Response to RC1 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 comments and suggestions which we believe will greatly improve the manuscript and strengthen the motivation for the method. The reviewer's comments have been summarised and numbered for clarity. The authors' responses are in blue.

1. The paper is difficult to follow due to dense mathematical notation and long, complex sentences.
   We will revisit the notation in the paper. We note that, where possible, we already use standard mathematical notation for data assimilation following Ide et al. (1997). We will add this reference to Section 2. We will reduce the number of symbols where possible, such as the symbol for the augmented observation operator and some superscripts. We will also remove the additional notation used to describe the approach for dealing with non-negative values and estimating the initial error ensemble as this notation is only used in Sections 4.1 and 5.3, and instead describe this with words.

   We will review the text in the manuscript and shorten particularly complex sentences throughout.

2. There is an overuse of jargon without sufficient introductory explanation for a broader hydrology audience.
   We will review the manuscript and remove unnecessary jargon, such as "spatiotemporal consistency". Where data assimilation technical terms are deemed by us to aid the description of the method, such as state augmentation, we will add a clear description of the term (see also comment 4).

3. The structure could be more concise, with a clearer division between methodology and results.
   Thank you for this comment. 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. Where possible we will remove repetition and unnecessary detail (see comment 12).

4. The state augmentation approach is described in a way that makes the approach seem unnecessarily complex.
   We will add the following sentences: "*State augmentation is a technique used in data assimilation to estimate the state and parameters of a system simultaneously. An augmented state is defined by appending the parameters to the state vector allowing both to be updated by the data assimilation method.*" We will also move the definition of the error ensemble to Section 2, streamlining the state augmentation description.

5. The assumption of constant error propagation is not well justified.
   We will add the following to Section 3.1 and add a discussion of this assumption 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)."

6. Also, related to this, the use of precomputed model outputs instead of an evolving state might introduce additional errors, which are not sufficiently discussed.
   We will amend the discussion of this assumption (line 219) to: "The assumptions made in Eqs. (18) and (19) make our system sub-optimal from a data assimilation perspective but are necessary to avoid rerunning the hydrological model. Importantly, we aim to estimate the error of the precomputed model output at each lead time. Therefore, while the lack of state evolution makes the hindcast component update sub-optimal, the update of the error ensemble remains mathematically consistent".

7. Inflation:
   We will restructure Section 5.2 to address the following comments.
   i I find the inflation method to be heuristic with little to no mathematical rigor. For instance, the assumption that the hindcast variance is a proxy for error growth does not account for potential biases in the raw ensemble itself.
   The reviewer is correct that the covariance inflation technique used is a heuristic method. We initially tried a simpler multiplicative inflation but found that this was not suitable due to the large variations in hindcast spread as a function of lead-time.
   Our new approach is a practical solution to the issue of filter divergence that is inspired by blending techniques such as the RTPP method (Zhang et al., 2004). We will add a comment to Section 5.2 to make this clearer. The limitations of using the hindcast uncertainty as a proxy for the uncertainty in the error estimate are discussed in the discussion section (lines 646-656).

   ii Unlike RTPP, the proposed inflation blends analysis and "estimated" perturbation information without explicitly evolving them. What motivates such an approach?
   Our goal is to use this approach for post-processing with pre-computed ensemble forecasts. Explicit online evolution of perturbations is not feasible in this scenario. Hence, the error ensembles are evolved between timesteps using a persistence model (see comment 5). Blending the evolved perturbations with an estimated perturbation method is inspired by palaeoclimatological reanalysis work such as Valler et al., (2019) where a climatological error-covariance matrix is blended with the background error-covariance matrix. Rather than a climatological matrix we use "estimated" perturbations. During the development of this method, it was found that the "estimated" perturbation matrix must 1) preserve the spatial structure of the river network and 2) must be adaptive, to avoid filter divergence and maintain physically plausible estimates along the river network. The hindcast ensemble perturbation matrix satisfies both these requirements. We will add this motivation to Section 5.2.

   iii The inflation parameter, alpha, is computed from a 3 steps-average of the hindcast (eq. 28). Why this choice is appropriate? I recommend testing with different alpha values through sensitivity experiments.
   Sensitivity experiments were conducted in the development of this method, but we did not include the results in the original manuscript for brevity. Our analysis indicated that 1) a constant alpha value was not suitable at different lead-times due to the change in hindcast spread, and 2) a lead-time dependent constant alpha value was not suitable for different flow situations. We therefore selected a method that is forecast dependent. An average across 3 steps was selected to ensure a smoothly changing alpha mitigating instabilities. We will add comments along these lines to Section 5.2.

iv If inflation is not localized along the network, that should be clarified and justified. The inflation factor does not vary in space (although it does vary in time). It is applied to perturbation matrices that themselves vary in space, consistent with the river network. Thus, the inflated covariances provide physically plausible error-variances and error-correlations between locations. While spatially varying inflation methods are available, (e.g., Kotsuki et al, 2017) this would result in many more inflation parameters to estimate, and the results may not be spatially consistent as each element gets inflated separately. We will make this clear in Section 7 (discussion).

8. Spread: It's clear that the method tends to overcorrect at short lead times but yields underconfident ensembles at longer lead times (as shown in Figs. 4, 5). In general, one expects the ensemble spread to accurately represent the forecast uncertainty but the issues the authors face could be related to the ad-hoc inflation.
We agree with the reviewer that the inflation method is a reason for limited reliability of the ensemble-spread at longer lead-times. However, we also note that it is very rare that the ensemble spread accurately represents the forecast uncertainty for all lead times and locations (e.g., see Kotsuki et al, (2017)'s comparison of inflation approaches). For our proof-of-concept study, we have not carried out extensive tuning experiments, instead making some pragmatic choices. A study comparing different inflation approaches for the context of post-processing ungauged locations is left for future work. The limitations of using the inflation method are discussed on lines 646-656. We will extend the discussion on the spread in Section 7.

9. I would also note that real hydrological errors are dynamic, but the paper assumes the errors to remain constant between cycles. A flow-dependent error propagation model and perhaps an adaptive inflation approach could address these issues.
The error covariance propagation is flow dependent (based on an ensemble of precomputed hind-casts) and the inflation factor is adaptive in time. Please see our response to comments 5 and 7.iii.

10. Localization: The choice of the length scale (262 km) should be better justified. There is no sensitivity analysis to determine whether this choice is optimal or whether smaller/broader radius would improve the results.
Sensitivity experiments were conducted during the development of this method. It was found that the optimal length scale varied by location, lead-time, and tuning metric of choice, but overall the differences were small for length scales from 65km to 786km. We therefore decided to define the length scale as the maximum distance between any grid-box and its closest river gauge which for our case study is 262km. This definition ensures all grid-boxes are updated by the LETKF and allows the method to be transferred to other catchments and models without the need to perform computationally expensive tuning experiments. We will make this clear in Section 5.1.

11. Also tangential to this, the authors need to revisit the equal error correction assumption in upstream and downstream locations. Overall, upstream locations are less dependent on distant downstream observations. Obviously, downstream conditions are often affected by accumulating upstream flows.
The propagation of the error-correction along the river network is determined by the ensemble covariances and the localisation applied. This means that while the localisation length is equal

upstream and downstream, the actual analysis increments are not. This can be seen in Figure 2 where we show the analysis increments for single observation experiments. We will add comments to the discussion of Figure 2 to make this point clearly.

We agree with the reviewer that the relationships upstream and downstream are different. This is shown in Figure 3a. The cross-correlations between the hindcast and error ensembles are strongest along the river stretch near the station and decrease at longer distances. The larger correlations downstream of the station are along the flow path of the river whereas upstream the correlations show a more branch like structure because the station location is impacted by accumulation of flows from all upstream tributaries. We will add comments to the discussion of Figure 3 to make this point clearly.

The localisation also dampens the influence of distant observations. Emery et al., (2020) investigate the use of localisation along the river network. They found that an observation can be beneficial to both upstream and downstream locations particularly for distances for which the flow transit time is less than the time between analyses. We will extend the discussion regarding the localisation in Section 7.

12. Figures: The figures are well-intended but too dense and overloaded with information, making them difficult to interpret and extract keys findings. I suggest splitting the complex ones (e.g., Figs. 3, 4, 6) and definitely simplify the annotations
Thank you for this comment. To address this comment, and comment 1 of RC2, we will reduce the content of some of the figures. We will make the following changes to the figures:
    - Figure 3: We will remake this figure removing panels c and g. Some of the information from c and g is duplicated in the hydrographs in Figure 5. Removing these panels will shorten the discussion (see comment 1 from RC2). This will also allow more room for remaining panels of Figure 3 making key details clearer.
    - Figure 4: We will simplify the annotations as request by the reviewer. We will combine the legends making the comparison between panels easier and more clearly indicate the difference between rows 1 and 2.
    - Figure 6: We will remove the panels a, d, g, and j, and related discussion as this information is also shown in the remaining panels. We will also remove the river names, which are already shown in Figure 3. We will combine the legends of panels c, f, i, and l and clearly label the metric shown in each column.

13. Line 6: “Error vector for each ensemble members” seems vague and unclear.
Thank you for this comment. We will change this sentence to read: “Our new method employs state augmentation within the framework of the Local Ensemble Transform Kalman Filter (LETKF). Using the LETKF, an error vector representing the forecast residual is estimated for each ensemble member.”

14. Line 12: The term “proxy” could mean a lot of different things. Clarify the nature of updates, whether that’s real data assimilation experiment or an OSSE.
Our experiment uses real gauge-data and meteorological forcings and we will make this clear in the introduction. We will change this statement to read “A spatial cross-validation strategy is used to assess the ability of the method to spread the correction along the river network to ungauged locations”

15. Line 160: I would use "cycled" instead of "iterated"
    Thank you. This will be changed.

16. Line 160: Replace "at each timestep" with "at each observation time"
    We use "timestep" rather than "observation time" as the analysis times are dictated by the availability of the precomputed hindcast data as well as by the availability of observations. For clarity, we will change it to "at each hindcast timestep for which observations are available".

17. Line 178: Replace "weights" with "weighs"
    Thank you. We will rephrase this sentence.

18. There are too many "see section xxx". This made navigation frustrating; I kept going back and forth. Consider restructuring for better flow.
    We will remove some cross-references and reword others to e.g., "(described in Section X)" as a signpost for the reader for the more novel components of the method (see comment 8 in RC2).

19. The word "improved" is overused in my opinion. Consider other synonyms "enhanced", "refined", ...
    We will reword sentences where appropriate to be more specific about the effect being described.

20. Explain technical terms more clearly, for instance "spatiotemporal consistency"
    See comment 3.

References

Emery, C. M., David, C. H., Andreadis, K. M., Turmon, M. J., Reager, J. T., Hobbs, J. M., ... & Rodell, M. (2020). Underlying fundamentals of Kalman filtering for river network modeling. *Journal of Hydrometeorology*, *21*(3), 453-474. https://doi.org/10.1175/JHM-D-19-0084.1

Ide, K., Courtier, P., Ghil, M., & Lorenc, A. C. (1997). Unified notation for data assimilation: Operational, sequential and variational (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, *75*(1B), 181-189. https://doi.org/10.2151/jmsj1965.75.1B_181

Kotsuki, S., Ota, Y., & Miyoshi, T. (2017). Adaptive covariance relaxation methods for ensemble data assimilation: Experiments in the real atmosphere. *Quarterly Journal of the Royal Meteorological Society*, *143*(705), 2001-2015. https://doi.org/10.1002/qj.3060

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

Valler, V., Franke, J., & Brönnimann, S. (2019). Impact of different estimations of the background-error covariance matrix on climate reconstructions based on data assimilation. *Climate of the Past*, *15*(4), 1427-1441. https://doi.org/10.5194/cp-15-1427-2019

Zhang, F., Snyder, C., & Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Monthly Weather Review*, *132*(5), 1238-1253. https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2