



EGUsphere, referee comment RC2
<https://doi.org/10.5194/egusphere-2022-536-RC2>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-536

Anonymous Referee #2

Referee comment on "Extending Ensemble Kalman Filter Algorithms to Assimilate Observations with an Unknown Time Offset" by Elia Gorokhovskiy and Jeffrey L. Anderson, EGU sphere, <https://doi.org/10.5194/egusphere-2022-536-RC2>, 2022

This manuscript explores data assimilation in the presence of observation time errors (such errors are present in handwritten data used in historical reanalyses). The authors tackle an interesting question and present intriguing solutions. The presentation is at times unclear and can be improved. I suggest that the authors make the revisions suggested below and resubmit.

Major comments:

1. This paper makes numerous assumptions which are not clearly stated. Please go through the manuscript and clearly state all assumptions. Some examples are described below (and also in reviewer #1's comments).
2. The notation in this manuscript is at times inconsistent and not all notation is defined. Again, some specifics are given below, but I have not documented all cases.
3. This one is a bit more open-ended. In this manuscript you incorporate time errors through a correction to the ensemble mean and variance of the prior estimate of the observations. Could you instead update your ensemble members directly (rather than their summary statistics)? For example, instead of using $y_n = h(x_n(t_i))$ could you use $y_n = h(x_n(t_i + e_{i,n}))$ where $e_{i,n}$ is drawn from a distribution of time errors? How would this compare to your current method in terms of computational cost and ease of implementation?

Minor comments:

- L 75: Why is it helpful to think of t^a_k as being equal to $t_k * P$? I agree this is true (assuming $t_0 = t^a_0 = 0$, which is a reasonable), but the statement is confusing to me.
- L 77: "where linear interpolation is used to compute χ between the discrete times". I would rephrase this. You may use linear interpolation to compute χ at non-integer-multiples of the time step, but this an approximation (unless the dynamics are linear). Something like the following would be more correct: "The k th observation is $y_k \sim N(\chi(t_k^o), R)$. In general χ need not be linear and since χ is modeled only at

discrete time steps we do not necessarily know χ exactly at any non-integer multiple of the time step. In this study we make the assumption that the time steps are small enough so that the dynamics are approximately linear between two adjacent time steps. Note that without this assumption the performance of an ensemble Kalman filter may not be very good anyway. In practice, we use linear interpolation to compute χ between the discrete times $\{t_i\}$."

— L 77: I would explicitly state that your forward operator/observation operator is the identity. (Also noted by reviewer #1).

— L 81: The notation for time offset here is different from the notation in L76. I suggest making the notation consistent.

— L 83: I suggest defining "ensemble prior estimate of observations" sooner since it may not be clear to all readers. I see that you define it in equation (2).

— L 110: Is τ known, or is the distribution of τ known?

— L 120: Please explain the equation for the distance.

— Eq. (4): Define the variables used in this equation. I think it is worth writing out what you are doing here in a few steps (you can condense some of the algebra below if you need more space). As I see it, you are looking for the conditional probability that $\epsilon^t = \tau$ given that the difference (which is a random variable itself, call it D) is equal to d . From conditional probability this is proportional to the probability that $\epsilon^t = \tau$ and $D=d$, or equivalently, $\epsilon^x = d - \tau \nu$. What you have here is somewhat confusing because ϵ^t and ϵ^x are independent, but conditioned on $D=d$ they are not at all independent.

— Eq. (5): Remind the reader that you are assuming a normal distribution.

— L 129: If I understand correctly you introduce this term so that you can complete the square and simplify the expression. The word "absorbed" is used in different contexts here and in L 126. Consider using a different word here.

— L 141: This assumption is tricky with the time offset, but perhaps it is okay with your assumption of linearity between two adjacent time steps?

— L 146: Observation error generally means the difference between the observation and the truth, but that is not how I understand ϵ^p . Please explain.

— Eq. (12): See comment about Eq. (4)

— Eq. (15): I don't follow this equation. Please explain and be specific about your notation.

— L 171: Why do you make the choice to use a smaller timestep?

— L 175: Is 100 time steps enough to reach a statistically steady state?

— Eq. (17): "." Is used instead of ","

— Fig. 1: Check colors for consistency with text. Also check that they are colorblind-friendly.