full smoothing solution (~~now with~~ where superscript s ~~as~~ is used in main text~~although TC96 and Table A1 retain a~~) in a *forward-looking* perspective (with ℓ to represent forward-looking intervals):

$$\mathbf{x}_{k|k+L_-}{}^{\mathrm{sa}} = \mathbf{x}^{\mathrm{a}}_{k|k} + \sum_{\ell=1}^{L} \mathbf{K}_{k|k+\ell}\left(\mathbf{y}_{k+\ell} - \mathcal{H}_{k+\ell}(\mathbf{x}^{\mathrm{f}}_{k+\ell|k+\ell-1})\right) \tag{A7}$$

(c.f. Eq. (10) in the main paper). Similarly, recursively applying (A5) over the lag window leads to the following explicit covariance of this ~~smoothing estimate:~~ fully smoothed estimate:

$$\mathbf{P}_{k|k+L_-}{}^{\mathrm{sa}} = \mathbf{P}^{\mathrm{a}}_{k|k} - \sum_{\ell=1}^{L} \mathbf{K}_{k|k+\ell}\mathbf{H}_{k+\ell}\mathbf{P}^{\mathrm{fa}}_{k+\ell,k|k+\ell-1} \tag{A8}$$

(c.f. Eq. (17)). We will now use Eqs. (A7) and (A8) to show the necessary approximations needed to give the simplified smoothing equations shown in Sects. 2.1 and 2.2.

The fundamental approximation ~~used in this work is to approximate the cross covariances mentioned above~~ is applied to the cross time error covariances (as they appear in Eq. (A6))~~with~~, using a temporal covariance decay ~~by writing~~and writing;

$$\mathbf{P}^{\mathrm{af}}_{k-j,k|k-1} \approx \gamma^j \mathbf{P}^{\mathrm{f}}_{k|k-1} \tag{A9}$$

(c.f. Eq. (12)). This is equivalent to writing $\mathbf{K}_{k-j|k}$ using Eq. (A6) and (A3), as

$$\mathbf{K}_{k-j|k} \quad \approx \quad \gamma^j \mathbf{P}^{\mathrm{f}}_{k|k-1}\mathbf{H}^{\mathrm{T}}_k \left(\mathbf{H}_k \mathbf{P}^{\mathrm{f}}_{k|k-1}\mathbf{H}^{\mathrm{T}}_k + \mathbf{R}_k\right)^{-1} = \gamma^j \mathbf{K}_{k|k},$$

which, by re-indexing, is $\mathbf{K}_{k|k+\ell} \quad \approx \quad \gamma^\ell \mathbf{K}_{k+\ell|k+\ell}$ (A10)

(c.f. Eq. (13)). Equation (A10) is in a forward-looking form, allowing it to be used in Eq. (A7):

$$\mathbf{x}_{k|k+L_-}^{\mathrm{sa}} \quad \approx \quad \mathbf{x}^{\mathrm{a}}_{k|k} + \sum_{\ell=1}^{L} \gamma^\ell \mathbf{K}_{k+\ell|k+\ell}\left(\mathbf{y}_{k+\ell} - \mathcal{H}_{k+\ell}(\mathbf{x}^{\mathrm{f}}_{k+\ell|k+\ell-1})\right),$$

$$= \quad \mathbf{x}^{\mathrm{a}}_{k|k} + \sum_{\ell=1}^{L} \gamma^\ell \left(\mathbf{x}^{\mathrm{a}}_{k+\ell|k+\ell} - \mathbf{x}^{\mathrm{f}}_{k+\ell|k+\ell-1}\right), \tag{A11}$$

where the last line follows from the first using the filter, Eq. (A1), allowing us to relate the simplified smoother updates to the later filter increments. Equation (A11) is analogous to the simplified scheme of Eq. (1), where $\mathbf{x}^{\mathrm{s}}_{k|k+L}$ ~~$\mathbf{x}^{\mathrm{a}}_{k|k+L}$~~ is the post hoc smoothing analysis (S_0 in Eq. (1)), $\mathbf{x}^{\mathrm{a}}_{k|k}$ is the previous filtering analysis (A_0), and the terms in brackets form the filter analysis increments at future times (I_j). Compare also to Eq. (15).

It is also possible to make the equivalent approximations to the smoothed covariances, Eq. (A8). Using Eqs. (A9) and (A10) gives

$$\mathbf{P}_{k|k+L_-}^{\mathrm{sa}} \approx \mathbf{P}^{\mathrm{a}}_{k|k} - \sum_{\ell=1}^{L} \gamma^{2\ell} \mathbf{K}_{k+\ell|k+\ell}\mathbf{H}_{k+\ell}\mathbf{P}^{\mathrm{f}}_{k+\ell|k+\ell-1}.$$

From (A2), $\mathbf{K}_{k+\ell|k+\ell}\mathbf{H}_{k+\ell}\mathbf{P}^{\mathrm{f}}_{k+\ell|k+\ell-1}$ in the above is equal to the difference between the forecast and filter analysis error covariance matrices, making the above

$$\mathbf{P}_{k|k+L_-}^{\mathrm{sa}} \approx \mathbf{P}^{\mathrm{a}}_{k|k} - \sum_{\ell=1}^{L} \gamma^{2\ell} \left(\mathbf{P}^{\mathrm{f}}_{k+\ell|k+\ell-1} - \mathbf{P}^{\mathrm{a}}_{k+\ell|k+\ell}\right). \tag{A12}$$

This is the same as Eq. (19) in the main body of the paper, but ~~in~~ using the TC96 notation.

Code and data availability. Implementations of the L63 system for the Ext KF/KS codes and the Ensemble KF/KS codes are available on Zenodo doi:10.5281/zenodo.7675286. The implementation here is based on the Python-based data assimilation templates DAPPER: https://github.com/nansencenter/DAPPER

Author contributions. BD and KH initiated the research idea and designed the experiments. All authors contributed to the theoretical developments presented. BD and YC implemented the codes and developed the diagnostic results. All authors contributed to the writing of the manuscript.

Competing interests. The authors are not aware of any competing interests.

Acknowledgements. All authors acknowledge support from the NERC NCEO LTSS strategic programme in DA for their contributions to
500 this work.

References

Balmaseda, M. A., Mogensen, K., and Weaver, A. T.: Evaluation of the ECMWF ocean reanalysis system ORAS4, Quarterly journal of the royal meteorological society, 139, 1132–1161, 2013.

Bishop, C. H., Etherton, B. J., and Majumdar, S. J.: Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Monthly Weather Review, 129, 420 – 436, https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2, 2001.

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Quarterly Journal of the Royal Meteorological Society, 140, 1521–1535, https://doi.org/https://doi.org/10.1002/qj.2236, 2014.

Burgers, G., Van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, Monthly weather review, 126, 1719–1724, 1998.

Cosme, E., Brankart, J.-M., Verron, J., Brasseur, P., and Krysta, M.: Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation, Ocean Modelling, 33, 87–100, https://doi.org/https://doi.org/10.1016/j.ocemod.2009.12.004, 2010.

Desroziers, G., Arbogast, E., and Berre, L.: Improving spatial localization in 4DEnVar, Quarterly Journal of the Royal Meteorological Society, 142, 3171–3185, 2016.

Dong, B., Haines, K., and Martin, M.: Improved high resolution ocean reanalyses using a simple smoother algorithm, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 626, 2021.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans, 99, 10 143–10 162, https://doi.org/https://doi.org/10.1029/94JC00572, 1994.

Evensen, G.: The ensemble Kalman filter for combined state and parameter estimation, IEEE Control Systems Magazine, 29, 83–104, https://doi.org/10.1109/MCS.2009.932223, 2009.

Evensen, G. and van Leeuwen, P. J.: An Ensemble Kalman Smoother for Nonlinear Dynamics, Monthly Weather Review, 128, 1852 – 1867, https://doi.org/10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2, 2000.

Evensen, G., Vossepoel, F. C., and van Leeuwen, P. J.: Data assimilation fundamentals: A unified formulation of the state and parameter estimation problem, Springer Nature, 2022.

Feng, L., Palmer, P., Bösch, H., and Dance, S.: Estimating surface CO_2 fluxes from space-borne CO_2 dry air mole fraction observations using an ensemble Kalman Filter, Atmospheric chemistry and physics, 9, 2619–2633, 2009.

Hamill, T. M., Whitaker, J. S., Fiorino, M., and Benjamin, S. G.: Global ensemble predictions of 2009's tropical cyclones initialized with an ensemble Kalman filter, Monthly Weather Review, 139, 668–688, 2011.

Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Monthly Weather Review, 126, 796–811, 1998.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B.: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations, Monthly Weather Review, 133, 604 – 620, https://doi.org/10.1175/MWR-2864.1, 2005.

Khare, S. P., Anderson, J. L., Hoar, T. J., and Nychka, D.: An investigation into the application of an ensemble Kalman smoother to high-dimensional geophysical systems, Tellus A: Dynamic Meteorology and Oceanography, 60, 97–112, https://doi.org/10.1111/j.1600-0870.2007.00281.x, 2008.

Lorenz, E. N.: Deterministic nonperiodic flow, Journal of the Atmospheric Sciences, 20, 130–141, 1963.

MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A., Gordon, M., Vellinga, M., Williams, A., Comer, R., et al.: Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system, Quarterly Journal of the Royal Meteorological Society, 141, 1072–1084, 2015.

540 Pan, Y., Zhu, K., Xue, M., Wang, X., Hu, M., Benjamin, S. G., Weygandt, S. S., and Whitaker, J. S.: A GSI-Based Coupled EnSRF–En3DVar Hybrid Data Assimilation System for the Operational Rapid Refresh Model: Tests at a Reduced Resolution, Monthly Weather Review, 142, 3756 – 3780, https://doi.org/10.1175/MWR-D-13-00242.1, 2014.

Petrie, R. E. and Dance, S. L.: Ensemble-based data assimilation and the localisation problem, Weather, 65, 65–69, 2010.

Pinnington, E., Quaife, T., Lawless, A., Williams, K., Arkebauer, T., and Scoby, D.: The land variational ensemble data assimilation frame-
545 work: LAVENDAR v1. 0.0, Geoscientific Model Development, 13, 55–69, 2020.

Ravela, S. and McLaughlin, D.: Fast ensemble smoothing, Ocean Dynamics, 57, 123–134, https://doi.org/10.1007/s10236-006-0098-6, 2007.

Sakov, P., Counillon, F., Bertino, L., Lisæter, K., Oke, P., and Korablev, A.: TOPAZ4: an ocean-sea ice data assimilation system for the North Atlantic and Arctic, Ocean Science, 8, 633–656, 2012.

Skachko, S., Ménard, R., Errera, Q., Christophe, Y., and Chabrillat, S.: EnKF and 4D-Var data assimilation with chemical transport model
550 BASCOE (version 05.06), Geoscientific Model Development, 9, 2893–2908, 2016.

Todling, R. and Cohn, S.: Some strategies for Kalman filtering and smoothing, 1996.

Uppala, S. M., Kållberg, P., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al.: The ERA-40 re-analysis, Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography, 131, 2961–3012, 2005.

555 van Velzen, N., Altaf, M. U., and Verlaan, M.: OpenDA-NEMO framework for ocean data assimilation, Ocean Dynamics, 66, 691–702, 2016.

Wang, X., Parrish, D., Kleist, D., and Whitaker, J.: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments, Monthly Weather Review, 141, 4098–4117, 2013.

Whitaker, J. S. and Hamill, T. M.: Ensemble data assimilation without perturbed observations, Monthly weather review, 130, 1913–1924,
560 2002.

Zeng, Y. and Janjić, T.: Study of conservation laws with the local ensemble transform Kalman filter, Quarterly Journal of the Royal Meteorological Society, 142, 2359–2372, 2016.

Zhang, M. and Zhang, F.: E4DVar: Coupling an ensemble Kalman filter with four-dimensional variational data assimilation in a limited-area weather prediction model, Monthly Weather Review, 140, 587–600, 2012.

565 Zhu, Y., Todling, R., Guo, J., Cohn, S. E., Navon, I. M., and Yang, Y.: The GEOS-3 Retrospective Data Assimilation System: The 6-Hour Lag Case, Monthly Weather Review, 131, 2129 – 2150, https://doi.org/10.1175/1520-0493(2003)131<2129:TGRDAS>2.0.CO;2, 2003.