Anonymous referee #3

First, we would like to thank anonymous referee #3 for his/her interest in our paper and for the comments, which allowed us to improve the original version of the manuscript. We report below our replies (denoted by AC, Authors' Comments) to the referee's comments (denoted by RC, Referee's Comments).

Brief description of the modifications to the manuscript

The major modifications to our manuscript are

- the lagged regularized particle filters are improved in terms of kind of kernels, aggregation method of lagged weights and the computational speed; and
- the experimental design of the study is improved to compare forecasts for varying lead times and process noises and independent observations with additional events.

Major Comments:

RC: Although the authors describe well the theory and implementation of the lagged regularized particle filter, their results are either unsatisfactory or insufficient to support the conclusions. The authors claim, that "the accuracy of lagged RPF is higher than the other filters in the calibration period" (p. 3402). However, the difference between the Nash-Sutcliffe model efficiency of the SIR and the lagged RPF is only 0.05, which is negligible. Furthermore, the performance of the SIR and the lagged RPF is clearly equal for the validation period according to the Nash-Sutcliffe model efficiency. The authors also claim that "lagged filtering is evaluated to have more proper probabilistic bands, whereas SIR reproduced more diffuse probable density" (p. 3403). However, this finding is not supported neither by the predictive QQ plots (Fig. 10) nor the forecasts (Fig. 8). Fig. 8 only shows that the uncertainty bands are narrower for the lagged RPF in comparison to the SIR. This does not necessarily mean that they are more appropriate, as is shown in the predictive QQ plots. Fig. 10 shows that lagged RPF tends to underpredict the observations (most likely due to the narrower uncertainty bands!) whereas SIR tends to overestimate the predictive uncertainty. Hence, the only result which is shown by the authors is that both approaches (SIR & lagged RPF) have a similar overall performance (see Nash-Sutcliffe efficiency) but generate different problems for the predictive uncertainty. Finally, the authors claim that the lagged RPF preserves particle diversity better than the SIR. This is indeed proven by Fig. 11. Unfortunately, the authors mention only once (p.3398) that this gain comes with a 2.6 times higher computational cost than compared to the SIR. Hence, if I would double the amount of particles for the SIR approach, I would have a comparable computational cost and probably a similar or even better nratio and also maybe even better simulation results. Thus, *one could ask the question why implement a relatively complex approach (such as lagged RPF) when a simple increasing of particle numbers would most likely have the same effect (or even better)? The authors need to do a better job in justifying the use of the lagged RPF. The theoretical advantages of the lagged RPF are that it prevents particle filter degeneracy and sample impoverishment. For example, these problems become very relevant when a lot of error sources are included into the model (e.g., error in input forcing like precip, temp, evap; process errors; and parameter uncertainty). Hence, I suggest that the authors rethink their case studies in order to illustrate better what the advantages of the lagged RPF are and thus rewrite the chapter 5 and 6.*

AC: We thank you for your comment. Based on your suggestions, we reconstructed the experimental design and rewrote chapters 4 and 5 to illustrate the advantages of the proposed method. We have started this research because, first, we'd like to know what happens when particle filtering is applied to a distributed hydrologic model having heavy non-linearity and capability for distributed forcing data. The forecasts by conventional SIR show its basic implementation. Although forecasts by SIR were better than deterministic modelling results, SIR forecasts were prone to deteriorate because optimal parameters related with noise were different, according to simulation periods. Then, we planned to apply advanced particle filtering techniques such as the regularized particle filter. We also found that different time scales and lagged responses of a distributed hydrologic model reduce the accuracy and stability of forecasts by particle filter, and the regularized particle filter should be performed in the lagged time window.

In the revised manuscript, we reconstructed the experimental design, focusing on forecasts for varying lead times up to 24 hours. As a result, LRPF showed consistent forecasts regardless of the process noise assumption. Meanwhile, SIR had different optimal values and sensitive variations of confidential intervals for the process noise. Please refer to Figs. 12 and 13 for differences between the two filters. As shown in Fig. 10, in the case of SIR, confidential intervals of forecasts widened rapidly, and the ensemble mean became unstable when the value of the process noise increased. In comparison, the mean and confidential intervals of LRPF showed stable results despite changes of process noise. Improvement of LRPF forecasts compared to SIR was found, especially for rapidly varied high flows, as shown in Fig. 11. Preservation of sample diversity from the kernel appears to be one reason. In general, the accuracy of particle filters can be improved when the number of particles increases. However, statistical stability of the lagged regularized particle filter is hard to achieve in the conventional SIR. As mentioned at the end of the third paragraph in Section 4.4, LRPF was also improved in terms of computational speed, at 1.6 times higher than the privious version by reducing the number of synchronizations within the MPI code in the lagged filtering.

Additionally, analysis for predictive QQ plots and n_ratio was eliminated in the revised manuscript. We found that predictive QQ plots have limitations when one is comparing probabilistic forecasts having thick tales, which will be addressed in another article. In the case of LRPF, sample diversity can be obtained in the regularization step even if sample impoverishment occurs in the resampling step, as is newly illustrated in Fig. 2. If the sample impoverishment is very relevant due to a lot of error sources, forecasts via the SIR particle filter should be carefully checked or avoided. In the case of SIR, when this problem arises, a very limited number of particles survives, and forecasting may deteriorate severely at the important time point, usually around the high flow. Even worse, until the particle diversity is recovered, the reliability of forecasts for the adjacent time steps via collapsed particle diversity becomes very limited. In contrast, in LRPF, different particles are generated from the kernel and are evaluated in the regularization step, which is equivalent to the resampling step of SIR. As shown in Fig. 11, LRPF is recovered from sample impoverishment faster than SIR, which gives improved model efficiency in Fig. 12(b).

Although there are still limitations related with slight reductions of forecast accuracy in the small size of the flood event, LRPF has benefits for sequential data assimilation of a process-based distributed model in terms of the improved forecasts for rapidly varied high flows and the stability of confidential intervals, regardless of the process noise assumption.

Minor Comments:

RC: Please correct references Salaman and Feyen into Salamon & Feyen

AC: The error is corrected accordingly.

RC: Describe in more detail why you chose 0.05 for alphasoil and 50mm for betasoil. Are these typical error ranges reported in literature? Are those values based on a sensitivity analysis?

AC: In the revised manuscript, sensitivity analysis was implemented, as shown in Figs. 12 and 13. In the case of LRPF, differences of model accuracy for various lead times according to process noise were negligible when the flood peak was high and varied rapidly. However, SIR showed different forecast accuracy according to changes of the process noise, and optimum values were also different, according to simulation periods.

RC: Abstract: the abstract only contains a general description of what methods/models have been used in this work, and only in the very last sentence you mention something about the result. Please rewrite the abstract so that you write more about the results (which is supposedly the most important part of a paper) and only very briefly mention the methods and models applied.

AC: We thank you for this comment. We totally reconstructed the abstract to contain the major findings of this study.

RC: There are numerous errors in English grammar and syntax which I will not list here in detail. I strongly suggest that before resubmitting this work is proofread with a focus on English grammar and syntax.

AC: According to an English-speaking editor, we have made significant changes to the paper to make sure the quality of the grammar is acceptable.