Response to Interactive comment Anonymous Referee #2

General comments:

This paper summarizes the application of the widely used HBV hydrologic model to streamflow forecasting in a Polish mountain river. The project uses ECMRWF ensemble weather forecasts to drive the streamflow model, and explores both pre- and post-processing of the ensembles for bias correction. Useful results are obtained, and the study has significant potential. I recommend that the paper is accepted pending major revisions.

We thank the reviewer for the assessment. We appreciate the reviewer's opinion about the potential of the study and the valuable suggestions to improve the manuscript. Below are our responses to the comments and points raised.

Detailed comments:

Comment: *1. The paper repeatedly refers to HBV as a spatially lumped model. This isn't just terminology, as around lines 20-25 of page 15, the manuscript seems to imply that the model assumes a snowpack to be present (or absent) across the entire model domain. There are a few versions of HBV, but it's normally viewed as semi-distributed, using (at a minimum) elevation bands.*

Reply: We appreciate the comment but see little opportunity to soundly add more detail in representing elevation bands. We have chosen to apply a lumped version of the HBV model, without elevation bands, because the available measurement data does not justify to enter multiple bands. For the area only five meteorological stations are available, which cannot be used to represent multiple elevation bands over the complete elevation distribution of the catchment. Following a first analysis on streamflow simulation results, there was no clear signal that model performance is largely affected by lumping so we considered it plausible to rely on the lumped model approach. If requested by the reviewers we could add a comment to Sect. 3.1.1 to address these considerations.

Comment: *2. The manuscript makes a good point on lines 29-30 of page 1 about socio-economic development increasing the impacts of extreme hydrometeorological events. It also probably bears mentioning that climate changes, both natural and anthropogenic, may further exacerbate these impacts. See Perkins, Pagano, and Garen, "Innovative operational seasonal water supply forecasting technologies," Journal of Soil and Water Conservation, 2009; and Fleming, "Demand modulation of water scarcity sensitivities to secular climatic variation: theoretical insights from a computational maquette," Hydrological Sciences Journal, 2016.*

Reply: We thank the reviewer for the comment and refer to P1 Line 29-30:

"Accurate forecasting becomes increasingly more important, ~~since~~ because frequency and magnitude of low and high streamflow events are projected to increase in many areas in the world as a result of climate change (IPCC, 2014). Due to socio-economic development ~~also~~ the impacts of extreme events further increase (Bouwer et al., 2010; Fleming, 2016; Rojas et al., 2013; Wheater and Gober, 2015)."

The first sentence aims to mention that climate change exacerbate both low and high streamflow events. Following the reviewer's comment we will add "as a result of climate change" to make the statement more explicit. The paper by Fleming (Hydrological Sciences Journal, 2016) is a good reference for the second sentence.

Comment: *3. Terms could stand to a little better defined. For example, most flood and water supply forecasters who I know would regard "short-term" forecasts as having lead times of 0-10 days, and "long-term" forecasts as having lead times of weeks to months. So what the authors refer to here as "medium-term" would be referred to as "short-term" by many if not most others working in the field. And no effort is made here to distinguish medium-*

term from short-term hydrologic forecasting. More broadly, some of the wording throughout the manuscript would benefit from a re-think for better clarity and precision.

Reply: To be consistent with respect to forecast windows, we explicitly define "medium-range" forecasts and follow the definition for "medium-range" by the World Meteorological Organization, which is also followed by ECMWF (ECMWF, 2012). WMO defines medium-range as forecasts with lead times from 3 days to 10 days, and we also refer to Olsson and Lindström (2008), Renner et al. (2009), and Roulin and Vannitsem (2005). We note that Bennett et al. (2014) refer to this range of lead times as "short-term" forecasts, so there is ambiguity. We opt to keep the term "medium-range" instead of changing it to "short-range", to remain consistent with definitions commonly used in meteorology.

In this paper the term "medium-range" is just used as a generic term to characterize the forecasting system. We do not explicitly distinguish short-range forecasts and medium-range forecasts, because in the analyses there is always referred to specific lead times.

Comment: *4. Why is only meteorological forecast uncertainty incorporated into the ensemble model? It's commonplace in the research literature for forecast models to include both meteorological uncertainty (NWP ensemble) and hydrologic model parameterization uncertainty (ensemble of hydrologic parameter values). This work is starting to make its way into operational practice too. Providing some justification for this choice might be a good idea.*

Reply: We agree to the comment, but argue that only meteorological forecast uncertainty is incorporated because this study aims to identify effects of the ECMWF meteorological forecasts on the quality and skill of streamflow forecasts. Additionally incorporating hydrological model uncertainty, parameter uncertainty and initial condition uncertainty would (partly) obscure this relation. In addition, Bennett et al. (2014), and Cloke and Pappenberger (2009) state that uncertainties in meteorological forecasts are the largest source of uncertainty beyond 2-3 days, and that only uncertainty in meteorological forecasts is incorporated in many studies (Bennett et al., 2014). We will add the above in Sect. 3.1.

Comment: *5. The description of the model implementation isn't quite adequate. What was the calibration-testing split, and what were the model performances during both phases? And it's stated that the objective function selected for calibration is "Y", which apparently combines the Nash-Sutcliffe efficiency with a volumetric error measure. Objective function selection is a key step in model calibration, and more information needs to be provided, starting with an explicit mathematical definition for "Y".*

Reply: We refer to P5 Line 16-23 where the calibration procedure is explained. The equations for Y, NS and E_{RV} are directly accessible in the cited references and we therefore hesitate to add the equations. The calibration and validation performances are listed in Table 4 and referred to on P10 Line 25-28.

Comment: *6. The updating of initial states was performed here for the slow-runoff and fast-runoff reservoirs. That's interesting and useful, but why was SWE not selected as the object of this data assimilation exercise? It seems like it would be a more rewarding, and certainly more conventional, choice in this northern continental European mountain catchment.*

Reply: We thank the reviewer for this thoughtful comment. If the catchment would have exclusively or mainly a snow regime, we would agree that updating of the snow storage would be a more logical choice. However, the catchment does not have an exclusive snow regime, but it has a mixture of regimes (also represented in Figure 4). Moreover, essential to the success of the updating procedure is the availability and quality of data on snow cover, and we consider this investigation to be out of the focus of this paper.

We have used streamflow measurements on the day preceding the forecast issuing day to update the slow and fast runoff reservoirs. This is possible because in the HBV model there is a direct connection between these reservoirs and discharge. Such a direct connection does not exist with the snow storage reservoir. Daily streamflow measurements commonly have a high autocorrelation, so it can be expected that observed streamflow on day *t-1* provides information about the storage in the slow runoff reservoir and fast runoff reservoir on day *t*. We expect that the correlation between snow water equivalent on day *t* and streamflow on day *t-1* will be much lower, and therefore updating of the snow storage using streamflow measurements will be less effective.

Comment: *7, The literature review of ensemble hydrologic forecasting, pre- and post-processing for bias corrections, and data assimilation and model updating, is a good start but seems a little light. Citing more work would provide valuable context to the paper. A reasonable place to start might be recent work by Dominique Bourdin at the University of British Columbia and Hamid Moradkhani at Portland State University.*

Reply: We thank the reviewer for suggesting these sources of additional relevant literature, especially the work by Moradkhani about pre- and post-processing (e.g. Khajehei and Moradkhani, 2017; Madadgar et al., 2014) and updating and data assimilation (e.g. Liu et al., 2012; Pathiraja et al., 2015; Yan and Moradkhani, 2016), and the work by Bourdin which contains recent developments in ensemble streamflow forecasting (Bourdin et al., 2012; Bourdin and Stull, 2013). We will further study the papers by Moradkhani and Bourdin and use this to further extend the context of the paper on ensemble streamflow forecasting, pre- and post-processing and updating procedures.

Comment: *8. Some of the specific conclusions seem a little surprising. That's great, but it also means they'd benefit from additional discussion. In particular, the paper concludes in section 4.2.1 that the quality of the forecasts at lead times of less than 3 days is dominated by hydrologic initial conditions, and the weather forecasts become the dominant source of predictive skill after that. This would be a reasonable conclusion for a large or flat basin, but for a small, steep mountain river it seems a little surprising – these are typically flashy systems that respond to rain or snowmelt inputs within a day or so. Indeed, a few pages later near the end of section 5, the paper states that "in the hydrological model the lag time between a rainfall event and the streamflow peak is set to 1 day." It also seems that conclusions like this, which attempt to attribute predictive skill (and therefore also predictive error) to various different sources, might be difficult to make convincingly without using a more statically sophisticated and exhaustive data assimilation procedure, incorporating ensembles of hydrologic models and/or model parameters, etc.*

Reply: We thank the reviewer for this comment and will further explain the observations of the reviewer. Regarding the comment that "the quality of the forecasts at lead times of less than 3 days is dominated by hydrologic initial conditions, and the weather forecasts become the dominant source of predictive skill after that" is "a little surprising": our results show that this depends on the streamflow category and the streamflow generating process. Short-rain generated high streamflows, snowmelt generated high streamflows and snow accumulation generated low flows are skillfully forecasted by the meteorological forecasts after 1 or 2 days, which could be expected for these fast processes and confirms the expectations of the reviewer. For long-rain generated high streamflows, medium streamflows and precipitation deficit generated low streamflows the maximum skill is observed at larger lead times, because for these processes both the forecasts and the alternative forecasts are dominated by the initial conditions at small lead times.

References

Bennett, J. C., Robertson, D. E., Shrestha, D. L., Wang, Q. J., Enever, D., Hapuarachchi, P. and Tuteja, N. K.: A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9 days, J. Hydrol., 519, 2832–2846, doi:10.1016/j.jhydrol.2014.08.010, 2014.

Bourdin, D. R. and Stull, R. B.: Bias-corrected short-range Member-to-Member ensemble forecasts of reservoir inflow, J. Hydrol., 502, 77–88, doi:10.1016/j.jhydrol.2013.08.028, 2013.

Bourdin, D. R., Fleming, S. W. and Stull, R. B.: Streamflow Modelling: A Primer on Applications, Approaches and Challenges, Atmosphere-Ocean, 50(4), 507–536, doi:10.1080/07055900.2012.734276, 2012.

Bouwer, L. M., Bubeck, P. and Aerts, J. C. J. H.: Changes in future flood risk due to climate and development in a Dutch polder area, Glob. Environ. Chang., 20(3), 463–471, doi:10.1016/j.gloenvcha.2010.04.002, 2010.

Cloke, H. L. and Pappenberger, F.: Ensemble flood forecasting: A review, J. Hydrol., 375(3–4), 613–626, doi:10.1016/j.jhydrol.2009.06.005, 2009.

ECMWF: Describing ECMWF's forecasts and forecasting system, edited by B. Riddaway, ECMWF Newsl., 133, 11–13 [online] Available from: http://old.ecmwf.int/publications/newsletters/pdf/133.pdf, 2012.

Fleming, S. W.: Demand modulation of water scarcity sensitivities to secular climatic variation: theoretical insights from a computational maquette, Hydrol. Sci. J., 61(16), 2849–2859, doi:10.1080/02626667.2016.1164316, 2016.

IPCC: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Core Writing Team, R. K. Pachauri, and L. A. Meyer, IPCC, Geneva, Zwitzerland. [online] Available from: http://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_FINAL_full.pdf, 2014.

Khajehei, S. and Moradkhani, H.: Towards an improved ensemble precipitation forecast: A probabilistic post-processing approach, J. Hydrol., 546, 476–489, doi:10.1016/j.jhydrol.2017.01.026, 2017.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., Seo, D. J., Schwanenberg, D., Smith, P., Van Dijk, A. I. J. M., Van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O. and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16(10), 3863–3887, doi:10.5194/hess-16-3863-2012, 2012.

Madadgar, S., Moradkhani, H. and Garen, D.: Towards improved post-processing of hydrologic forecast ensembles, Hydrol. Process., 28(1), 104–122, doi:10.1002/hyp.9562, 2014.

Olsson, J. and Lindström, G.: Evaluation and calibration of operational hydrological ensemble forecasts in Sweden, J. Hydrol., 350(1–2), 14–24, doi:10.1016/j.jhydrol.2007.11.010, 2008.

Pathiraja, S., Marshall, L., Sharma, A. and Moradkhani, H.: Hydrologic modeling in dynamic catchments: A data assimilation approach, Water Resour. Res., 52(5), 3350–3372, doi:10.1002/2015WR017192, 2015.

Renner, M., Werner, M. G. F., Rademacher, S. and Sprokkereef, E.: Verification of ensemble flow forecasts for the River Rhine, J. Hydrol., 376(3–4), 463–475, doi:10.1016/j.jhydrol.2009.07.059, 2009.

Rojas, R., Feyen, L. and Watkiss, P.: Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation, Glob. Environ. Chang., 23(6), 1737–1751, doi:10.1016/j.gloenvcha.2013.08.006, 2013.

Roulin, E. and Vannitsem, S.: Skill of Medium-Range Hydrological Ensemble Predictions, J. Hydrometeorol., 6(5), 729–744, doi:10.1175/JHM436.1, 2005.

Wheater, H. S. and Gober, P.: Water security and the science agenda, Water Resour. Res., 51(7), 5406–5424, doi:10.1002/2015WR016892, 2015.

Yan, H. and Moradkhani, H.: Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling, Adv. Water Resour., 94, 364–378, doi:10.1016/j.advwatres.2016.06.002, 2016.