Dear Jeff,

I would like to thank you very much for your constructive criticism. I am confident that your comments will help us to improve the manuscript and I have no doubt that they will move the discussion forward.

It was our objective to propose a novel method for combining remote sensing-derived water level data with hydraulic models using the particle filter. To achieve this goal we adopted a two step approach. First, it was our intention to test and develop the method through a serious of synthetic experiments. Obviously, the results of our study relate to the assumptions we have made. We believe that the assumptions are consistent with the objectives of the paper (and the title). Although we believe that most of these assumptions are realistic, it is necessary to test and challenge these assumptions with real data. This second part of the study is under its way and we already realize that some assumptions may be questionable (e.g. forcing errors are not the only source of uncertainty in hydraulic modelling). As it was mentioned in the answers to the other reviewers, I agree that we need to put more efforts in clarifying and justifying the assumptions we made and in discussing the limitations with respect to real event data sets.

Comment 1: P1786, L4-5: What evidence is there that the temporal frequency of these SAR constellations is compatible with real time forecasting requirements and is this scale dependent? To my knowledge the repeat frequency used in this study is not about to become available.

Regarding the sampling rates of currently available SAR sensors, it is true that no single satellite provides water stage data with standard deviations of 0.3 m every 12-24 hours. However, existing (e.g. ALOS PALSAR, ENVISAT ASAR, COSMO Skymed constellation, Radarsat-2, TerraSAR-X) and future satellite constellations (e.g. Sentinel-1, Radarsat constellation) do provide coverage over Europe and many other parts of the world in less than two days. Hence, by combining data sets from different missions it is possible to get the kind of high-resolution data that we refer to in this paper. Sentinel-1 and Cosmo Skymed may even provide these sampling rated with a same sensor.

In an operational context, the frequency requirements clearly depend on catchment scale. Sampling rates of 24 hours seem to be acceptable for most (meso- and macro-scale) river basins affected by large scale flooding.

Comment 2: P1788, L19-23: "As a matter of fact, in order to be of relevance to flood forecasting systems, uncertainties associated with remote sensing data should be smaller than simulation uncertainties." I don't see why this is the case unless you are saying the measurement bias is larger than the simulation model, surely this is one of the main reasons for using a data assimilation approach. Perhaps this section on accuracy is less clear cut given that the reference (Arya et al., 1983) pre dates all the assimilation methods discussed in this section.

I realize that this statement is not correct. We will remove it from the re-submitted manuscript.

Comment 3: P1792, L7-8: I think this section needs more care if you have a non-Gaussian distribution from an ensemble and assimilate data with an EnKF you get a non-Gaussian posterior pdf. The clarification on this section seems to describe the Kalman Filter.

One of the assumptions of the Kalman filter is Gaussianity of the observation and model errors, which is frequently not met in practice. This can lead to a suboptimal functioning of the algorithm. In the particle filter, the assumption of Gaussianity is relaxed. We will clarify the differences between the KF and PF.

Comment 4: L11: "for a full representation of the probability distributions" What does this mean? This might be correct but I don't understand; obviously as an ensemble method it must be an approximation of the pdf.

This comment relates to comment 3. The Ensemble Kalman filter reduces the representation of the probability distributions to the standard deviation and the mean (thereby assuming a normal distribution). The particle filter, on the other hand, through a varying number of particles, represents the probability distributions more accurately. However, I agree that this is still not a "full representation" but rather an approximation of the pdf. The text will be modified accordingly.

Comment 4: P1793, L13-31: Is it necessary to assume all errors in measurement data are Gaussian and normally distributed? If the weighting method is based on Gaussian likelihood and the measurements assimilated have Gaussian errors it appears this method makes lots of assumptions about the pdf that are not obvious from the introduction.

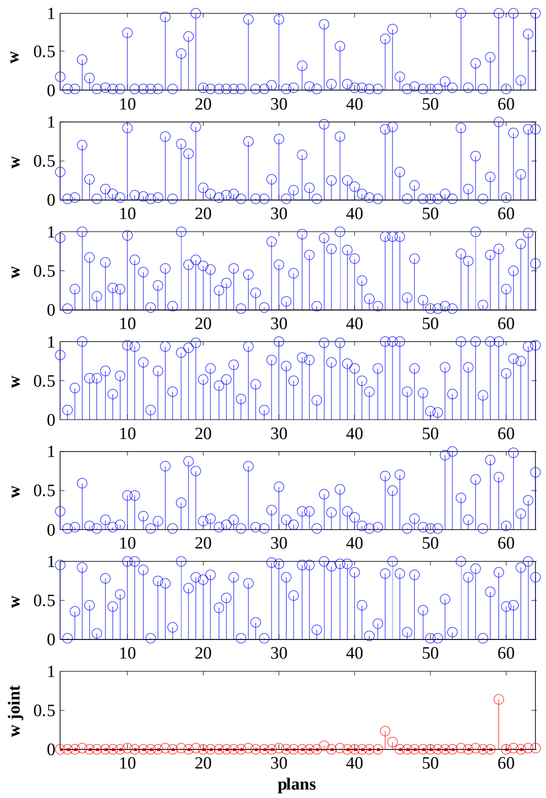
No, this is not necessary. One way to do the weighting of each particle is through the use of a Gaussian likelihood. However, any kind of likelihood function can be considered, which is a strength of the Particle Filter. For the sake of simplicity and because of a lack of information on more realistic pdfs of remotely sensed water stages, here we assume a Gaussian distribution. We will clarify the assumptions.

Comment 5: P1797, L9-11: Were there big differences in the error characteristics at high flows and what are the implications for the ensemble generation?

The verification measures were computed over a 6 month period. It was our objective that on average the ensemble spread differs from the observation by a value that is equal to the time average of the ensemble spread. Hence, we did not consider any differences between high flow and low flow periods. This means that for a particular flood event, the verification measures might not give a satisfying spread of state variables. The corresponding over- or under-estimation of model uncertainty could lead to an over- or underestimation of the improvement obtained through the assimilation of satellite data. There might be a significant impact on the results if model uncertainty is not represented in a realistic way. The adopted procedure for generating a meaningful ensemble in hydrologic modeling and data assimilation has been explained in Moradkhani et al. (2005) and Pauwels et al. (2009).

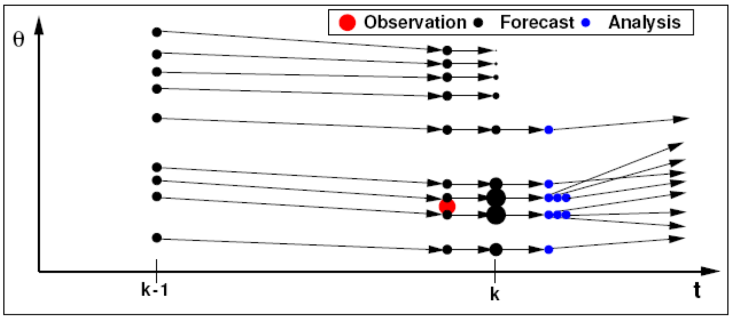
Comment 6: P1800, L2: Roughly how many particles are retained after assimilation and what was distribution of weights used... a typical example should suffice. From the figures it looks like very few particles are retained and it is not clear what the weights are for the forecast ensemble members that are subsequently simulated. How are the weightings considered in the calculation of the ensemble mean?

Here is a typical example:



The first 6 plots give the weights of each particle at 6 representative cross sections. The last plot gives the joint weight (equation 3 in the manuscript). In this case only three particles out of 64 were retained, with one particle carrying almost all the weight.

The SIR algorithm replicates the 3 particles in proportion to their weights. The resulting 64 particles will all have the same weight and the ensemble mean thus corresponds to the mean of the 64 re-sampled particles. The model is re-initialized with the water surface lines corresponding to the re-sampled particles. Due to differences in forcing data, the 64 particles will spread again. The following plot illustrates the functioning of the SIR algorithm.



Comment 7: L14-17: Are the image repeat frequencies considered realistic?

See also comment 1.
By combining data from different sensors (with different characteristics) it is possible to have water stage data every 12-48 hours. The upcoming Sentinel-1 constellation will provide data sets from the same sensor every 3 days. The considered repeat frequencies are realistic, but admittedly difficult to achieve at the moment. The main difficulty consists in having access to different data sets in near real-time data. In the future, however, the access to data in near real-time every 48 hours should be achievable.

Comment 8: P1801, L1-5: Is this correct? Surely each ensemble member should have its own Qerror and not be updated using an average? Also, as most of the ensemble members are re-sampled using sequential importance re-sampling how are the Qerrors related to the ensemble members in any sensible way? Should the flows not be updated individually and allowed to relax back towards the open loop values over a period of time?

We tested the alternative for inflow correction that you propose. This approach results in a conversion of all particles on the expected discharge value. In other terms, we would assume to have zero uncertainty at the end of the analysis step (the following time steps the spread would be to small) . To keep some uncertainty around the expectation, we updated each particle individually (Equation 11) but with a constant relative error term. Another alternative consists in correcting all particles in the same way using a same absolute error term. However, this can lead to negative inflows. Of the three variants that we tested, the proposed one gave the best results. We will describe the three options and discuss their advantages and limitations.

Comment 9: L 20: I'm not convinced by the use of "best estimate" here as there is no justification for the method of propagating model errors except that it is simple.

We cannot prove that the proposed error estimate is the "best estimate" if no discharge data is available for calibrating, for instance, auto-regressive models. Our experience, however, shows that relative errors tend to remain constant during periods of recessions. Many studies show that most river basins act like linear reservoirs (in the absence of rainfall). If this is the case and if the model error is due to an over- or underestimation of the water storage, the relative error would remain constant. I don't agree with the statement that there is no justification for this model (P1801 L20-25). I do agree with you that we cannot back-up the statement that this is the "best estimate". The text will be changed accordingly. We are not aware of any better error propagation model (unless there is training data available).

See also comment 16.

Comment 10: P1802, L1-5: I think you should have a non-normal test case.

We applied the filter on water stages that are not normally distributed. The results in a synthetic experiment are very similar as the particle filter easily adapts to any kind of probability distribution function. In this paper we assume that the water stages are normally distributed because there is no clear evidence on the type of non-normal distribution to consider. We are currently working on a real event data sets with empirical pdfs. and would prefer to put the non-normal test case in this framework. We will add a discussion on the consequence of non-normally distributed observations to the manuscript.

Comment 11: L6- : So the proposed approach has updated the states to represent current conditions then forces those conditions with the same ensemble of flows used before the assimilation, but shifted by the average difference between the prior and posterior ensemble mean. Meaning the ensemble spread rapidly returns to the original variance but with different mean? This is OK for the theoretical test case but is this likely in reality?

Each particle is updated individually. Hence, the term "shifting" may be misleading. It is true that with our approach the ensemble spread returns to the original variance very quickly. This is what we wanted to achieve since the updating should not influence the model errors on the long term but only remove bias due to temporary over- and underestimations of water storages. We tested also two alternative approaches (see comment 8) but the results were less good.

Comment 12: L26: This is rather obvious. The main question is how assimilating the measurements impacts on forecasts of the near future?

Assimilating measurements has positive impacts on forecast of the near future. However, the increased frequency of assimilations has only a small impact during the recession periods (which proves that the inflow correction method is able to remove systematic errors). During the rising limb the frequency needs to be higher because the relative model error does not remain constant. In our opinion, this is not an obvious result. It demonstrates one of the main shortcomings of the proposed method for updating the forcing data. We conclude that the sampling rate needs to be reasonably high to accept the underlying assumption of relative model errors remaining constant. We notice that performances tend to improve with higher sampling rates, more precise observations and with assimilation schemes that adjust the fluxes at the upstream boundary. See also comment 14. We would prefer to calibrate a more powerful auto-regressive model, but if the necessary discharge data is available there would be (in our opinion) no need to update the hydrologic model with discharge estimates obtained via remote sensing observations. In this case it would be better to use the measured discharge directly.

Comment 13: P1803, L8: This is true but you have not assessed the temporal correlation in model errors, instead you have applied a persistent shift in the mean to the boundary.

We assume that the relative errors remain constant. We further assume that we have no discharge data that would allow us to assess the temporal correlation in model errors.

Comment 14: L10: Again obvious, but what about the forecasts?

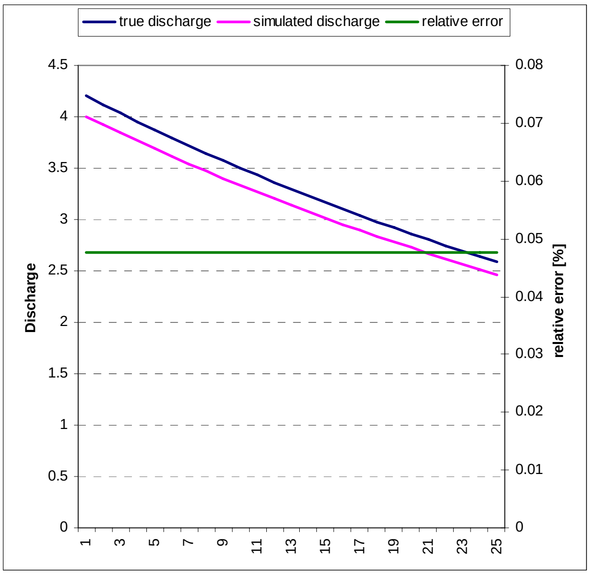
The RMSE is computed over the entire time window, with the exception of the time steps of the analysis that are not included. Only the forecasts are considered. Hence, the reduction of the RMSE shows that the assimilation improves the forecasts. We will clarify the way the RMSE is calculated. We believe that this result is not obvious.

Comment 15: L12 Discussion: "Our study demonstrates that the information contained in radar flood images can lead to improved flood inundation modeling". This is already reasonably well established in the published literature, whilst the study cannot be used as additional evidence to confirm this point because the data are simulated using an error model that is not typical of SAR imagery.

We agree. This statement will be reformulated. In general, we will be more careful with respect to the conclusions. We would like to mention that the assumptions remain realistic with respect to the proposed SWOT mission (i.e. normally distributed water stages, spatially non correlated errors). I agree that they appear unrealistic with respect to water stages obtained through a fusion of DEMs and SAR-derived flood boundaries. We further believe that the assumptions are consistent with the objectives of the paper (and the title).

Comment 16: L8-11: Your error model doesn't do this, the boundary flows are only adjusted to change the mean and as far as I can work out (so clarification may avoid this problem) the flows have little relation to the water levels. I don't see how the proposed error forecast model is particularly well suited to prediction in ungaged basins because it implies a scalar shift in discharge is appropriate at all forecast times. To make this conclusion you must demonstrate that it is appropriate; obviously in this case where a constant bias is added to the boundary the proposed method will do alright in terms of forecasting the mean. This and conclusion 1 are rather large overstatements and need to be toned down.

I agree. The formulation is not correct since there is no regression involved. However, we still believe that the proposed inflow correction is appropriate (see also comment 9). The assumption is that relative errors remain constant, which is true if the catchment behaves like a linear reservoir and storage errors are the only source of error. These storage errors are mainly the result of under- or overestimating rainfall amounts. The conclusion will be toned down.



Many catchments are known to act as linear reservoirs (e.g. Fenicia et al., 2006)*. Hence, we believe that the proposed inflow correction model may be used in ungauged catchments to estimate errors unless calibration data for more sophisticated models are available.

I don't understand your comment that "the flows have little relation to the water levels". The expected discharge value in Equation 10 is obtained through the stage-discharge relationship of the upstream cross section. Also I would like to mention that each particle is updated individually.

Obviously, we need to improve the description of the proposed error correction model.

*Fenicia, F., Savenije, H.H.G., Matgen, P., & Pfister L., 2006. Is the groundwater reservoir linear ? Learning from data in hydrological modelling. Hydrology and Earth System Sciences, 10 : 139-150.

Comment 17: L21: "closes the overall water balance" what does this mean? It seems unlikely that you have closed the catchment water balance. Do you actually mean that the hydraulic model update conserves mass for each ensemble member?

Yes, you are right. We will clarify this statement.

Comment 18: L22:25: It is a bit difficult to understand what the message of this sentence is. The Andreadis example also estimates boundary discharge (but uses a well established autoregressive model for the boundary errors). It also reduces uncertainty in the boundary discharge which is not the case here, although the PF can do this I assume? As for parameter estimation this is not implemented and no discussion of how this would be done is included.

We believe that the PF has some advantages over the EnKF because the updates conserve the mass for each particle. The Kalman filter calculates an innovation (difference between observation and simulation), and maps this onto the state space. Consequently, the value of the state variables changes in the analysis (assimilation) step.

This means that with the PF we can infer the input data and model parameter sets associated with the re-sampled particles. This comment should be considered as an outlook rather than as a conclusion and we have to make this clear. It is also worth mentioning that the Andreadis study assumes that discharge data is available to regress future error values against current model error. We believe that if satellite data are to be used in the future to routinely update hydrodynamic models, it is fair to assume that no *in situ* data are available. Hence, we cannot assume that there will be enough data available for using well established auto-regressive models. If such data are available, satellite data will probably not provide useful information in an operational context and there is no reason to update coupled hydrologic-hydraulic models *via* stage data obtained from remote sensing observations.

The usefulness of combining *in situ* hydrometric station data and remote sensing observations needs to be investigated in future studies.

Again, I thank you very much for your very helpful comments. I hope that I could give you a satisfying answer to all your comments. Don't hesitate to contact me if you need any further information.

Sincerely,

Patrick Matgen