1 Review 1

I like this paper. The authors have done a great job at giving a walk through on exactly what is going on in a way that is easy to understand. It goes through each method nicely with no real assumptions made that might make the reader confused.

Major Comments:

None **Thank you for your kind comments and interest in this work.**

Minor Comments:

Line 19 - A comma after ”However” is probably needed here. **Done.**

Line 59 - A comma after ”In addition” is probably right here **Done.**

Line 59:60 - ”that has been developed over thousands of person years” sounds really odd to me. **Changed**

Section 3.1:

a) What is the timestep size? I am guessing it’s 0.01 (?), but I think it should be noted. **Now stated explicitly (it was previously implicit).**

b) Why did you not observe z? I am curious about why this choice was made instead of observing all variables.

The true answer here is that we had some legacy code for L63 set this way. But in fact it reveals some interesting smoothing results because z errors are not strongly correlated with x,y errors climatologically. The full KS manages to deal with this and improves z as it follows how correlations change with time which the simplified smoother does not, thus revealing an aspect of the algorithms which may not have been apparent otherwise.

c) Why choose x=5 and y=20 to be your observation frequencies?

This frequency is somewhere near the divergence limit with these error settings etc, as can be seen the RMSEs in Figs 2,4 do rise and fall significantly. It was important to have a system which does not completely converge under the filter leaving room for the smoother to significantly improve results.

d) Why retain 5%? I know you said it’s to avoid divergence, do you have a citation for this or was it tested by you?

This was obtained by testing in the early stages. With essentially the same idealised system in Dong et al (2021) but using 3DVar we were using climatological error covariances throughout which was working well but the Kalman filter uses flow-dependent error covariances, and this sometimes leads to irreversible error collapse in the very nonlinear L63. The retention of a climatological component prevents this, ensuring new data always gets to be used.

Points b-c are just me curious about why the choices were made. I don't think more experiments are needed, but maybe add something to give context to why you decided on these parameters. Could be easy like referring to papers whose setup you are using, or just a small sentence here and there. **We have also added the L63 model set up reference at the beginning of the model description.**

Line 183 - I like L=40 is good here, how general do you think that is?

We found that the full KS RMSEs have mostly converged at L=40 with the timestep, observation frequencies and observation error choices. This will also be a very specific function of the L63 system related to the x,y oscillations and the lobe transition frequency, although we have not investigated in detail.

Line 198 - What do you mean by "true RMSE" here? I don't understand what these dotted blue/green uncertainty estimates are in agreement with. **Changed to RMSE against truth.**

Line 252 - Say what "i" refers to. **now defined -is the normalised anomaly of the *i*th ensemble member from the mean.**

Line 286 - "Fig. 2, reflecting the improved assimilation approaches." Is this supposed to be a stand alone sentence? It feels like something is missing to make it a complete sentence. **there was a . instead of a , now corrected.**

2 Review 2

"Simplified Kalman smoother and ensemble Kalman smoother for improving reanalyses" By Bo Dong, Ross Bannister, Yumeng Chen, Alison Fowler, and Keith Haines

Referee comment

The manuscript introduces a simplified smoother algorithm which can be used to post- processes the analysis state in the filter-based reanalysis datasets. Based on the authors results, this algorithm could fix the time-discontinue problems in the filter-based reanalysis, and further reduce their errors. The algorithm can be easily applied to existing filter-based datasets, and therefore are of interest to the ocean and earth system reanalysis community. However, there are still some problems and deficiencies that the authors should clarify and fix before the manuscript is accepted and published. I suggest the authors further improve the languages and grammar since I feel like some of the statements cannot be understood easily. I suggest a Minor revision.

Thank you for your careful reading and comments on this work. We have done our best to address these below.

General Comments: 1. Introduction: The readers likely prefer to see how much more computational, storage, memory requirements of KS and EnKF than KF and EnKF. This is a major reason for simplifying the original EnKS in this study. I see that there are already studies using KS (L50-L55), probably in small models. Please clarify how many times the original KS algorithms than the KF roughly in L55, so we can see that it is not possible to use KS or EnKS algorithms in large operational forecasting systems.

To run the full Kalman smoother requires the cross time error covariances to be calculated for each time lag as the main extra step. If L=40 as used here the ExtKS requires an extra 40 runs of the model compared to the ExtKF to carry the cross time covariances forward in time. For the EnKS this requires 40 additional matrix.matrix multiplications to fully smooth each filter analysis field, Asch et al (2016) DA, Methods, Algorithms and Applications.

2. Methods

1) For the purpose of introducing a new algorithm or method, in equation (1)-(5), the authors should clarify Why they use the coefficients in the forms of $\gamma^0, \gamma^1, \gamma^2, \gamma^3$? May I use other forms of decaying coefficients like in ln forms and Gaussian Forms? And how do I decide the parameter γ?

From my expectations, the errors of nonlinear systems grows exponentially, and their growth rate depends on singular values locally or Lyapunov exponent globally. Therefore, it is more nature to use a *ln* forms of the decaying coefficients. In this sense, based on the predictability of the nonlinear systems and their variables, we can use equation (5) to roughly decide the γ for different state variables.

Thank you for pointing out this aspect. Based on Eq (5), $\gamma = e^{\frac{-\delta t}{\tau}}$, which implicitly satisfies a ln form. At the ℓ-th time step, the polynomial form of γ^{ℓ} is simply a multiplication of δt. This implies that, in a global sense, the (leading) Lyapunov exponent $\lambda = \frac{1}{\tau}$. Certainly, the choice of γ is a simplified approach to parameterise the error growth of the model. A more explicit use of the Lyapunov exponent, a Gaussian form, or a time-dependent use of local Lyapunov exponent is possible and are potential directions for further exploration, which are out of the scope of this work.

2) In L100, please clarify: do I need to run the assimilation system, especially the forward model again? Or the algorithm (1)-(4) just do an weighted average on the analysis fields and their errors without running the forward model again (e.g. give S0-S1 back to the forward model and run the model again to produce the model state between the analysis time t-t+1)?

The model does not need to be run again. Most operational systems only save state fields once per analysis time and the Dong et al 2021 system was aimed at smoothing products at the same frequency. However if forward modelling products are available between analysis times these can also be smoothed using fractional increments, along with a redefined γ per timestep as mentioned later in the paragraph. See also responses around L130 below.

Specific comments:

L35: please clarify the word “analysis”. I think these studies use data assimilation to produce the initial condition for predictions. But there are systems that produce reanalysis datasets now, for instance, the TOPAZ system by Sakov et al, 2012. Therefore, please clarify that they are used to optimize the initial conditions. **For clarification, the word ”analysis” is the terminology to mean the updated state from a data assimilation procedure. Whether the analysis is used for initial conditions for a forecast, or as a reanalysis product, the assimilation procedure is largely the same, so it does not seem to be necessary to distinguish the application here.**

L 35, “For example ......”, incomplete sentence, please consider rewriting the sentence. **We added a : to the beginning of the sentence so that this is now continuation of last sentence.**

L105, What do you mean by saying “which is not given an explicit maximum cutoff"? "given as an explicit maximum cutoff"? or which is similar as an maximum cutoff? Please clarify. **This is now explained noting that Eqs 1,2 do not have fixed lag cutoff, unlike the fixed-lag smoothers.**

L 110,"in which a tangent linear model is used when the model is nonlinear", this statement is confusing. I think tangent linear model is only used to propagate the model errors in the Kalman gain matrix, right? **Agreed, clarified in text as "a tangent linear model is used for error covariance propagation when the model is nonlinear".**

L115, "The nonlinear operators have to be replaced by their tangent linear approximations in the forecast and analysis". Do you mean that in $y - H(x)$, H should also be replaced by their tangent linear approximation? The model matrix M and its tangent linear form M are not explained explicitly. **We removed the sentence and added a sentence after the Kalman gain equation Eq(7) to explain the use of TLMs. "Here, instead of the non-linear observation operator, the tangent linear approximation, $\mathbf{H}_k \in \mathbb{R}^{m\times n}$, is used in the gain below, and the tangent linear model, $\mathbf{M}_k \in \mathbb{R}^{n\times n}$, is used for the propagation of the forecast error covariance matrix, $\mathbf{P}_k^f = \mathbf{M}_k \mathbf{P}_{k-1}^f \mathbf{M}_k^T$."**

L120 the meaning of T in equation (7) is not explained. **Transpose defined now as "transpose operator".**

L 130, "We note that index ℓ could be defined as future analysis timesteps rather than model timesteps if data are only introduced at regular analysis intervals" This statement is confusing. I believe that the analysis steps must be in model timesteps right? Please clarify that ℓ means analysis timesteps, I think this is enough.

The analysis timesteps are a subset of model timesteps. It is not necessary to explicitly calculate a smoother analysis for every timestep if that is not required. Any operational (filter) system will not usually store state fields for every timestep. Therefore only the stored filter states need to be smoothed, generally at the same frequency as filter assimilation increments are being analysed. We suggest the following text change, hoping this is clearer: "We note that if filter states are only stored at some assimilation frequency, eg. once per day, the index ℓ can be defined at the same frequency as these filter increments if smoother states are only required at the same frequency as the stored filter states."

L140, "cross time error covariance", do you have any cross variable error covariance in KS or ExtKS? I think in you simplified algorithms with time-lag, the parameter γ doesn't include cross-variable information. Please clarify in the comparison between KS and you simplified algorithm.

In original KS or ExtKS, there is cross variable error covariance. However, in our simplified smoother, the cross-variable error covariances are defined entirely through the filter covariance i.e. are contained in the P^f in Eq 12. It is awkward to refer to this explicitly here but we note this explicitly in response to the L150 point below.

L150, ", with the spatial covariances being determined by the KF equations, but the temporal covariances (from times $k+\ell$ to k) being approximated by a simple decay" How about the cross-variable covariances? **Cross-variable error covariances, along with spatial errors, are defined via the filter, which is now stated.**

L175 "error standard deviation of 2" Is that mean "add Gaussian errors with standard deviations of 2"? please clarify. **Agreed, now clarified.**

L180, "Dong et al. (2021) used 3DVar for assimilation into L63 and they used a fixed background error covariance from a climatological L63 run" This statement makes no sense. **We have changed the text to "Dong et al. (2021) used 3DVar for assimilation into L63, with a fixed background error covariance prescribed as the time mean error covariance in a separate L63 run". Now in the text "climatological" is no longer used.**

Figure 1,2,4, please consider include the dashed lines in the legends. Also, please clarify that KS/KF means the extended KS/KF in the legends. Please added what the x-axis means? Time units. **The legends and x-axes will be corrected now.**

L205 "RMSE time series in Fig. 1 where the red line declines sharply where data are available" Do you mean the extended KF results? I don't think it is a red line. **Yes the Kalman filter line is perhaps orange. Explicitly noted this is the ExtKF now.**

Figure 3, please clarify that the x axis means time steps rather than time units! **Figure updated to define time axis - it is actually in time units.**

L 235, "the TLM is not always reliable for a system as non-linear as the L63 model" Until now, no error models away work for the nonlinear system. But they work at limited time steps or conditions. Therefore, I suggest the authors to be more precise " the TLM can represent the model error propagation within limited time", maybe.**Thanks for the suggestion – this is now rephrased as "the TLM reliability declines sharply with propagation time."**

L 235 "This improves the quality of the forecast error covariance matrix". I don't agree with this statements. If you have any citations for this statement, please add it. For L63 model, I guess TLM should be more accurate than small ensembles (e.g. 10s) within short assimilation window (e.g., 0.4 time units).

Agreed. The text around L235 has been modified to say that ensemble methods can give improvements over methods using a TLM under the correct circumstances. It is not desired to go into quantifying details for the L63 system as that is not main focus of paper.

L 240 “While ensemble filter methods are starting to be adopted for larger environmental models, the cost to store, update and apply posterior ensemble covariances still makes ensemble smoother methods generally infeasible” What is the logic of this statement? The application of ensemble smoother is not related to ensemble filters. Please rewrite the statement. **Text has been clarified.**

L290 “the effective temporal smoothing window timescales are generally short reflecting atmospheric timescales.” What do you mean “reflecting atmospheric timescales”? I think the predictability of atmosphere system should be 2 weeks. The reasons for using a 24-hour reanalysis window is mostly related to the usefulness of the adjoint model, since the parameterizations processes the accuracy of the adjoint model significantly. Please added citations here. **Agreed, we now explicitly refer to validity of tangent linear and adjoint modelling of atmosphere.**

L 315 “This shielding of longer lag influences if shorter lag data are available is missing in the simple smoother as presented and could cause the application of the smoother to give poor results when very frequent observations are available.” Not sure what the authors mean, please rewrite and clarify.

shielding has been replaced by reduction. This is a key difference and is described through the full KS algorithm as described in the Appendix.

L320, “In particular localisation is often required to remove unrealistic error covariances arising from limited ensemble sizes eg. Petrie and Dance (2010), and when extended to ensemble smoothing that localization may need to vary with lag for the cross time error covariances eg. Desroziers et al. (2016).” The citation should be (eg. Petrie and Dance, 2010), (eg. Desroziers et al., 2016.), please confirm. **Agreed done.**

After reading the method, I feel like the parameters work the same way as a decay time localization. I think the more challenging things in ensemble smoother are localization spatio-temporal together. In your simplified algorithm, you use the localized analysis from the filters, and assume time-decay γ. So. I am not sure whether there are some challenges when applying to a larger model. As the authors mentioned, the more flexible and challenging things should be deciding different γ with different variables and potentially different locations and time.

There certainly will be more challenges in tuning in order to use “optimal” γ smoothing, which could vary spatially for example, on large model

output. Also state variables may be locked into dynamical relationships (eg. Dong et al discuss T, S and sea level smoothing). However, this paper's introduction of the γ time decay smoothing parameter, operated through post processing, opens up the possibility of testing smoothing methods more routinely in a range of dynamical systems that have been originally designed essentially as sequential filters.

L345 This paragraph is not clear, please consider rewrite it in a clear way. For instance, please clarify, you use extended and ensemble Kalman smoothers. What is the meaning of "improved RMSE results"? What is "from the full smoothing"? is that means smoothers? "when the truth comparison data is not independent," is that means the assimilated data? The statements can be expressed in a more neater ways.

We have rewritten to clarify the application to each of the smoother methods. "Improved RMSE results" means smaller RMSE wrt the truth. The updated text is: "In both the extended and ensemble Kalman smoother cases, using the full smoother equations give the best RMSE results against the truth. However in each case the simple smoother method still gives substantially reduced RMSE values compared with the respective Kalman filters, eg. $\sim$70% of the error reduction of the extended Kalman smoother. We also include the RMSE evaluated only at filter analysis times, when the truth comparison data are not independent (observations from these times, albeit with added errors, have been assimilated) and still find that the smoother results provide substantial improvements over the filter."

Technical corrections

L85, should be In Dong et al. (2021), a simple smoother m **Done.**

L 185 please consider deleting "The reasons for this appear later." **Deleted.**

L190 "Across the 100 member ensembles" **We have altered this to "assimilation runs" to avoid using "ensemble" when referring to the ExtKF/KS results to avoid confusion with EnKF/KS methods used later.**

L240 "are starting to be adopted for larger environmental model" have been adopted **Done.**

L245 (Bishop et al., 2001, ETKF) (ETKF, Bishop et al., 2001) **Done.**

L285 "20 time units of the Ensemble runs"– ensemble **Done.**

L295, "However there are still further challenges to applying smoothing in real large systems" smoother algorithms **Done.**

L300 "when the same increment gets repeatedly assimilated by the filter" is this means repeatedly produced by the filters?**"assimilated" altered to "produced".**

L305-L310, “Another option not explored here,” —“because L63 is too simple, Another option not explored here” **Done.**

L310” eg after deployment of new observing platforms.”—“for instance” “and in doing so” and therefore.**Done.**

L325 “decay away” decay with time. **Done.**

L330 “for both operational work and for reanalysis systems”—“for both operational and reanalysis systems” **Done.**

L 330”it seems sensible” is that means “it makes sense”? **Done.**

“At the same time” , there should be “,” behand it. Please consider remove “therefore”. and “also”. **Done.**

L 335 “treat as an exponentially parameters”? “rather than to model it directly”? please consider rewrite it . **Rewritten as ”rather than seeking to calculate these covariances with the system model.”** “post processed provided the increments (changes) in the error covariances between the” “post processed, provided that the increments (changes) in the error covariances between the” “and analysis for each filter assimilation window are also stored”—please consider remove “also”. **All Done.**

“error variance”. what is error variance? Please clarify **This means final smoother analysis error covariance. Only the final uncertainties (error variances) of independent state variables are likely to be of interest therefore cross variable error ”covariances” need not be retained.**

L 340 “Lorentz 1963” “Lorenz 1963”. **Done.**

L350 “We also demonstrate the smoothing of the uncertainty estimates in both systems” Please clarify what “both systems” means? **ExtKS and EnKS are now referred to.**

L 370 “the increments (change from filter forecast to analysis)”—“the analysis increments” should be enough. **Done.**

3 Review 3

General comments

The manuscript presented a simplified (ensemble) Kalman smoother as a post-processing for improving reanalysis. It derived smoother equations (including uncertainty estimation) under a simple decay assumption and demonstrated the proposed method in a Lorenz system (1963). RMSEs were significantly reduced, which is very promising. I find the manuscript is well written and reader-friendly. I have some specific comments and suggest the authors revise the manuscript.

Thank you for your interest in this work. It has been useful to consider the points you raise about what EnKF data may be available for smoothing.

Specific comments

L34 ”Ensemble Kalman filters” should say ”The EnKF” **Not done.**

L170 In reality, model has bias and is not perfect. I am wondering when the authors consider model bias, will they converge to the same conclusion?

We briefly note potential problems with model bias in the Discussion section of this paper. Dong et al (2021) noted that it is possible to correct for model bias if the bias component of the filter increments can be identified and these could then be removed prior to applying smoothing. It is possible that better ways of identifying bias could be developed by following error covariances as described here however we must leave that as future work.

L259–270 I understand in the KS case you should introduce the simple approximation (i.e. Eq 12 or 13). But in the EnKS case, all needed information (ensemble mean, covariance, time cross covariances...) can be derived from the ensemble X^a that have been restored any way during the simulation. I do not understand why the simplification is necessary. Do the authors store only he ensemble mean and covariance (uncertainties) rather than all individual ensemble members? That is not common.

In fact only saving the EnKF analyses alone is not sufficient to perform the smoothing. The reason for this is that it is not possible to separately identify the contribution of the observations, i.e. the innovations, to the analysis ensemble, and therefore it is not possible to use that information for smoothing. One could additionally save the EnKF forecasts immediately doubling the expense, or at minimum the ensemble mean increments (Eq 16) along with the ETKF transformation matrices (Eq 26 in our paper), or in the simple smoother framework you need the IP error covariance increments in our Eq 19.

In our paper we point out explicitly that the full past ensembles are NOT

NEEDED for the simple smoother. We agree that current groups investigating smoothing have tended to save all ensemble members but we do not think this is truly practical for big operational models at full resolution with large ensembles due to the massive storage requirements, and in any case it is still not enough to allow smoothing. The approach presented here would save a huge amount of computer storage allowing for example larger ensembles of higher resolution to be run and still smoothed. (See also comments to Reviewer 2, point 1.)

L286 ”The increments are smaller than those from the ExtKF/KS (Fig. 2)” **Done**

L287 ”the ensemble runs” **Done**