Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-458-RC2, 2018

Hydrology and Earth System Sciences Discussions

Open Access

HESSD

Interactive comment

Interactive comment on "Subseasonal hydrometeorological ensemble predictions in small-and medium-size mountainous catchments: Benefits of the NWP approach" *by* Samuel Monhart et al.

Anonymous Referee #2

Received and published: 23 October 2018

General comments This is an interesting and thorough assessment of an ensemble streamflow forecasting system in snow-affected mountainous catchments. The system pairs NWP forecasts with a distributed hydrological model that includes detailed accounting of cryogenic processes. The system is technically advanced and in my opinion of high interest to the readership of HESS. The study is well conceived and very clearly written. A particular strength of the study is the extensive and thorough verification of the forecasts, encompassing multiple appropriate measures of performance that are described and discussed in clear and interesting ways. In general, the

Printer-friendly version

Discussion paper

authors' conclusions are strongly supported by their analyses. I have one quibble with the use of QM as a means for downscaling meteorological forecasts, which amounts to a minor revision. Other than this, I have no hesitation in recommending this study for publication.

Major comments It appears that the authors use QM to downscale NWP predictions from a coarse grid (∼30/60 k) to a fine grid (∼2 k). It's well established that QM is not theoretically ideal for this practice, because of so-called variance inflation (Maraun 2013). The authors appear to be aware of this, as they discuss this issue in an accompanying paper (Monhart et al. 2018). However, it is more salient in this paper, because of the hydrological modelling that is carried out. Variance inflation is only an issue when quantile mapped/downscaled meteorological forecasts are spatially re-aggregated, which is exactly what the hydrological model does. So it will not show up in the analyses carried out by Monhart et al. 2018 (where variables are not re-aggregated), but it could well be an issue in this study. In addition, and as Maraun shows, the variance inflation problem is only strongly evident for extremes. Extremes are not the focus of the analysis carried out in this paper, which is fine. But this means it's hard to tell if variance inflation is present in streamflow forecasts. As the forecasting system could be used for flood prediction, this may be a serious issue. Accordingly, I recommend two changes to the manuscript:

1) The authors should clearly describe how they bridge the gap in spatial resolution from a ∼30/60 k horizontal grid (NWP forecasts) to a ∼2 k horizontal grid (observations). And if, as I've assumed, they use QM for this purpose: 2) The authors should briefly acknowledge the issue of variance inflation in the discussion, including a discussion of possible implications for their system (perhaps alongside recommendations for dealing with these implications).

I also encourage the authors to consider measuring the impact of variance inflation on their system in future work.

Specific comments Page 2 L7-9 "For a given target day of a reforecast the correction is derived from the distribution of all the reforecasts within a three weeks window around the same lead day and the corresponding observations, hence the correction depends both on the lead time and on the period of the years". Is the QM cross-validated in some way? How are zero values in precipitation handled in the QM?

Page 5 Section 2.4 For each score used, please note the range of values taken and the orientation of the score (e.g. -infinity to 1 for NSE, with 1 being perfect) This allows easy intepretation of, e.g., Figure 4.

L16 "we use the spread to error ratio (SprErr) as an indicator for the forecast reliability" Please briefly describe how this is calculated.

Page 8 L18 Figure 3. It's very difficult to see the different colours in this figure, especially in the right hand panel - i.e., it's not possible to distinguish QM from raw. Please replot so this is clearer (e.g., with different colours/box outlines, and/or perhaps restrict the vertical axis in the rh panel to [-0.2 0.3]).

L22-23 "After bias correction the skill is higher with positive CRPSS up to three weeks in winter and spring." From the figure, precip skill looks to be negligible in DJF after week 1. Skill scores will of course be a little noisy (in time, as well as in space; the authors have only considere spatial variation), so I don't think the authors should describe forecasts as 'skillful' if they have CRPSS values only very slightly above zero.

L28 NSE, Bias - I assume these are calculated on the mean of the ensemble? Please state this in Section 2.4.

Page 9 L7-8 "The negative biases of the ESP approach indicate an underestimation of the streamflows for all lead times in the Verzasca catchment." It's not clear to me why ESP predictions would be biased. ESP forcings, by construction, are unbiased. As the bias in predictions is calculated against model climatology, there should be no bias, as occurs in the Thur. Please briefly explain what is going on here.

L15 "The spread error ratio of the ESP predictions is below 1 for all lead times indicating overconfidence." There are two issues here. First, as I already alluded to, it would be beneficial to readers unfamiliar with the spread-error ratio to offer a brief explanation of the range of values it can take, and which direction indicates over/under confidence in Section 2.4. Second, I can't understand why the ESP forecasts are not reliable. ESP forcings are by construction reliable, so the spread-error ratio for the ESP forecasts should be close to 1. But this is not so in the two smaller catchments. Why is this?

Page 10 L26-27 "The rank histograms for the ESP predictions do provide more uniform rank histograms with a weak tendency of a negative bias." Again, I would be interested in a brief explanation of this bias in ESP forecasts.

Page 12 L15 "QM indeed is able to provide reliable ensembles" To me this sounds as though QM is responsible for the reliable ensembles, and this isn't really correct. QM can improve reliability to the extent that it is negatively impacted by bias (including conditional bias - i.e., biases at different points in the marginal distribution). You have shown (like Zhao et al.) that QM does not correct for overconfidence in short lead-time forecasts. This is because the underlying forecasts are overconfident - QM can't correct this (by construction). At longer lead-times, the QM forecasts are reliable because the spread in the underlying forecasts is appropriate (notwithstanding conditional biases). The same goes for coherence (discussed in the accompanying Monhart et al. 2018 JGR paper evaluating the NWP forecasts) - QM is not capable of correcting negative skill wrt to climatology in forecasts, other than that due to bias (again, by construction). The forecasts presented in this study are coherent because the underlying (raw) forecasts appear to be neutrally skillful at long time scales. In other words, it is the combination of NWP model and the QM that create the reliable and coherent forecasts shown in this study, not just QM. I think it would be better to reword what's written here to reflect this fact.

P14 L9-10 "to correct the errors and biases of the hydrological simulation" and, presumably, to account for additional uncertainty induced by the hydrological model in the ensemble?

P14 "To do so we verified the streamflows of the reference simulation of the PRE-VAH model against observations." One thing not discussed here is reliability. I assume when assessed against observations, the ensembles are highly overconfident because uncertainty in the hydrological model is not included in the ensemble (see, e.g., Bennett et al. 2014). This is especially true at very short lead times (perhaps <3 days), when hydrological model uncertainty may be the dominant source of uncertainty in the forecasts. This may be worth mentioning.

Typos/Grammar Page 1 L14 "Prior of" should be "Prior to" or more simply "Before"

Page 2 L11 "both," delete comma

Page 5 L2 "comparison of to" delete "of" L28 "year" should be "years"

Page 7 L3 "station" should be "stations"

Page 8 L9 "evaluate of operational" delete "of"

Page 9 L9 "and reach" should be "and reaches" L11 "enhance" should be "enhances" L11 "elongates positive up" I think "skill" is missing here - i.e. "elongates positive skill up" L27 "are shown" should be "is shown"

Page 13 L1 "in seasonal meteorological can" I think this should be "in seasonal meteorological forecasts can"

Page 14 L11 "prediction" should be "predictions" L23 "exhibit" should be "exhibits"

Page 15 L14 "enhance" should be "enhances" L24 "EPS" should be "ESP" L25 "both, temperature" delete comma

Page 27 L5 "whereas" should be "where"

References Bennett JC, Robertson DE, Shrestha DL, Wang QJ, Enever D, Hapuarachchi P, Tuteja NK. 2014. A system for continuous hydrological ensemble fore-

Printer-friendly version

Discussion paper

casting (SCHEF) to lead times of 9 days. Journal of Hydrology 519: 2832-2846. DOI: 10.1016/j.jhydrol.2014.08.010.

Monhart, S., Spirig, C., Bhend, J., Bogner, K., Schär, C. and Liniger, M. A.: Skill of Sub-seasonal Forecasts in Europe: Effect of Bias Correction and Downscaling using Surface Observations, J. Geophys. Res. Atmos., 1–18, doi:10.1029/2017JD027923, 2018.

Zhao, T., Bennett, J. C., Wang, Q. J., Schepen, A., Wood, A. W., Robertson, D. E. and Ramos, M. H.: How suitable is quantile mapping for postprocessing GCM precipitation forecasts?, J. Clim., 30(9), 3185–3196, doi:10.1175/JCLI-D-16-0652.1, 2017.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-458, 2018.

Printer-friendly version

Discussion paper