Response to reviewer's report 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 providing a second review of this manuscript and once again providing thoughtful comments and suggestions to improve the clarity of the manuscript. The authors' responses are in blue. Line number, sections, and figures refer to revised manuscript.

In addition to the reviewers' suggestions, we have corrected typos identified in our own review of the manuscript.

Specific comments

1. L84 "Therefore, we refer to these ensembles as hindcasts for clarity." Sorry - I meant to pick this up in the previous revision: what is meant by 'these' here? It appears that here 'hindcasts' refers to something other than 'ensembles of river discharge that we are error correcting', as defined in the opening sentence of this paragraph. Please clarify.

The "hindcasts" do refer to the ensembles of river discharge. In this paper, the ensembles are operational EFAS ensemble forecasts from 2021. However, when we apply the error-correction method we use observations from the forecast period. This is not possible in an operational system as these observations would be in the future. We therefore chose to use the term hindcast to highlight that these ensembles are not valid forecasts. We have added this clarification to lines 88-91:

"In this paper, these ensembles are past operational EFAS forecasts (see Section 6.1). However, when we perform the error-correction we use observations that are available within the forecast (hindcast) period. Observations are not available during the forecast period in an operational system, since these timesteps are in the future. Therefore, we refer to these river discharge ensembles as hindcasts to indicate that the ensembles are not valid forecasts."

2. L199 Figure 1: In the plot accompanying Step 4, b(i)_(k-1) appears to be smaller than b(i)_k. I'm not sure if this is an error, but on face of it this seems to contradict L147 "we assume a simple persistence model, such that b(i)_k = b(i)_(k−1)". Should the figure be showing "b^hat^{(i)a}_k"?

Thank you for spotting this error. Yes, the figure should show "b^hat^{(i)a}_k". Figure 1 has been updated.

3. L220 Para starting with "The Kalman filter is not constrained to enforce non-negativity..." thanks for clarifying this. I would imagine that if corrections are applied to the falling limb of the hydrograph (i.e. where the error is computed at high flow and

propagated to low flow) flows would be corrected to zero quite often, especially as b(i)_k = b(i)_(k−1). This problem would get worse if the method is used to correct multiple lead times. I'm not suggesting a change here - the authors have stated what they have done clearly (at least, if I've understood this correctly!) - but I'd note this is likely to be a serious short-coming for operational deployment, particularly in flashier catchments where flows can vary rapidly over short periods of time.

The error vector is updated at each lead-time with the assumption of a constant error (i.e., b(i)_k = b(i)_(k−1)) only assumed during the propagation step of the LETKF. However, the reviewer is correct that the method does struggle to update the error vector correctly for the falling limb of hydrographs as shown in Fig. 5b. This is due to the small ensemble spread (as discussed in lines 584-586). Potential solutions to address the negative discharge issue are discussed in Section 8. The need to correct negative discharge is most often due to the spread of the ensemble not being corrected sufficiently (see lines 622-627) and to the non-Gaussian distribution of river discharge (see lines 658-663).

4. L234 "b^hat^{i}_k" I think this should be "b^hat^{(i)a}_k"?

Thank you for spotting this typo, it has been fixed.

5. L458 A useful addition to this plot would be the location of the two gauges being discussed, plotted on the right-hand map.

Thank you for this suggestion. The map has been updated.

6. L520 i) "This assumption does ensure the analysis hindcast component is always physically plausible" Not necessarily suggesting a change here, but I could not follow why this this enforces non-negativity. ii) The other question I have here what is causing the update is changing with lead time? Is it that the covariance matrix changes because of changes in the raw ensemble? Or have you assumed the availability of observations (from the statement at L89 in the introduction: "when we perform the error-correction we use observations that are available within the forecast (hindcast) period")?

i) The assumption made is that the precomputed hindcast ensemble is a good approximation for the analysis hindcast component. The precomputed hindcast ensemble is the raw output of the LISFLOOD hydrological model and is constrained by the model's physics. I have made this link to non-negative river discharge more explicit (lines 521-522).: "e.g., the river discharge is always positive as this is a constraint within LISFLOOD".

ii) The reviewer is correct that we have assumed the availability of observations, and hence we therefore refer to the river discharge ensemble as hindcasts as they are not valid forecasts after the error-correction (see comment 1). As the reviewer notes the updates at each lead-time are also impacted by the changing covariance matrix of the raw ensemble. The impact of the lead-time dependent updates of the error vectors is discussed in Section 7.1.2.

7. L538 "As demonstrated in Figs. 4c and 4f, this can result in the error ensemble spread being large for the rising limb of an event and smaller for the falling limb." I would say this is a desirable property for a skewed variable, so long as the ensemble spread is still reliable.

We agree with the reviewer that if the ensemble spread is reliable then this property of the ensembles is beneficial. However, the ensemble mean can still be biased. The update of the error vectors is impacted by the spread of the ensemble, resulting in very small updates to the error vector when the spread is narrow (as shown in Fig. 5b). Discussion of the limitation of the method is in lines 646-657.

8. L556 Figure 6. A nice addition to this very informative figure would be the proportion of gauges where corrected forecasts outperformed the raw ensembles for (a)-(d), now that the histograms have been removed. It's sometime hard to make out how many stations have black rings around them or not - e.g. in (b) - so this would neatly summarise this information.

Thank you for this suggestion. The proportion of degraded gauges has been added to the figure caption.

9. L586 "The decrease in N-RMSE, despite an increase in mean bias, suggests that the error-corrected ensembles consistently underestimate flow, while the raw hindcast ensemble fluctuates more between under- and overestimation, which can compensate for each other in the mean bias metric." I'm not sure I agree with this interpretation. 6(b) shows that while there is slight tendency to underestimate flows at some gauges, many sites are unbiased and some have positive biases. Measures of mean squared error applied to skewed variables like streamflow tend to emphasise errors at high flow; better performance at high flow does not suggest a more general tendency to underestimate flow.

Thank you for highlighting this error. The reviewer is correct. This paragraph should have been updated when the NMAE was replaced with the N-RMSE. We have revised the text (lines 587-598):

"Overall, the error-corrected ensemble reduces the N-RMSE but there are 14 stations where the skill is reduced. Typically, these stations are on the upstream reaches of their respective rivers (Fig. 6d; see discussion on correlation). Interestingly, the N-RMSE does not follow the same spatial pattern as the mean bias. This divergence indicates that the correction method is more effective at reducing large errors than at addressing systematic biases. One possible explanation is that the error vectors adjust too slowly to changes in forecast errors between time steps. This slow adjustment is particularly problematic when errors fluctuate around 0 m^3s^{-1}, since alternating positive and negative deviations may not be corrected quickly enough and can accumulate into a worsening mean bias. When the error magnitude is large, the gradual adjustment is less detrimental because the sign of the error is usually captured correctly even if its magnitude is not. However, at upstream stations, where rivers are smaller and respond 595 more quickly to rainfall, large errors often persist for shorter durations, making the slow adjustment of the error vectors more detrimental. This likely contributes to the increase in N-RMSE observed at these upstream stations. Further development of the method—for example, allowing the error vectors to evolve during the propagation step of the LETKF in addition to the update step—could enable faster adaptation to changing forecast errors."

10. L596 "7.2.2 Skill of the ensemble distribution" and Fig 7. It wasn't clear to me which set of predictions is used to generate the plots discussed in this section. Is it the leave-one-gauge out cross-validated predictions (which would be preferable, as the main contribution of the paper is for error predictions in ungauged regions)? Please specify.

Yes, it is the leave-one-gauge out cross-validated predictions that are used. This is stated in lines 545-546.

11. L658 "A transformation between river discharge and specific discharge (river discharge divided by upstream area) could be used to ensure that the ensemble covariances more accurately represent the true relationship between locations." Ok, and I'm not suggesting a change here, but this could mean that more weight is given to upstream gauges in the error analyses, as errors in the timing and location of rainfall tend to cancel over larger areas, resulting in relatively larger errors in headwaters. Using specific discharge would reduce the dominance of larger rivers in the error update, since their larger ensemble variances would have less influence, and the update would instead rely more on the correlation between gauged and ungauged locations. However, as the reviewer notes, a consequence of using the specific discharge could be that upstream gauges are given disproportionately high weights. A study of the benefits and drawbacks of specific discharge, as well as other possible transformations, is left for future work as discussed in lines 661-669.

Typos etc.

12. Figure 1: "Step 3: Update the error ensemble at by assimilating observations" either there is something missing after 'at' (perhaps 'k+1'?) or delete 'at'. Done. Thank you.

13. L261 "Appendix ??" missing cross-reference

The cross-reference has been fixed.

14. L368 "The minimum value across the stations is 0.516 m3s−1 and the maximum value is 7662.917 m3s−1." suggest rounding these numbers: "The minimum value across the stations is <1 m3s−1 and the maximum value is 7663 m3s−1."

Done. Thank you.

15. L489 "with which the correlation" should be "for which the correlation"

Corrected. Thank you.

16. L491 "the correlations begins" should be "the correlations begin"

Corrected. Thank you.

17. L519 "although, the perturbations" delete comma

Removed.