Response to RC2 for hess-2024-3989: Matthews, G., et al. Error-correction across gauged and ungauged locations: A data assimilation-inspired approach to post-processing river discharge forecasts

We thank the reviewer for their insightful comments and helpful suggestions which we believe will greatly strengthen the evaluation and discussion of the new method. The reviewer's comments have been summarised and numbered for clarity. The authors' responses are in blue.

1. The paper is too long, in particular the description of the methods.
   We will shorten the descriptions of methods by restructuring and condensing the descriptions in Sections 4.1, and 5.1-5.3. Addressing some of the comments from both reviewers such as comment 1, 3, and 12 from RC1 will also reduce the length of the manuscript.

2. I am a bit concerned about how applicable these methods are outside the case study attempted (see specific comments below). For example, the use of a very large catchment is likely to allow the authors to make simplifying assumptions such as that residuals will be normally distributed, or that errors can be characterised using a 10-day window. I would be interested in some discussion of how generalisable these methods are.
   Thank you for this comment. We will extend the discussion on the generalisability of the method to include the estimation of the initial error ensemble (where the 10-day window is used) and the assumption of Gaussianity. See comments 5, 6, and 18.

3. L58 "ensemble Kalman Filters are common data assimilation methods for hydrological applications" this is true for hydrological research, but (to me at least) it remains a curiosity as to why data assimilation within hydrological models - including with ensemble Kalman Filters - remains to my knowledge quite rare in operational streamflow forecasting systems.
   This is a good point. We will specify in "hydrological research applications". In our experience, data assimilation cannot be used in large-scale operational hydrological forecasting due to data latency issues. For weather forecasting, significant international infrastructure has been set up to share and distribute observations rapidly via the GTS, so that observations can be processed in the critical path for forecast production. For gauge data, some observations are available in near-real time, but many are distributed much more slowly. In this study we investigate the use of data assimilation techniques within a post-processing environment as a first step towards an operationalizable post-processing method for ungauged locations. We will make this clearer in the introduction.

4. L90 "Hydrological ensemble forecasts consist of N potential realizations referred to as ensemble members" I think it would be good to state explicitly which variable(s) you are discussing here, as it wasn't clear to me - I'm assuming streamflow (or runoff, as it's on a grid?)?
   The variable of interest is streamflow or river discharge. We will change this to "The hydrological ensemble forecasts consist of N potential realizations of future river discharge referred to as ensemble members."

5. L107 I would have thought with a strongly skewed (and potentially zero bounded) variable like streamflow, an additive error only generally holds after a normalising transformation has been applied (and, if applicable, zero values have been dealt with).
   We will add normalising transformations to the discussion (see comment 18). The reviewer is correct that the assumption of Gaussianity limits the applicability of an additive error by resulting occasionally resulting in negative discharges. We deal with negative discharge values

as described in Section 4.1. The impact and potential solutions are discussed in Sections 6.2 and 7.

6. L112 Similar to the above criticism at L107, Equation 6 appears to assume that errors are normal and homoscedastic. If my understanding of what is being assimilated is correct, this is highly unlikely to hold for streamflow, for which residuals are almost always non-normal and heteroscedastic. See e.g. Smith et al. 2015, among many others.
   We agree that streamflow residuals are often non-Gaussian and heteroscedastic. Our framework does employ updates based on Gaussian assumptions as we use the LETKF, but the resulting distribution is not necessarily Gaussian (Reichle et al., 2002). However, we do not assume homoscedasticity: the ensemble spread evolves dynamically and reflects state-dependent and lead-time-dependent error variability. We will change Equation 6 to express the distribution of the updated ensemble more accurately.

7. L145 "we adopt the common assumption that the error is constant" I would not have said this is common. I would say it's much more common to use autoregressive models (often AR1) to describe the autocorrelation between residuals in streamflow. I understand why this is a pragmatic simplification, but errors often do change with lead time as the value of forecast information decays.
   Thank you for the comment. We will add the following to Section 3.1 and add a discussion of this assumption (and possible developments) to Section 7: “As the true evolution of the error vectors at all grid-boxes is unknown, we assume a simple persistence model, such that $\mathbf{b}_{\mathrm{k}}^{(\mathrm{i})} = \mathbf{b}_{\mathrm{k}-1}^{(\mathrm{i})}$. This is a common assumption used in state augmentation (Pauwels et al., 2020; Rasmussen et al., 2016; Ridler et al., 2018; Martin, 2001).”

8. L149 "define the propagation" I'm not sure what 'propagation' means here, given the error is assumed constant in time. Can the authors clarify? Nevermind - the authors do this in Section 3.2! The authors may want to flag that the explanation for this is coming.
   We will highlight that the propagation equation defined on line 150 if for use in the LETKF described in Section 3.2

9. L180 "(see Eqs. (8) and (9) in Bell et al., 2004)" I feel that if the authors need to specify equations from another study to describe these methods, the equations should be present in the paper (in an appendix is fine) - especially Eq (9) of Bell et al. which the authors later describe as 'key' to the method. (Unless they are included later?)
   We will add the decomposition of the Kalman gain matrix as an appendix.

10. Figure 1 - this is a really nice, clarifying figure.
    Thank you!

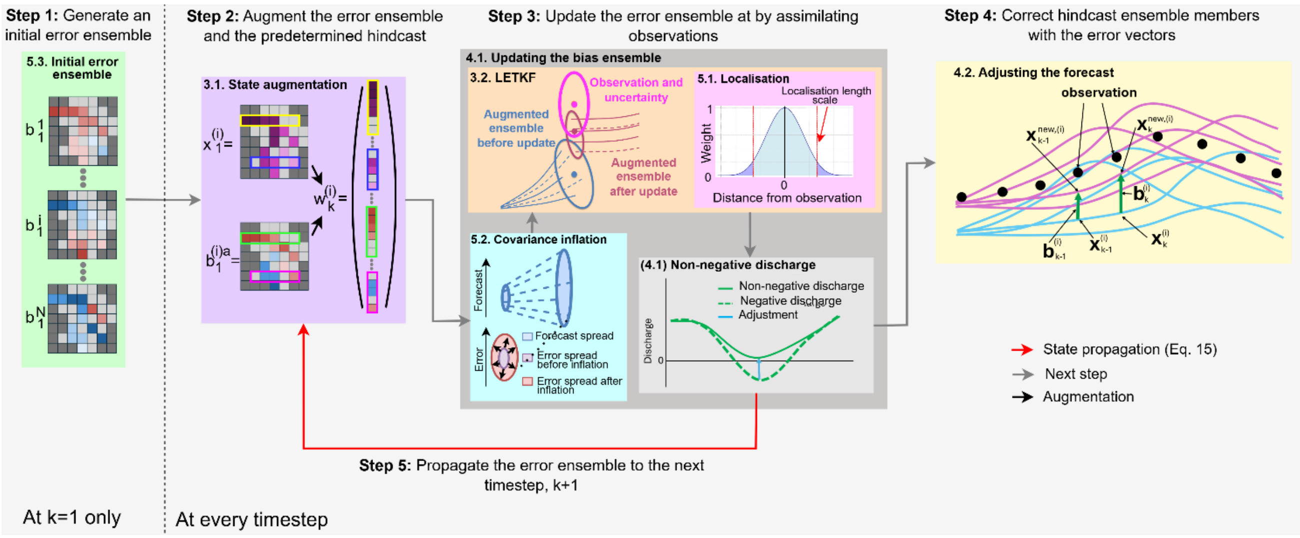
11. L223 "We enforce non-negativity by further adjusting the error ensemble members after the LETKF update step (Fig 1)." This indicates that zero values are present in output state, indicating that errors are not continuously distributed. I realise not everyone handles zeros, but it would be good to acknowledge the limitation of this assumption (as noted above).
    We will extend the discussion regarding the assumption of Gaussianity for river-discharge. See comment 5.

12. L265 "Eq. 4.10 in Gaspari and Cohn (1999)" - I think the authors should include this equation, as well as discussing (briefly) why they thought the form of this equation appropriate for this task. The regionalisation of errors is in my view the major contribution of the paper.
The Gaspari and Cohn correlation function is a commonly used localisation function in data assimilation as it smoothly decreases to a definable radius. We will add the Gaspari and Cohn function as an appendix.

13. L272 "We propose instead for the localisation length scale to be defined as the maximum distance between any grid point and its closest observation." This seems like a sensible choice.
Thank you.

14. L325 "(here 10 days)" This is a long period over which to assess an error - some use periods of this length for bias correction (e.g. Bennett et al. 2021). I'm assuming this really only works for larger catchments where rivers have slower varying errors; I would have thought for small headwater gauges shorter periods would be more appropriate. It also explains why errors are assumed not to vary with lead time, above. This is all fine, but the authors may wish to mention this in their discussion.
This is a good point. The 10-day period is used to generate the initial error ensemble mean. This initial estimate is not used to correct the river discharge ensembles directly, but rather to provide a starting point for the LETKF. The LETKF then updates the error ensemble at each timestep. We will update Figure 1 to better clarify this process (see Figure below).

We selected a 10-day period to capture the consistent biases of the hydrological model but also to allow for seasonal/dynamic variation in this bias. We agree that shorter periods may be more appropriate for smaller, fast-responding catchments and will clarify this in the discussion.



15.

a. L408 "we assume that the observation errors from different gauge stations are uncorrelated" I'm not suggesting a change here, and I think this is a reasonable suggestion without additional information. But I suspect the long-range nature of the errors (a 10 day period) may undermine the assumption somewhat.
The observation errors we refer to on line 408 are the errors described in Equation 7. The sources of these errors are instrument uncertainty, observation processing, observation operator error and scale mismatch between the observations and the model resolution.

Observation errors are assumed to be uncorrelated with the prior errors which is a standard assumption in data assimilation. We will clarify this in the text.

b. I'm also curious what happens when errors are propagated in space: what happens when you get a point equidistant (or close to equidistant) from two gauges, and the errors from the two gauges interact in some way (e.g. cancel each other, or sum).
The Kalman gain matrix governs the spreading of observation information in space. A weighted mean is calculated. The weight of an observation is determined by the cross-covariances between the hindcast and error ensembles, the distance from the observation (via the localisation), and the uncertainty in the observation itself. The left and central panels in Figure 2 show single observation experiments, indicating how information from one observation is spread spatially. The panels on the right of Figure 2 show how observation information is propagated when all available observations are assimilated. A study into the impact of the specific locations of the observations is left for future work. We will add this discussion to Section 7.

16. L438 "forecast mean is decreased by the proposed method we use the Normalised Mean Absolute Error" It's preferable to apply measures of absolute error to the ensemble median. See, e.g., Taggart (2022).
Thank you for this very helpful comment. As it is the ensemble mean that we want to evaluate we will use the normalised RMSE and make the appropriate changes to Sections 6.4 and 7.1.

17. L526 "However, this assumption is necessary to propagate the hindcast to the next time step without the use of a hydrological model (Section 3.1)." Perhaps, but one application of ensemble predictions is to sum ensemble members through time (e.g. to assess cumulative inflows to reservoirs). From this figure, it seems this would result in highly unreliable accumulations. This may not be an application of EFAS (I don't know), but if the method is to have more general applicability this is a serious weakness.
The hydrographs shown in Figures 4b and 4e are not the final hydrographs resulting from this method but instead are an intermediate step used to investigate the impact of the methodological assumptions made. We will restructure Section 7 to make a clearer division between the results that show how the method works and the results that show the skill of the resulting ensemble (see comment 3 from RC1).

18. L660 "Future work could look into applying anamorphosis to make the ensemble distribution more Gaussian-like" I'm not familiar with the concept of anamorphosis, but a conventional way of doing this is to use normalising transformations, of which many are available for hydrological variables.
Anamorphosis is very similar to normalising transforms used in hydrology. We will add the use of normalising transformations to this paragraph.

19-23. Typos and grammatical errors.
These typos will all be corrected. Many thanks to the reviewer for catching them!

References
Martin, M. J. (2001). *Data assimilation in ocean circulation models with systematic errors* (Doctoral dissertation, University of Reading).

Pauwels, V. R., Hendricks Franssen, H. J., & De Lannoy, G. J. (2020). Evaluation of State and Bias Estimates for Assimilation of SMOS Retrievals Into Conceptual Rainfall-Runoff Models. *Frontiers in Water*, *2*, 4. https://doi.org/10.3389/frwa.2020.00004

Rasmussen, J., Madsen, H., Jensen, K. H., & Refsgaard, J. C. (2016). Data assimilation in integrated hydrological modelling in the presence of observation bias. *Hydrology and earth system sciences*, *20*(5), 2103-2118. https://doi.org/10.5194/hess-20-2103-2016

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **130**, 103–114, https://doi.org/10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.

Ridler, M. E., Zhang, D., Madsen, H., Kidmose, J., Refsgaard, J. C., & Jensen, K. H. (2018). Bias-aware data assimilation in integrated hydrological modelling. *Hydrology Research*, *49*(4), 989-1004. https://doi.org/10.2166/nh.2017.117