Response to Interactive comment on "State Updating and Calibration Period Selection to Improve Dynamic Monthly Streamflow Forecasts for a Wetland Management Application" *by* Matthew S. Gibbs et al.

Anonymous Referee #2

General:

The manuscript of Gibbs et al. evaluates the effect of calibration setup on the GR4J model performance within 2 Australian basins on 1-month lead-time hydrologic forecast. Authors draw mainly following conclusions based on 2-basin analysis using 5 indicators: (I) the length of calibration period does not necessarily should be as long as possible, in particular when changes in flow regimes are observed. Additionally, (II) the authors state that a simple model state updating improves hydrological forecast at 1-month lead time.

Based on my review, I consider the overall topic to be relevant for HESS, however, some parts of the manuscript need to improved and clarified, as further suggested below.

Thank you for the constructive comments that will improve the clarity and contribution of the manuscript.

Major comments:

1. In general, it is not surprising that updating model initial conditions has benefits on hydrologic forecast. Additionally, it is not surprising that changes of physiographic conditions may change the catchment's response, indicating different information content/validity of observed discharge data on model parameters. This only confirms observations of previous studies, which some of them are cited. Would be nice to more clearly demonstrate benefits over existing/operational approaches (in terms of costs etc). In particular, when the title includes words like "Management Application".

Commentary on the benefits over the existing approach will be added to the Discussion section. The approach currently used by the management authority is very conservative. Essentially forecasts are not considered and changes in water management are made only once downstream requirements have been met. With the forecasting models and methods developed, it is now possible to provide a probabilistic estimate of how likely it is that the downstream requirements will be met in the next month, to allow the risk in changing management actions earlier to be assessed. This is expected to improve the outcomes achieved from the volume of fresh water available, by enabling decisions on management changes to be made earlier in the season.

The Discussion section will be expanded to contrast the findings in this work against the relevant references cited (Brigode et al., 2013; Luo et al., 2012), to highlight where there are similar or different findings.

2. Unfortunately, the analysis is limited to two basins, which really can't be used to draw any conclusions (as authors also recognise in the end). I would strongly encourage authors to enlarge the number of basins and events. Another two basins may yield completely different results; therefore, generality should be avoided.

It is agreed that general results cannot be drawn from the results based on two basins. The drainage system considered was based on a user need for seasonal forecasts, which resulted in the limited application to two basins. Although the procedure has not been tested on a wide range of basins, the catchments selected are ephemeral and therefore are known to be more challenging to provide good quality forecasts (Ye et al., 1997). So even though only two basins have been considered, they are expected to be 'challenging' basins.

Future work will consider if, and under what conditions, the results for state updating at the monthly scale and effect of the length of the calibration period apply, beyond the case study considered. This point, on the limited generality of the results and need for further work on more basins, will be included in the Discussion section.

3. Would be nice to relate your results with another study, which details CRR state updating for 1-month forecast. However, I wonder, whether it is really the effect of state updating here. It is well recognized that the effect of initial conditions (based on discharge observations), in such small basins, diminishes after a couple of days. Please, comment on this.

The result that model updating still improved forecast performance at the lead-time of one month is considered one of the key results of this work. The controlled study for testing the model with and without state updating for all the cases considered (observed and forecast rainfall, two lengths for the calibration period and two basins), demonstrates that for the scenarios considered the effect can be attributed to state updating. The point at Comment 2, on the potentially limited generality of this result, is relevant and will be included.

It is agreed that the majority of the literature focused on short term flood forecasting has found limited effect of initial conditions after a few days (e.g. Berthet et al., 2009). However, forecasting of flood peak and timing is a different application to that considered here, forecasting water volume available for environmental use. A number of data driven modelling studies have demonstrated that monthly streamflow lagged by one (or more) months provided some useful information for forecasting a one month lead time (e.g. Bennett et al., 2014; Humphrey et al., 2016; Yang et al., 2017). This difference in application, and relevant literature, will be included in the introduction of the revised manuscript.

The description of basin and data is way too long and detailed (3 pages), in particular when the discussion and conclusion do not come to those details at all.

Reviewer 1 raised a similar issue, and as such Section 2 will be shortened to be more targeted and remove surplus detail, such as that of the water balance model. One of the topics of interest for the sub-seasonal to seasonal hydrological forecasting special issue was user needs for seasonal forecasts. As such, more detail than typical on the case study application to wetland management was included in the manuscript. However, it is agreed that this is a distraction from the more generic scientific issues of interest to most readers. As outlined at Comment 1, commentary on the benefits over the existing approach will be added to the Discussion section.

4. Please, place error bars into figure 4, in the same way as the uncertainty is presented in following figures and provide discussion.

The uncertainty in the streamflow time series had been produced and assessed (e.g Figures 4 & 5) because it is a direct of the output of the modelling setup. However, all of the streamflow replicates are used in the calculation of the performance metrics (reliability, precision and CRPS in particular), and as such the uncertainty in the values of the metrics is not easily derived. To the authors' knowledge, it is no common to provide the uncertainty in performance metrics related to streamflow forecast uncertainty (e.g. Alfieri et al., 2014; McInerney et al., 2017; Pappenberger et al., 2015; Renard et al., 2010).

We do acknowledge that there is uncertainty in the values for the metrics, due to finite number of replicates and the finite length of the observed data (McInerney et al., 2017) and that in the presence of strong autocorrelation in the streamflow error time series this uncertainty may be significant. However, to develop an approach to quantify the metric uncertainties is beyond the scope of this study. Instead, we follow the guidance provided by McInerney et al. (2017) who suggested that these uncertainties are quite small and are unlikely to impact on the conclusions of the study. We do acknowledge that further work is needed in this area.

6. Discussion about alternative types of observations (besides Q) may be provided in the manuscript.

The literature review will be expanded to refer to previous studies that have used other types of observations that could be used for state updating (e.g. either remotely sensed or ground based soil moisture), and why streamflow was focused on in this study.

7. What is the main applicability of your findings? Are they going to be used operational, if yes, what are the benefits over existing forecast method? Please, clarify.

See response to Comment 1.

Minor:

- Third sentence from the Introduction regarding Drain M catchment should be moved somewhere towards the end of Introduction.

This change will be made.

- P 6, L.5: evapoconcentration => evapotranspiration?

This typo will be corrected

- Section name 2.2 "Streamflow and streamflow data": sounds a bit repetitive

The section will be changed to "Streamflow data"

- P 9, L.21 "burn-in" into quotes

This change will be made.

- Eq. 12 is wrong, sum in the denominator is missing

Correct, the equation will be updated.

- Caption of figure 2: "POAMA" => "POAMA-2"

This change will be made.

References

Alfieri, L., Pappenberger, F., Wetterhall, F., Haiden, T., Richardson, D., and Salamon, P.: Evaluation of ensemble streamflow predictions in Europe, Journal of Hydrology, 517, 913-922, 2014.

Bennett, J. C., Wang, Q. J., Pokhrel, P., and Robertson, D. E.: The challenge of forecasting high streamflows 1 & 3 months in advance with lagged climate indices in southeast Australia, Nat. Hazards Earth Syst. Sci., 14, 219-233, 2014.

Berthet, L., Andreassian, V., Perrin, C., and Javelle, P.: How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments, Hydrology and Earth System Sciences, 13, 819-831, 2009.

Brigode, P., Oudin, L., and Perrin, C.: Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change?, Journal of Hydrology, 476, 410-425, 2013.

Humphrey, G. B., Gibbs, M. S., Dandy, G. C., and Maier, H. R.: A hybrid approach to monthly streamflow forecasting: Integrating hydrological model outputs into a Bayesian artificial neural network, Journal of Hydrology, 540, 623-640, 2016.

Luo, J., Wang, E., Shen, S., Zheng, H., and Zhang, Y.: Effects of conditional parameterization on performance of rainfall-runoff model regarding hydrologic non-stationarity, Hydrological Processes, 26, 3953-3961, 2012.

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., and Kuczera, G.: Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors, Water Resources Research, 53, 2199-2239, 2017.

Pappenberger, F., Ramos, M. H., Cloke, H. L., Wetterhall, F., Alfieri, L., Bogner, K., Mueller, A., and Salamon, P.: How do I know if my forecasts are better? Using benchmarks in hydrological ensemble prediction, Journal of Hydrology, 522, 697-713, 2015.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resources Research, 46, 2010.

Yang, T., Asanjan, A. A., Welles, E., Gao, X., Sorooshian, S., and Liu, X.: Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information, Water Resources Research, 53, 2786-2812, 2017.

Ye, W., Bates, B. C., Viney, N. R., Sivapalan, M., and Jakeman, A. J.: Performance of conceptual rainfall-runoff models in low-yielding ephemeral catchments, Water Resources Research, 33, 153-166, 1997.