hess-2015-54: Authors' response (in *blue italics*) to comments by Anonymous Referee #1 (in black)

General Comments:

This manuscript titled "Operational aspects of asynchronous filtering for hydrological forecasting" compares the traditional Ensemble Kalman fitler (EnKF) with its modified version, asynchronous EnKF (AEnKF). The main difference is in the update step, with the traditional EnKF using observation data from a single time step whereas the AEnKF uses observation data from multiple time steps in the past.

The paper is well written and structured in a systematic manner. The results clearly support the findings and make an important contribution to the data assimilation literature.

Authors' response: We appreciate that Referee #1 sees our manuscript as well written, systematically structured and having important contribution to DA literature. We discuss and answer her/his comments in detail below.

However, my main concern is how to identify the specific contribution from the AEnKF approach. That is, are the improvements in the discharge forecast purely from using multiple data points in the past at the update stage?

Authors' response: Figure 7 shows the forecast accuracy of three different numbers of data points in the past (W=0, 5 and 11). This figure shows clearly that the differences in the forecast improvement of these various setups are purely due to using multiple data points in the past at the analysis step. We will stress this more clearly in the revised manuscript.

This question arises from the observation (in Figure 7) that the forecast improvements in the AEnKF are very systematic for most lead times. Given that the forecast improvements are systematic, I think it is safe to say that the improvements at the update stage are also systematic. If this is true, that means that there is a linear relationship between the updates made with a single observation data and those made with multiple data points in the past. Practically, this is unlikely because you expect different levels of updates especially for transition periods like dry to wet and vice-versa (or low to peak flows and vice-versa).

Authors' response: Results in Figure 7 show the average behavior (over many forecasts) of improved initial state on the forecast accuracy of the different filters. We agree that updates may vary in time. This is a valid approach, since data assimilation is expected to improve forecasts in average sense. What we see is in line with many other DA studies that show similar kind of behavior (not only in hydrology). That the improvements in forecast accuracy decay with lead time in a systematic fashion is to be expected. It would be very strange if they wouldn't, as the model slowly returns to the model climatology each time a forecast is made (no observations, so no updating). This is not a sign that a constant bias is present, but a sign that an update of the initial condition lasts for a while.

On a related note, the level of improvement from EnKF to AEnKF seemed to be relatively constant irrespective of the lead time (Figure 7). For me this is a worrying sign because it looks as if you are able to identify and quantify this constant difference (maybe call it bias) then you can get a better forecast. In other words, the AEnKF seemed to have a better treatment of bias. It will be interesting to know how the AEnKF will perform for low flows, normal flows, and peak flows independently. That is, will updating the model with multiple data points in the past always have a positive impact on forecasts made during low and peak flows?

Authors' response: We will state in the revised manuscript that for the high flows, the AEnKF with a longer time window W is able to make corrections that last longer on average (this is not a constant bias) with respect to the shorter time window W. Additionally, we will change the title to "Operational aspects of asynchronous filtering for flood forecasting", to ensure the presented study investigates high flows only. We agree with the reviewer that characterization of the statistical properties of the temporal flow dynamics (i.e. typical time scales of flood peaks as compared to low flows) is a relevant issue. The length of the time window W has to be seen relative to the time scale of the river flow dynamics. We assume that for low flow conditions, the improved skill of longer W with respect to shorter W will become negligible, as low flows exhibit less temporal dynamics than high flows.

I think these are important questions the authors need to address, even if they do not have results to support it at least a clarification is needed about which conditions their methodology will mostly apply.

In Figure 6, the no-update looked pretty accurate and almost comparable to the assimilation, but in Figure 7 the evaluation measures for no-update is very poor. This is a stunning difference, please clarify.

Authors' response: We agree that the no-update scenario matches the magnitude of the major peak quite well, although it has quite large spread. Additionally, when we consider the ensemble mean of the no-update scenario with respect to the DA scenarios, the accuracy deteriorates. This is shown in Figure 7 for the set of eight flood peaks, while Figure 6 shows only one individual peak.