Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC2
https://doi.org/10.5194/hess-2020-642-RC2, 2021

Comment on hess-2020-642

Anonymous Referee #2

Referee comment on "Ensemble streamflow data assimilation using WRF-Hydro and DART: novel localization and inflation techniques applied to Hurricane Florence flooding" by Mohamad El Gharamti et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-642-RC2, 2021

The paper presents an application of streamflow data assimilation for improving hourly streamflow forecasts with a focus on two methodological innovations:
(1) improving the model forecast/prior via covariance inflation that changes in time and space
(2) improving the model update/posterior by better capturing spatial correlations between states in the model river network via along-the-stream (ATS) covariance localization.

Overall, this is a well written paper that clearly illustrates the benefits of the proposed DA methods (ATS localization and spatial-temporal inflation) for streamflow forecasting. However, the paper can and should be further improved taking into account the comments listed below.

Main comments:

1. Novelty: the abstract lists the two methodological innovations mentioned above. From the literature review in the introduction it is however not entirely clear to what extent these are innovations. For example, the text (lines 60-67) does not specify whether space-time adaptive covariance inflation has been applied to flood forecasting, except that "the impact of inflation on streamflow predictions is not fully understood", without providing a reference. Similary, regarding ATS covariance localization, the introduction (lines 78-82) does not mention whether this technique is proposed here for the first time or if it has been applied in other studies. So, my suggestion is to make the exact contributions of the paper more explicit.

2. Title of the paper suggests that main innovation of the paper is data assimilation during an extreme rainfall/flooding situation (hurricane), whereas abstract/intro focuses on methodological innovations as the main contribution (see point 1). Please clarify/make consistent.

3. Methodology: key novel parts of the methodology are not described in sufficient detail, specifically the ATS covariance localization strategy in section 3.2 and the inflation method in section 3.3 (see more details below).

4. Results: the provided results and figures do a good job of illustrating the benefits of the proposed methods, so no major comments in this respect. I do however have some comments on providing additional results, see below.

Detailed comments (more or less in chronological order):

-line 83: typo in "hydrological"

-line 164: "The coefficients in equation 2 can be found in the literature", please provide a reference

-line 221: include units for these values

-line 224: "We did not investigate the effect of ensemble size on the results within this study". Ok but it would still be helpful to address whether a larger ensemble size would change the results and conclusions of the paper. E.g. since the methods aim at fixing sampling errors due to small ensemble size (among other things), does their benefit decrease with larger ensemble size?

-line 227: priors on stream channel parameters are missing from fig. 2

-line 269: not clear what is meant by "bucket distributions"

-eq.14: subscript k is not defined

-I was wondering whether the USGS streamflow data (rating curve based) are still accurate during a hurricane. Reading section 2.7 it seems the answer is "no", since the paper uses revised streamflow data. On line 253 it is stated that using original vs revised streamflow data had significant impact on the results, yet only results with revised data are presented. Do the methods proposed here still work when using the original non-revised data? And if not, why not and what is needed? This should then be addressed in the discussion as an open problem for DA under realistic hurricane conditions.

-line 260: "observation error plays a somewhat secondary role in the quality of the assimilation, ". I guess that is after revision of the original streamflow data!?

-line 276: define 'observation increments'

-line 279: "alpha is computed...", this sentence remains very cryptic and unclear at this point. Suggest to refer to section 3.2 where it is explained in more detail.

-eq. 17: in this paper, is inflation applied to forecast, analysis, or both? Edit: ok later in the paper this becomes clear, but good to briefly mention here as well.

-section 3.2 (localization): this section is a key contribution of the paper and needs to be better explained. The section describes the approach in words and illustrates what it looks like in figure 4 (nice figure), but it doesn't actually show how to implement it. Please include relevant mathematical expressions so that the methodology is reproducible.

-line 303: ATS localization assumes flow of information only travels downstream not upstream. Why is this a reasonable assumption and why is this assumption needed?

-line 331: why don't you show these results to corroborate your conclusion that r=100km works best? It may further be helpful to plot forecast performance as a function of localization radius for all gauges to corroborate the statement on line 332-333 that

forecasts deteriorate for r<100 and r>100.

-section 3.2.2: this is a nice section that reports promising results of ATS vs regular localization (table 2)

-section 3.3 (inflation): similar to section 3.2, the inflation method should be better explained. Only Eq. 18 is now given, which shows the standard prior times likelihood formula (but note that the posterior on the left is proportional to, not approximately equal to, prior times likelihood). I suggest to provide more details on the likelihood function: what does it look like, what are the underlying assumptions, why are the assumptions valid. Also the difference between prior and posterior inflation should be better explained.

-line 392: "To find the updated value of the inflation, $p(\lambda|d)$ is maximized and the resulting value is used as a prior for the next DA cycle". But after maximization one (optimal) value is obtained, how can this single value be used as a prior distribution in the next DA cycle? It sounds like this approach does not keep track of the posterior of the inflation factor? Please clarify.

-line 546: "running WRF-Hydro with a land surface..." missing 'model'?