Anonymous referee # 2

We are grateful to anonymous referee #2 for his/her helpful review and recommendations and suggestions. We report below our replies (denoted by AC, Authors' Comments) to referee's comments, followed by brief descriptions of the changes we have made.

Brief description of the modifications to the manuscript

The major modifications to our manuscript are:

- the lagged regularized particle filter 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 Changes/Questions/Concerns:

The authors assess the applicability of three particle filtering approaches (with increasing complexities) in streamflow simulation at one test basin by assimilating observed streamflow into a widely-used hydrologic model. I have several major comments:

1. The paper, as it stands, does not specify the generation of gridded (250 m resolution) model forcing and parameters as well as the connectivity between neighboring grids (flow direction in individual grids, the conflux of flow from different grids). Additionally, soil moisture states at different layers and different grids are perturbed using a same error definition (Equations 22 and 23) (so is the overland flow state). Those observations elicit a set of probing questions which require comprehensive answers:
a) Is this study really a distributed case or more like a lumped case? Are the paper subject and the context overstating the work and thus confusing the readers? Note that a distributed model can be applied at a lumped scale.

AC: This study is focused on the applicability of particle filtering for a distributed hydrologic model. And to our limited knowledge, this is the first trial to apply a fully spatially distributed hydrologic model for flood forecasting using particle filters except for the parameter estimation cases by Salamon and Feyen (2009, 2010). We also agree that a distributed model can be applied at a lumped scale. However, in this study, our concern is to determine what happens when particle filtering is applied to a fully distributed model (that will be forecasts by SIR in the manuscript) and what should be considered to improve forecasts in the short-term. We found that different time scales and lagged responses of a distributed hydrologic model may reduce the accuracy and stability of forecasts of particle filtering and suggested a lagged regularized particle filter as one candidate. Although there remains a gap to completing data assimilation for a distributed hydrologic model, the proposed approach is one advance in that direction.

b) What is the rationale of picking up only soil moisture and overland flow as model states to be updated? In such a humid basin, interflow and baseflow could significantly contribute to the flood volume.

AC: In the case of the Katsura catchment, located in the upland as one of tributaries of the Yodo river, the contribution of groundwater to the flood volume is relatively small, and the response time is short (4-8 hours). Therefore, only soil moisture and overland flow are selected as control state variables.

c) What if r_s in equation (22) is greater than 1 and thus produces unrealistic soil moisture (e.g., equation (23) provides soil moisture values greater than 1)? Applying a same multiplier to all layers and all girds is problematic (same for the overland flow).

AC: We thank the referee for this comment. There is no problem only if γ_s is greater than 1. A problem may arise when $\hat{\theta}_j^l$ becomes greater or smaller than the physical limitation. Therefore, if the perturbed soil moisture at each grid and layer, $\hat{\theta}_j^l$, reaches an unrealistic value, $\hat{\theta}_j^l$ is adjusted as its maximum (i.e., porosity) or minimum (i.e., wilting point). We added the description below Eq. (21). (Please note that the order of the equation was changed.)

d) Erroneous uncertainty definition (for soil moisture and overland flow states) very likely creates pseudo correlations between measured variables (to be assimilated) and model states (to be updated). As such, the update could be unfaithful. Should this be investigated and justified (e.g., at a small basin in a lumped fashion)?

AC: We thank the referee for this comment. We don't think the uncertainty definition of this study, adapted from previous studies, is erroneous. A similar noise definition for soil moisture has been successfully applied for state update of a distributed hydrologic model (Kim et al., 2003). As mentioned in the manuscript, we also adapted noise definition for overland flow from Seo et al. (2009). For investigation and justification, we compared forecasts via particle filters for varying lead times and the process noise, which will be discussed later.

e) The big size (1100/0.25ˆ2 = 17600 grids) of the test basin makes the problem more intractable. How the lag time associated with each grid can be reasonably defined (all grids sharing a single lag time is no better than treating the whole basin as one lumped grid)? Are the correlations between the outlet streamflow and the soil moisture states at three layers (and/or overland flow) of each grid physically sound or just purely statistical? Would assimilating *streamflow indeed improve estimates on current soil moisture/ overland flow and subsequently improve streamflow predictions in the future time?*

AC: We know lumping of a distributed hydrologic model has benefits for simplification and parameter estimation even in data assimilation. However, we think a fully distributed hydrologic model still has merits to explicitly deal with various distributed forcing data. Proper lag time for filtering should be estimated through sensitivity analysis for real data. We implemented sensitive analysis for varying lag times as shown in Fig 12(a). When lag time was larger than 4 hours, the difference of Nash-Sutcliffe efficiency (NSE) for the lag time became negligible. If we consider the correlation between the outlet streamflow and the soil moisture states at each grid, that should be statistically meaningful. However, statistically updated state variables can also contribute to improvement of streamflow predictions, which is one of achievements of this study.

As a matter of fact, many of above questions can be addressed by adding a simple study (e.g., via synthetic experiments at a smaller scale (1 grid), as conducted in the study of Weerts and El Serafy (2006) in the reference list).

AC: We thank you for your suggestion for a new experiment. However, the synthetic experiment at one grid scale is not as efficient for verifying lagged filtering. If we assume very small catchment (less than 10 km^2), the time lag from rainfall to runoff is negligible; therefore, we don't need to consider the lagged assimilation window. Instead, we modified the real experiments for extended data periods, varying lead times and various process noises, as shown in Table 1 and Figs. 7-13.

2. The experiment design of the study could be improved. As an example, selecting a calibration period and a validation period needs further clarification. Is the calibration period used to determine filter parameters (e.g., error configuration) and/or distributed model parameters (e.g., saturated hydraulic conductivity)? Either way, how the calibration is conducted (e.g., procedures, objective functions, algorithms etc.)? Why Fig. 10 and Fig. 11 present results from two different sub-periods of the calibration period (for the validation period as well)? What alternative periods could be selected as calibration period and validation period (the current periods representative?)?

AC: We thank you for your suggestion for experiment design. We thought about how to show the strengths and weaknesses of a proposed particle filter more generally. As a result, we decided to show all simulation cases for various process noises and lag times without dividing calibration periods separately. Sensitivity of lag time was analyzed using Nash-Sutcliffe efficiency, shown in Fig. 12(a), and sensitivity of process noise was analyzed in Figs. 12 and 13.

Another example is that the study focuses only on 2-h-lead predictions. The potential added value of DA to any operational forecasting system is that the technique could provide improved estimates on model states at the current time (forecast time) by assimilating the most recent observation(s). These "optimized" model states are expected to produce meaningful streamflow predictions with long lead time. In operations, the lead time for (typically called) "short-term forecasting" is up to 36 hours or even longer. If a DA technique only shows benefits for 2-h-lead predictions, for one, the technique would not be applicable in any operational environment; for two, most likely, the DA technique has significant flaws in science (refer to comment # 1). To be more specific about the second point, due to the fact that flow observations are assimilated so frequently (every hour) and the forecasting window is so short (2 hours), it is not surprising at all that the flow predicted highly resembles the observed. The flow prediction is actually dominated by the observed flow and not produced from the updated states at the forecast time. What if the update frequently is different (e.g., every 6 hours) and forecasting window is larger (e.g., 36 hours)? This study will make a high impact if it could show that these approaches used really improve estimates on model states which lead to improved flow predictions with a large lead time. The authors could refer to the study of Seo et al. (2003) (in the reference list) regarding flow predictions at various lead times. A further example is that the analysis only focuses on limited periods and events (e.g., Fig. 8). As such, the results are not generic (so are the conclusions). In operations, a reliable technique performs well in all events during any periods. Again, the authors are referred to Seo et al. (2003) paper on the analysis of multiple events.

AC: We thank you for your suggestion and the helpful reference. According to your suggestion, we reconstructed the experimental design to focus on forecasts for varying lead times up to 24 hours. We can confirm that updated state variables via two particle filters produce meaningful streamflow predictions that are sufficient for short-term forecasting, as shown in Figs. 12 and 13.

3. The manuscript, as is, is not self-explanatory in many ways. Examples, among others, include a) Equation (16) is incomplete. What if the denominator equals to zero?

AC: We thank you for your comment. We modified Eq. (16) to consider the case when the denominator equals zero. When the denominator equals zero, Eq. (16) will be 1.

b) /alpha is used in equations (16), (18), and 21)

AC: In the case of alpha, we provided different subscripts for each parameter and definitions below the equations. We also eliminated Eqs. (17) and (18) to simplify the lagging process.

c) S_k is not used but explained in equation (15). It appears in equation (22), though. Not all terms in equation (15) are defined.

AC: We thank you for your comment. It was a mistake to use the same character, S_k, in different equations. We changed S_k to L_k in Eq. (15) and added an explanation.

d) The authors should be cautious by stating "real-time streamflow forecasting" (L10, Page 3387). It requires forecasted model forcing produced from numerical weather models, which is not the case of the current study. The current study is more doing the "hindcasting" with all the forcing data and model output (streamflow) observed.

AC: We thank you for your comment. We modified the sentence accordingly.

e) A set of sentences/statements could be improved. Examples include, but not limited to: L1-2, Page 3384: "Applications. . ...used to. . ." L13-15, Page 3386 L24, Page 3386 and L18, Page 3403: "forecasting capability of streamflow" L3-6, Page 3390 L22, Page 3392 L24-25, Page 3393 L19, Page 3396 L6-7, Page 3404: any references?

AC: We thank you for your comment. We modified the sentence and added references according to your comment.

f) Table 1, Page 3409: Actually, the differences in NSE values of three methods are negligible. The flow predicted by these methods is tuned according to the observed flow assimilated into the model every hour, and thus resemble the observed flow more than the deterministic flow simulation (also refer to the 2nd example of comment # 2).

AC: We thank you for your comment. As mentioned above, we compared forecasts for various lead times to verify the applicability of particle filters. For time intervals of assimilation, we think a 12-hour or 1-day interval of assimilation can obviously improve computing efficiency in the drought season. However, since the main goal of this study was focused on flood forecasting, we kept the assimilation interval as one hour.

g) Rainfall is represented by bars (Fig. 7) and in lines (Figs. 8 and 9). Fig. 11, the validation period is not consistent with which defined in the context.

AC: We thank you for your detailed comment. In the revised manuscript, rainfall is represented by lines in all graphs. All statistics and hydrographs are drawn for the whole data period, except for Figs. 10 and 11, which are drawn for detailed comparison between two particle filters.

Reference:

Kim, S., Tachikawa, Y., and Takara, K.: Applying a recursive update algorithm to a distributed hydrologic model, J. Hydrol. Eng., ASCE, pp. 336-344, 2007.

Seo, D.-J., Cajina, L., Corby, R., and Howieson, T.: Automatic state updating for operational streamflow forecasting via variational data assimilation, J. Hydrol., 367, 255-275, 2009.

Salamon, P. and Feyen, L.: Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter, J. Hydrol., 376, 428-442, 2009.

Salamon, P. and Feyen, L.: Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation, Water Resour. Res., 46, W12501, doi: 10.1029/2009WR009022, 2010.