Reply to the reviewer comments RC2: 'referee comments', by Anonymous Referee #2

We would like to thank the Anonymous Referee #2 for the detailed review of our paper and the constructive comments. In the following, we answered each comment individually. The reviewer comments are printed in grey italic font and our replies, in black font.

Specific comments:

1- In general, I would say just a little more effort might be made to help the reader under-stand where this work falls in terms of a practical perspective. The comments below will go some of the way to achieving this.
As far as I understood, the synthetic forecasts are not compared against historic forecasts that have been made nor against a real forecasting system. I understand this comparison would be outside the scope of the paper. However, have any of the cited authors who work with synthetic forecasts made this comparison? It seems to me that understanding whether the biases in synthetic forecasts reflect the types of biases in real forecasts is important information for a reader to know regarding the practical value of synthetic forecast-based studies.

Reply: The reviewer understood it correctly. The synthetic forecasts are compared against observed flows to verify the technique used to create different biases in the forecasts. Forecasts from a real forecasting system are only used to apply the ECC method to retrieve temporal correlation, as presented in lines 179-188. These forecasts are based on operational models. We have not evaluated their quality in this paper (as the reviewer mentioned, it is out of the scope). However, as reference, the reader can find scores of forecast evaluation for a similar dataset in the paper by Zalachori et al. (2012), which focused on evaluating the impact of bias correction techniques applied to precipitation and to streamflow forecasts. This study shows that raw (without bias correction) operational streamflow ensemble forecasts display biases and under-dispersion, notably for the shorter lead times, with mean normalized RMSE values ranging from 1.7 to 2.4. These values are of the order of magnitude of the over-estimation biases introduced in our synthetic forecasts (Fig. 6, top).

Concerning the link between the quality of synthetically generated forecasts and the quality of forecasts used in practice in the studies we cited in our paper, we note the following:

- Lamontagne and Stedinger (2018) do emphasize that "Regardless of how they are generated, synthetic forecasts should replicate the important statistical properties of the real forecasts or the specified properties of a potential forecast product", referring to statistical properties in terms of mean, variance and accuracy. This is the basis for the conceptualization of their forecast generation approach. Their application is however illustrative. It supposes that a given forecast product has a given R2 and then generates synthetic time series of flows based on this information.

- Maurer and Lettenmaier (2004) worked with a measure of predictability developed in their previous studies as indication of potential seasonal predictability (fractional runoff variance explained by given predictors). They added errors to observed flows that are normally distributed with a mean of zero and a variance that is a function of this measure of predictability. Therefore, the comparison against practice relies on the fact that they use a range of values for this measure of predictability that are reported by the forecast skill evaluation study presented in Maurer and Lettenmaier (2003).

- Arsenault and Côté (2019) used a proxy for observed streamflow derived from the hydrological model initialized with empty reservoirs and driven by observed climate data (ESP-type forecasts). They then add bias to the ensemble means by multiplying the ESP forecast members by a factor that allows to shift the distribution upwards (factor >1) or downwards (factor <1). The factors chosen resulted in bias that ranged from −7 % to +7 %. As for the relation to biases seen in operational practice, the authors only mention that: "Larger values were excluded because they were not necessary for exploring the behaviour of biases on the hydropower system operation".

We fully agree with the reviewer that understanding how much the biases in the synthetic forecasts reflect the actual biases in operational forecasts is important regarding the practical value of the conclusions drawn from studies based on synthetic forecasts. It should also be noted that, in operational forecasts, bias may vary according to the time of the year, the physical phenomenon causing high or low flows, and the catchment under consideration. Additionally, biases are often dependent on lead time, which is not the case in our synthetic forecasts. In the same context of the work we present in Zalachori et al. (2012), we observed that in some catchments the hydrometeorological forecasts (forecasts issued in the period 2005-2008) tend to under-estimate the flows, while in other catchments, the tendency is towards over-estimation (PhD Thesis I. Zalachori, 2013). The fact that we introduce the same bias over all time steps and catchments in the synthetic forecasts of our study makes it challenging to compare the synthetic forecasts with actual forecasts. Despite these limitations, we note that we have created the scenario of under-dispersed ensembles in our synthetic generation, where high flows are underestimated and low flows are overestimated (UnD synthetic forecasting system), to capture a feature that is often observed in operational hydrological forecasts due to the difficulties to capture peak flows in extreme events and to model highly-influenced low flows. We will make this important point clearer in the revised version of the paper.

Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T., Gailhard, J., 2012. Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. Advances in Science & Research*, vol. 8, p. 135 – 141. doi:10.5194/asr-8-135-2012)*

Zalachori, I., 2013. Prévisions hydrologiques d'ensemble : développements pour améliorer la qualité des prévisions et estimer leur utilité. Thèse de Doctorat, Irstea (Antony), AgroParisTech (Paris), 398 pp. [In French]

Maurer, E. P., and D. P. Lettenmaier, 2003: Predictability of seasonal runoff in the Mississippi River basin. J. Geophys. Res., 108, 8607, doi:10.1029/2002JD002555

2-L234/5: This seems a rather arbitrary choice for reservoir storage capacity. In fact, I am left slightly confused as to whether the reservoirs under study are real or not (I think they are not, but '10 reservoirs in France' is used in the abstract). If they are not, that needs to be explained more clearly. If they are real, then surely real capacities can be identified?

Reply: The reviewer again understood it correctly. The reservoirs are not real reservoirs. We use actual observed streamflows of the 10 catchments that provide the inflows to the reservoirs, but we do not use the actual reservoir dimensions and operational characteristics. This will be explained more clearly in the revised version.

3-I was not clear from the explanation in 2.3, is there some accounting for the state of the reservoir storages at the end of the 7-day optimization period? It is OK if there is not, but the authors should note this, as it may lead to reservoirs becoming overdrawn in the long run.

Reply: It is well noted that if there is no accounting for the state of the reservoir storage at the end of the period, the management model would want to empty the reservoir at the end of the period to maximize benefits. In this study, we do not use water value (future benefits) to constraint the storage level at the end of the 7-day, but instead 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 also pointed out by Reviewer#1 and we will make it clear in the revised version.

Editorial comments:

L35/36: The word ’interesting’ is an unusual choice, I would suggest ’beneficial’

We will modify it in the revised version.

L38: Grammar should be ’..within integrated river basin management..’

We will correct it in the revised version.

L40: You can’t have the ’most’ optimal. Simply use the word ’optimal’, or if you want to avoid the implications that releases are truly optimal, then use the word ’best’.

We will correct it to ‘optimal’ in the revised version.

L44: Personally I would extend ’linear programming’ to ’linear and nonlinear programming’ since there are many nonlinear approaches (see any of the cited reviews for references) here that could not be said to fall under the term dynamic/heuristic programming.

We will modify it in the revised version.

Figures 11 & 12 - why switch to number of hours instead of % difference as in the other plots?

The aim is to give also the order of magnitude of the differences.