Reply to the reviewer comments RC1: 'referee comments', by Anonymous Referee #1

We would like to thank the Anonymous Referee #1 for the detailed review of our paper and the many constructive comments. In the following, we answered each comment individually. The reviewer comments are printed in gray italic font and our replies, in black.

Specific comments:

1- Line 205-206: Here it is written that three additional values of D were used, which corresponds to the spread factors of 0.01%, 1%, 2.25% and 4% (4 values). It should be integrated with the previous sentence or explained better that the first value of 0.01% is from the D value of 0.01.

Reply: Thank you for the remark. We will clarify this in the revised version.

2- Figure 2: The forecasts for the UnE and the UnD are eerily similar, which is explained later in the discussion for other aspects but the discussion should refer to figure 2 to show how the forecasts are similar by reducing the high peaks (Figure 2 is not discussed after line 219). Making this link near lines 366-370, lines 379-382, lines 443-445, etc.

Reply: We agree with the reviewer and we will modify accordingly in the revised version.

3- Lines 220-226: Could the LP solver be cited? Is it a commercial software such as XPRESS, CPLEX, GUROBI, etc.? or is it in-house? Perhaps open source?

Reply: The linear problem is solved by a third party library COIN Clp, which applies the simplex algorithm. The COIN Clp solver is called using a Python interface, called PuLP. We will add this information in the revised version.

4- Line 239-240: Here, do the authors mean that the ensemble mean of the streamflow is fed to the solver (I think this is the case)? Or is the solver run on all members and the average decision taken? The latter could be interesting also to consider non-linearities in the ensemble forecasts. If the former is true, which I think it is, then I wonder why bother to generate ensemble streamflow forecasts at all? Why not simply generate deterministic forecasts with the desired properties (slightly biased, over/under -stimated, etc.?) under-dispersion, if it is symmetric, would be the same as the unbiased system, as seen in many figures. I understand it is not necessarily symmetric but the method employed (skewness of the log-normal distribution) could be corrected by another distribution: Basically, I think the authors should justify the use of an ensemblebased methodology if the solver only uses the mean value and the rolling-horizons test-bed only uses the ensemble average as well. Any difference between the unbiased and under-dispersed values would be due to the generation of artificial streamflow. If the authors used real-world streamflows (or simulations of streamflow using observed/forecast weather) then this point would not bother me as much:

Reply: The solver is run with the ensemble mean in this paper. We used an ensemble of generated streamflows to approach the (real-world) ensemble operational forecasts used to recover the temporal evolution (ECC approach) and to retrieve ensemble characteristics (such as the under-dispersion generated with the UnD system) that are sometimes found in practice (as shown in Fig. 4). It should also be noted that we also investigated running the solver with each ensemble member (optimization step) and then averaging the decisions. We agree with the reviewer that this is also an interesting question to investigate. This is however the topic of another paper under preparation, where we used operational forecasts of different quality (with/without meteorological and/or hydrological post-processing) to feed the solver. With operational forecasts, we were also able to investigate the influence of the evolution of the quality of the forecasts with lead time, which we do not investigate in this paper since the quality (the bias) is identical at each lead time. This paper would be too long if we had also included these aspects. Therefore, we opted to separate the studies: while this paper focuses on the influence of systematic bias on economic value, our paper under preparation focuses on the influence of the 'mean versus all members' approach on the output of the reservoir management model.

5- Line 244-245: This means that there is no terminal water value, correct? i.e. the system would want to empty the reservoir at the end of the period to maximize benefits IF there was no constraint on drawdown volumes equal to the expected inflows? Perhaps the authors could comment on the impacts of this, as there would be no consideration of marginal head gains.

Reply: The reviewer understands it correctly; there is no the terminal water value. As also noted by Reviewer#2, if there is no accounting for the state of the reservoir storage at the end of the period, the management model would normally want to empty the reservoir at the end of the period to maximize benefits. In fact, in our study, we implemented a weekly production as a 'soft constraint', which should not be higher than the weekly volume of water entering the reservoir. This prevents the reservoir from being emptied. This is an important issue pointed out by both reviewers and we will make it clear in the revised version.

6- The fact that the optimization method is deterministic should introduce a deterministic bias, by which the optimization method does not account for uncertainty and thus is over-confident that it can maintain high-head without spilling (Philbrick and Kitandis 1999). Therefore slightly increasing the bias "tricks" the model into thinking there will be more water, forcing it to produce more energy and thus counteracting it's overconfidence on high water levels. In a real-world scenario, this would be observed as the reservoir head would increase efficiency and entice the optimization algorithm to maximize revenues this way. But the setup in this study uses constant efficiency (line 252) with no change caused by reservoir head, making it much less representative of an actual system. The optimization has no need to optimize water levels, just make sure the reservoir doesn't crash or overtop. I think the authors need to add a section to the discussion to highlight these differences, because the way the abstract, discussion and conclusion are written, it could seem like the authors are saying that overestimation of forecasts is the worst possible solution, whereas in the real world it is probably the best option if using a deterministic optimization algorithm to counteract the optimization deterministic bias. I think the presented results would be much more in line with what would be seen if the optimization algorithm where stochastic (SDP or in the same vein), as the uncertainty is inherently included, which is not the case for deterministic solvers. [Philbrick, C. R. and Kitandis, P. K.: Limitations of Deterministic Optimization Applied to Reservoir Operations, J. Water Res. Pl.-ASCE, 125, 135–142, https://doi.org/10.1061/(ASCE)0733-9496(1999)125:3(135), 1999.]

Reply: This is a very relevant remark and we thank the reviewer for drawing attention to it. We will include a discussion on this issue in the revised version.

7- For the soft constraints, is there a penalty term in the cost function? Or is it considered as a hard constraint during the solving and then dealt with during the simulation part? If there is a penalty, could the information be provided?

Reply: There are penalty terms in the cost function associated with spills (maximum volume) and minimum volumes. They only intervene in the objective function of the LP, and not in the simulation part or the economic valuation (evaluation of gains). Penalties are based on the order of magnitude of the gains per hm3 (taking the maximum electricity price into account). The minimum volume penalty is calculated to always be greater than the potential gains, and the spill penalty 10 times the minimum volume penalty. We thank the reviewer for pointing this out and will provide the information in the revised version.

8- Could the authors perhaps give an example (figure?) of the impact of increase in dispersion/spread using the D2 factor, compared to the effect of the under-dispersion? How does this affect the ensemble mean? Perhaps an example with a random date would help understand these impacts and differences (as the increase in spread is counteracted by the process of under-dispersion).

Reply: The figure below represents the evolution of the mean of the synthetic ensemble forecast for the under-dispersed forecasting system and for the different spread factors (fig. 1R). We can see that the stronger the spread factor is, the further the mean of the ensemble will be from the observed value. In other words, the higher the spread factor, the more the high flows will be underestimated and the low flows overestimated.

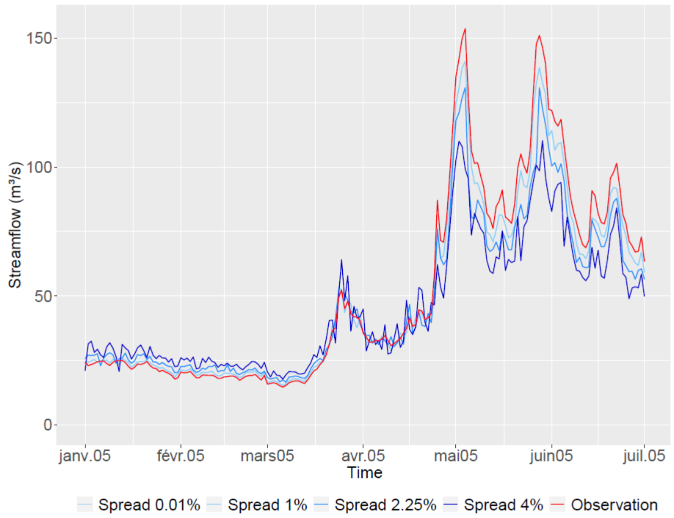
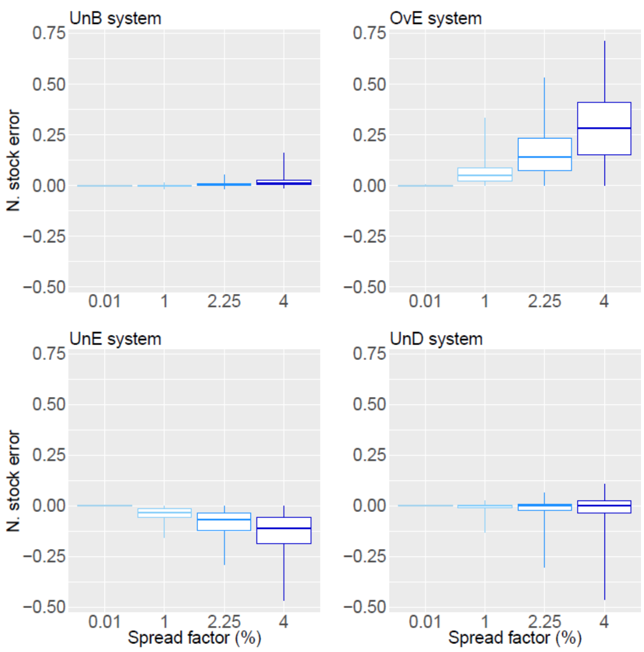


Figure 1R: Evolution of the mean of the synthetic ensemble forecast for the under-dispersed forecasting system and the different spread factors. Each color represents a spread factor value, ranging from light blue (spread factor = 0.01%) to dark blue (spread factor = 4%). The red line represents the observed flows.

9- Figure 13: I think this figure could be changed to a 2x2 panel figure as in the previous figures, to keep things simpler and cleaner.

Reply: We thank the reviewer for the suggestion. We will modify the revised version with the figure below (caption remains the same):



Typo/precisions needed:

1- Line 17: "approximately 3% to 1% (in MC)" Here, I suppose the authors mean that the value of 1-3% represents Millions of Euros, however the way I read it is as if the units were MC which would be moot as the difference is relative. I suggest writing it as "which represents millions of Euros"

We will consider the suggestion in the revised version.

2- Line 28: squared kilometers ! square kilometers

We will correct it in the revised version.

3- Line 281: non-respect ! violation

We will correct it in the revised version.

4- Lines 487-492: These few sentences were quite confusing to read. I think they are technically correct, but reading them and parsing the information was somewhat difficult. Perhaps separating into a few more sentences and clarifying? Especially for the last sentence (lines 490-492).

We will consider the suggestion and clarify it in the revised version.