Authors' reply to Reviewer 2
We thank the reviewer for their time in providing a review of our manuscript. We provide responses to each individual point below. For clarity, comments are given in italics, our responses in plain text.

This study demonstrates the use of a variant of the Ensemble Kalman Filter to assimilate streamflow measurements into a flow forecasting system. The topic is appropriate for the Hydrology and Earth System Sciences journal, but in some parts, the parts read like a user's manual (for example p. 9542, l. 2 or section 3.4); this is a scientific paper and should be written as such. The methodology followed is generally sound, although there are some issues that need to be addressed before publication (see below for details).

We are glad of the reviewer's generally positive assessment of our manuscript, and will address each point below. In our paper we tried to give a sufficiently full description of the method that other researchers would be able to reproduce the results. This is in agreement with the current public discussion of the importance of 'repeatability' in scientific results. However, we will revise and shorten some aspects of the method, particularly those highlighted by the reviewer, to improve the scientific tone of the paper.

- probably shorten the title by removing everything after "Filter".
We will shorten the title as suggested

- p. 9538 should be in an appendix. Perhaps adding a schematic for the REnKF would help.
We will move the algorithm to an appendix, and as recommended replace it with a schematic (thanks for this suggestion).

- It seems that the same observation is assimilated multiple times. Shouldn't the covariances developed by the model physics make re-running the model for each of the time steps before the observation time redundant? What does the pdf of the innovations look like? Can the authors add a comment on whether that pdf approximates a Gaussian since that will provide some insight on the optimality of the assimilation algorithm.
The reviewer is correct that we use the same observation multiple times as we iterate over the model states preceeding the observation (we will explain this more clearly in the revised paper and also provide a schematic as suggested). Although the model physics means that state updates are propagated forward in time, errors re-occur due to uncertainty in the model structure/parameters as well as in the precipitation and climate inputs. This results in update magnitudes which decrease with decreasing time before observation (see Fig 10) but are non-negligible at all times. We checked the distribution of updates (using the same data underlying Fig10) and found that distributions typically had the form of an unbiased bell-curve, but with fatter tails that a true Gaussian.

- p. 9539, l.12: I wouldn't use the term "directly", isn't the optimality of the filter governed mostly from the Gaussianity and the linear model assumptions?
Yes the reviewer is correct that the assumptions are also necessary for optimality of the filter. In the revised paper we will list the assumptions needed, and rephrase this sentence appropriately.

- p. 9540: the adequacy of the ensemble size depends on the state vector size. Probably, a 50-member ensemble should be good enough, but adding that information would be helpful.
As with the same comment from Reviewer 1, we will make the reason clearer, as the Clark et al., 2008 study cited found that 50 parameters was sufficient when using the same model structure in a catchment larger than any tested here.

- Section 2.2.2: why are the state variables perturbed along with precipitation? The correlations developed by the model itself will be physically consistent, however perturbing the state with additional noise could change the model error covariance significantly? Can the authors comment and demonstrate the consistency of the model states after perturbation?
The perturbation of state variables is a standard implementation of the EnKF, e.g. see discussion in section 3.2 of Liu et al. (2012). The ensemble of state variables should represent their prior distribution, so here we account for uncertainties in both inputs and model structure. An alternative is to perturb the model parameters. We also note that the perturbations are constructed to have physically realistic spatial and temporal covariances. It is possible, for example, that in a single subcatchment then water might be removed from the soil moisture store and added to the water table store, but this situation is not physically unrealistic as it could result from a situation where drainage was faster than model predictions. Similarly, a situation where water was added to both stores could result from incorrect precipitation measurement/forecast.

- p. 9541, l. 18: how were the parameters chosen for the other catchments?
The parameter choice is explained in Section 4.1. We will add a sentence to the paper at p. 9541 to direct the reader to this section.

- p. 9546, l. 2-3: why aren't the results from the first and second type simulation shown on Fig. 5?
We did not show all the results of 3 simulation types x 7 catchments as this created too many figures and made the paper excessively long. This will be even more the case in our revised paper as we consider a range of forecast lead times as suggested by the other reviewers. Instead we show the results of our method for all 7 catchments (Fig 6) and a comparison of the three types in Fig 7. We also summarise numerically the performance of all assimilation methods for all catchments in Tables 3 and 4.

- p. 9546, l. 9-15: how did each catchment's time of concentration relate to the chosen lag? Could that parameter have been estimated a priori (as was aforementioned in the text)?
We found experimentally that a lag time of 12 hours was suitable for a wide range of catchment sizes (i.e. catchment time of concentration smaller or greater than 12 hr), suggesting that optimal lag time might be controlled by catchment soil and geological characteristics rather than size.

- p. 9546, l. 16-17: why was the median chosen and not the mean of the ensemble?
We preferred to use the ensemble median as this statistic is robust to outlying ensemble members (especially important for the EnKF in which spikes could occur in some but not all ensemble members).

- p. 9546, l. 20-27: it's not very clear whether the percentage of times the flow measurement fell within the ensemble for the EnKF and REnKF refer to a posteriori ensembles. If so, why wouldn't those numbers be 100%, shouldn't the filter nudge the model towards the estimates (I suspect a reduced ensemble spread is the reason, but it should be verified by the authors).
The percentage of times that the flow measurement falls within the ensemble refers to the model running in forecast mode, i.e. assimilation has been completed up to the start time of the forecast, but the model ensemble is free-running during the forecast period. Hence it is not the *a posteriori* ensemble. We will explain this more clearly in the paper.

- Fig. 8: please add a legend at the bottom of the overall figure.
We will add this.

- p. 9548, l. 14-22: I'm having some difficulty understanding whether the change in the fractional error parameters improved the estimate because the filter did a better job or because the ensemble spread was wider and captured the observed streamflow values. - p. 9549, l. 10-12: the text seems to imply that a larger ensemble spread alone would improve the results, when in fact a better approximation of the spread would probably lead to better results.

We agree that both the size of the ensemble spread and its approximation of the model error are important. In cases where the spread does not capture the observed values, the spread is clearly too narrow and impedes the ability of the filter. Where increasing the fractional error parameter increases the spread and allows the spread to capture the observed value, and we believe this is the main mechanism for the improvement. However if the spread is already sufficiently wide, we would not expect an improvement. We will make these points more clearly in the revised paper.

- Section 4.4: I think the Ensemble Kalman Smoother might have been a more appropriate comparison.

We appreciate that there are a range of alternative data assimilation methods available to the hydrological community, including the Ensemble Kalman Smoother, particle filter, and variational methods as suggested by Dr Weerts. We will consider such methods as possible future directions, but in the case of this paper report on our existing choice of the EnKF with explicit consideration of catchment lag. We will comment more fully in the introduction regarding the alternative methods.