Review of the manuscript 'Extending Ensemble Kalman Filter Algorithms to Assimilate Observations with an Unknown Time Offset' by Elia Gorokhovsky and Jeffrey L. Anderson.

The manuscript explores the impact of observation time errors in data assimilation (DA) experiments, using the EKF method. Different approaches of taking into account time errors are explored and compared in a systematic way. The organization of the manuscript (introduction, methods, results, conclusions) is generally good but may be improved in many places. The recommendation is the acceptance of the manuscript after the correction/explanation of some issues, presented bellow.

1) Line 19-21 Authors must include recent references of DA applications in the subjects referred in the manuscript: sea ice, space weather, pollution, paleoclimate, earth's dynamo. long-term reanalyses, estimating prediction model error and evaluating the information content of existing or planned observing systems.

2) Line 29 The phrase: 'attempt to do a completely general diagnosis (Hamilton et al., 2019)' is quite vague. Please rephrase explaining in more detail.

3) Line 76 The time offset or time error ϵ_k^t is assumed to have an unbiased, symmetric, Gaussian distribution. Give arguments to assume those hypotheses. For example, clock errors are much likely to be skewed due to time delays of several origins (electronics, propagation etc.).

4) Line 77 Instead of 'observation error covariance $\mathbf{R}$ diagonal' say 'diagonal observation error covariance $\mathbf{R}$'. In every reference to R, the word matrix must be included.

5) Line 90 (Eq 1). The state vector is represented by letter x(t) while before that is represented by the Greek letter χ(t). The notation must be consistent, all over the manuscript.

6) Eq 2 is assuming a direct observation vector y of the state vector x, i.e. an identity observation operator corresponding to the identity operator. The authors must say that explicitly in order to well understand the difference between x and y. Moreover in eq. 2 and 3, a linear correction of the state vector due to time errors is assumed. Justify this approximation. The linearity hypothesis is only valid for small enough time errors or slow enough dynamics (small time derivative of the state vector). The smallness of time errors is not guaranteed for every type of observations, for instance when observation time is quite fuzzy (e.g. qualitative observations issued from old data, logbooks and diaries).

7) Eq. 6 Use the consistent notation τ^2 instead of τ^2.

8) Explain the title of section 4.3 'Impossible linear estimate' in this section, not in the following one.

9) Line 140. Define the matrix Σ within the text.

10) Line 150 In eq. 13 the vector d must be used instead of d-tilde.

11) Line 153. Authors say, 'It is statistically difficult to distinguish between observation error and error due to time offset' Again this is only valid when linearity hypothesis is assumed leading to the additivity of observation error variance with the error variance coming from the time offset. Say that in the text.

12) Line 170. Present the reason for the choice of the '40-variable Lorenz model'. It is worth to present the model equations here:

$dX_i\ dt = X_{i-1}(X_{i+1} - X_i - 2) - X_i + F;\ i = 1,...,40\ ,\ X_0 \equiv X_{40}\ ,\ X_{-1} \equiv X_{39},$ where the index is the positional index in a periodic ring.

Justify the use of parameter F=8, maybe due to the existence of a chaotic attractor. Present other works using the same model in DA experiments:

Evenson, G. and Fario, N.: Solving for the generalized inverse of the Lorenz model, J. Meteorol. Soc. Jpn., 75, 229–243, 1997.

van Leeuwen, P. J.: Nonlinear data assimilation in geosciences: an extremely efficient particle filter, Q. J. Roy. Meteor. Soc., 136, 1991–1999, 2010

and nonlinear statistics like:

Pires, C.A.L. & Perdigão, R.A.P. (2015) Non-Gaussian interaction information: estimation, optimization and diagnostic application of triadic wave resonance. Nonlinear Processes in Geophysics, 22, 87-108. DOI:10.5194/npg-22-87-2015

13) Line 192. The random draws of the time offset are not allowed to be larger than $P\Delta t$ in absolute value. In that case, authors halve artificially the error to $1/2P\Delta t$ which cuts the occurrence of large time offsets. The most correct procedure would be to take a truncated normal distribution (https://en.wikipedia.org/wiki/Truncated_normal_distribution). Authors must comment on that.

14) Line 205. Authors say: 'This is analogous to the extrapolation performed by Eq. (2) using the ensemble mean estimate as a base point and gives a feeling for how nonlinear the time offset problem is at a particular analysis time.' Explain that describing the behavior in particular times (e.g. near time=369.6 the tangent linear hypothesis has completely failed).

15) Caption of Figure 1 must be completely revised. Refereed symbols do not correspond to the Table inserted in Figure 1 (e.g. green symbols do not exist).

16) In eqs. 17 and 18 do not use the right curly bracket }, since, mathematically, {} is used for the description of a set of members.

17) Lines 255-259. The definition of d^m_i is quite confusing. Accordingly to the text, m is the positional index of the state vector. In lines 190-191, one says that all state variables are observed. Therefore, the observation vector has the same dimension than state vector and therefore the index i in d^m_i (The subscript i references the vector component for a given scalar observation) is quite confusing. Explain that much better, using different indices for different purposes.

18) Lines 260-265. The choice of the threshold T appears to be quite artificial and not generalizable to a generic model.

19) In the caption of Fig. 2, say that the RMSE in each of the 10 experiments (IC trials), correspond to a dot in the graphs, using a particular DA method (color), analysis period and time offset std.