

I want to thank the authors for thoroughly addressing the suggestions that I provided on an earlier version of this manuscript. Despite it has been improved considerably, I have a suite of specific comments that I would like the authors to address before this paper is accepted for publication.

We would like to thank the reviewer for their great contributions that improved the quality of the work. We appreciate their time and effort while reading our work. Please find detailed response to all raised comments and questions below.

1. L3: the authors clarified that they are referring to structural and parametric model uncertainties with “model deficiencies” in the response document, though not in the revised manuscript. Please include that clarification in the abstract.

Good catch! The clarification has been included in the revised manuscript which reads as follows: “Traditional data assimilation methods face challenges during extreme rainfall events due to numerous sources of error, including structural and parametric model uncertainties, forcing biases and noisy observations.”

2. L35: the authors clarified what they mean with “ensemble increments” in the response document, but not in the revised manuscript. Please include that clarification here.

The ensemble increments have been clarified as follows: “The EnKF employs a minimum variance estimator, utilizing observations to compute ensemble increments (i.e., difference between analysis and forecast) that are subsequently linearly combined with the predicted ensemble.”

3. L151, L185 and everywhere else: I do not think that a parameter perturbation approach is the same as a multiphysics approach. In order to achieve the latter, they should have used multiple model structures, which is indeed possible with Noah-MP and other modular modeling platforms like FUSE (Clark et al. 2008), SUMMA (Clark et al. 2015a,b), MARRMoT (Knoben et al. 2019), Raven (Craig et al. 2020), etc. Although I did not catch this in my first review, I think this is a critical point that needs to be revised to avoid confusion among readers.

In the revised manuscript, we omitted the use of the term multiphysics. Both occurrences have been modified accordingly. For instance, L151 now reads: “In addition to this time-varying uncertainty, we also generate an invariant ensemble of channel parameters similar to the configuration of El Gharamti et al. (2021).”

4. L153-154: do you mean that increasing the variability in the ensemble can help your DA scheme to effectively estimate hydrological model states. If yes, please rewrite the text to reflect this.

The text has been rephrased as follows: “Because ensemble DA depends on probabilistic forecasts, enhancing the variability within the ensemble can aid the method in accurately estimating the states of the hydrological model.”

5. L168-169: I recommend adding the calibration objective function as an equation in the paper. This is relevant information for the sake of reproducibility.

We have added the objective function and edited the text of the revised manuscript as suggested: “The objective function was one minus weighted Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970) and NSE of logarithmic streamflow (NSElog), both calculated based on the hourly streamflow simulations.”

$$\text{Minimize } J = 1 - 0.5(\text{NSE} + \text{NSElog}).$$

6. Equation 6: can the authors please clarify if (and how) this equation relates to the calculation of the model error covariance matrix P, and the observation error covariance matrix R?

In DART, the observations are assimilated serially and thus the full state and observation covariance matrices P and R are never constructed. This makes the computations a lot easier because matrix operations (such as SVD, inverse, etc) are not needed. The serial formulation of the ensemble Kalman filter in DART is separated into 2 steps: update in observation space and then regression onto state space. R (a diagonal element for a single observation) for instance is used to compute the observation increments in equation (5). The covariance σ_{xy} in equation (6) is one entry of PH^T corresponding to the covariance between observation y and the j^{th} element of the state (H is the observation operator). The details of this serial algorithm have been covered extensively in the literature, starting with the original work of Anderson (2003) which is cited right before equation (5).

We added text addressing the observation error: “the EnKF solution is obtained as a linear regression of the observation increments Dy on the entire state vector. We note that the assimilated observations are noisy with Gaussian errors and their observation error covariance is accounted for when computing the observation increments in (5).”

We also added a clarifying text at the end of Section 2.2.1 and it reads as follows: “In terms of implementation, we note that the full state covariance in eq. (6) is never constructed using this 2-step serial update scheme. This also applies for the observation error covariance matrix assuming that the observations are uncorrelated in space. For more details on the algorithm and its implementation, the reader is referred to the work of Anderson (2003).”

7. Figure 2: it is quite odd having panel (a) on the right and (b) on the left. Please consider switching the order of those panels.

The panels on Figure 2 have been switched as suggested.

8. L304, 374, 416, 422, 605, 621 and everywhere else in the manuscript: please revise the use of the word “significantly”. This is a point I raised in my first review, and I think that needs to be addressed more carefully.

Although we don't see any problem with the use of the word significantly, we decided to re-word as suggested. Please find the changes below:

L304: “It helps avoid assimilating inaccurate observations, and it prevents the inclusion of observations where the mean of the ensemble members is quite far from the observation value.”

L374: “both the prior and posterior ensemble estimates exhibit improved accuracy”

L416: “This area was affected severely by the flooding event under consideration.”

L422: “although it still shows superior performance compared to the EnKF (bottom panels).”

L605: “predictions at short lead times are aligned with the observations.”

L621: “we illustrate that the hybrid algorithm enhances the performance of the EnKF, notably improving prediction precision.”

9. L393: please re-word this sentence to make it clear that NSE and KGE share the same range of variation.

The sentence has been re-worded as follows: “It's important to emphasize that the KGE and NSE are not directly comparable metrics, despite having the same range of variation.”

10. L422-423: Are the authors referring to the shift in performance metrics between panels (a) and (b)? Note that panel (b) contains results for $\alpha = 0.1$ (i.e., 10% weight to the dynamic component of the covariance, according to equation 7). Then why did the authors write that “outstanding performance can be achieved by incorporating only 10% of the hybrid covariance from the climatology”? Should it be the opposite (i.e., 10% from the dynamic component)?

Related to this point, the authors refer to “The introduction of climatological information with $\alpha = 0.1$ ” in L434. However, according to equation (7) $\alpha = 0.1$ means that you are giving a 10% weight to the dynamic component, and 90% to the climatological component (i.e., $\alpha = 0.1$ means that the authors are introducing dynamic information and not the other way around). Based on this, I strongly recommend revisiting this interpretation of weights and/or correct equation (7).

Regarding the reviewer's first question, we're referring to panel (f) in which alpha is 0.9 and so 10% ($1-0.9=0.1$) of the background covariance come from climatology. This is now clarified in the revised text as follows: "Remarkably, such performance can be achieved by incorporating only 10% (i.e., $\alpha = 0.9$) of the hybrid covariance from the climatology in panel (f)."

As for the second comment, we think the use of the word introduction is confusing. We meant to say, the first-time climatological information is incorporated, and so eq. (7) is correct. To avoid confusion, we have revised the text as follows: "For $\alpha = 0.1$, most of the weight is placed on the climatological information resulting in a notable increase in the ensemble spread."

11. L442: the authors refer to an "overestimation", though it seems that there is underestimation because a large fraction of observations (nearly 40%) is larger than the simulated ensemble members according to the rank histogram. Please revise and re-word if needed.

The sentence has been revised as follows: "At Cowpasture River in the second domain (panel c), a large fraction of the observations (nearly 40%) appears to be larger than the simulated discharge indicating underestimation. The rank histogram of the EnKF also shows partial skewness to the right."

12. The authors refer to an "improved estimate of the uncertainty", though what they are getting is an improved ensemble spread based on the relative range of variation of the observations.

The term uncertainty is replaced with ensemble spread as follows: "The hybrid scheme successfully mitigates that bias and yields an improved ensemble spread."

13. L457-458: This is still VERY hard to visualize from Figure 9, especially in the left panel. Consider decreasing the size of symbols for the hybrid configurations.

We have followed the reviewer's suggestion and decreased the marker size of the hybrid runs. Please find the updated figure in the revised manuscript. We hope that the figure is clearer now.

14. L461: the authors state that "The EnKF-OI schemes yield comparable correlations, with $\alpha = 0.5$ consistently offering the best performance in both domains". I do not think this is true for FL, where $\alpha = 0.7$ offers the highest correlation value (according to Table 2).

The text is now revised, and it reads as follows: "The EnKF-OI schemes yield comparable correlations, with $\alpha= 0.7$ and 0.5 offering the best performance in FL and WV,

respectively.”

15. Figure 13: all the descriptions provided by the authors regarding this figure are still very hard to visualize. Because of this and the length of the manuscript (which contains a tremendous amount of information), I would consider removing these results or sending them to supplementary material. In any case, the authors should make the final choice on this matter.

We thank the reviewer for their suggestion. We really do believe that the spatial variations of the hybrid weight (and the connection to inflation) are an integral part of the story of this manuscript. As such, we would like to keep Fig. 13 and the associated discussion as part of the article.

16. L556-558: please revise this sentence, because I do not see the consistent improvement that the authors describe when comparing red and blue boxplots, especially in panel (a).

Thanks for pointing this out. We have revised the text to reflect the results in Figure 14 as follows: “Compared to the 20-member EnKF-OI with a fixed alpha, the adaptive variant demonstrates relatively similar accuracy with a slight advantage in WV’s low-flow diagnostics.”

17. L558: what is an outstanding score?

Here is the definition of the word outstanding from Cambridge Dictionary:

“outstanding: clearly very much better than what is usual.”

So, in this context the EnKF-OI with 80 members has a clear superior performance compared to the other schemes on the figure. While we don’t see any issue with the term, we have removed it based on the reviewer’s comments.

18. L141: “using level pool scheme” -> “using a level pool scheme”. *Done*

19. L142-143: this sentence reads repetitive. Maybe just write “we use a channel, reservoir, and conceptual groundwater submodel of the NWM, following...”. *Done*

20. L157-158: please place “distributions, e.g., gamma, inverse-gamma and exponential” between parentheses. *Done*

21. Caption of Figure 1: “Dotted box” -> “The dotted box”. *Done*

22. L170: “Summary statistics of the model statistics” -> “Summary model statistics”. *Done*

23. L170-171: awkward sentence. Maybe rewrite as “It should be noted that some model biases remain after the calibration process”. *Done*

24. L422 and everywhere else: please delete the word “outstanding”. There is no need to use bombastic adjectives. *Done*

25. 565: “until now” -> “so far”. *Done*