

## ***Interactive comment on “Relative effects of statistical preprocessing and postprocessing on a regional hydrological ensemble prediction system” by Sanjib Sharma et al.***

**M. Scheuerer (Referee)**

michael.scheuerer@noaa.gov

Received and published: 13 September 2017

This manuscript studies the relative roles of statistical preprocessing of meteorological inputs in a hydrological forecast system and statistical postprocessing of the resulting flow forecasts for four basins in the US middle Atlantic region. The paper is well written, the structure is good, and the conclusions are interesting and relevant. The methodology is sound with two exceptions detailed below. These are major in the sense that they are scientifically problematic and may have an impact on the conclusions, but they can probably be addressed quite easily.

Specific comments:

C1

- p6, l4:  $\pi_i$  is only a probability when  $y_i=0$ , otherwise a likelihood
- p7, l15: 'smallest mean CRPS is selected': I don't fully understand how this works. Apparently  $c_{i+1}$  changes over time, so what exactly is minimized here? The CRPS over some training data with a rolling training window? Please add some more explanation
- p8, l15-16: '... is focused on flood events ... by choosing flow amounts greater than ...': This kind of subsetting is very problematic and can lead to false conclusions about the relative predictive performance of different methods, see Lerch et al. (2017). Bellier et al (2017) give a discussion of pitfalls of sample stratification and make suggestions how one can stratify samples in a way that avoids these pitfalls
- Section 4.4.1: I'm not sure if this part of the analysis makes sense. In addition to the stratification issue (which demonstrably entails a bias), it is also known that the ensemble mean does not necessarily yield the best/appropriate point forecast when a relative error statistic is considered (see Gneiting 2011). I suggest either considering the mean error (over the entire verification data), or omitting this subsection entirely and maybe replace it by a subsection that studies reliability of threshold exceedance

Language and typos:

- p6, l31: hourly
- p7, eq (7):  $x_{i+1} \rightarrow x_{i+1}$
- p9, l15: It sounds weird to say that one basin outperforms the other, please reformulate
- p10, l24: Replace 'While' by 'The gains ... , on the other hand, ...'

References:

Bellier et al. (2017): Sample Stratification in Verification of Ensemble Forecasts of Continuous Scalar Variables: Potential Benefits and Pitfalls. Monthly Weather Review,

C2

<https://doi.org/10.1175/MWR-D-16-0487.1>

Gneiting (2011): Making and Evaluating Point Forecasts. Journal of the American Statistical Association, <https://doi.org/10.1198/jasa.2011.r10138>

Lerch et al. (2017): Forecasters dilemma: Extreme events and forecast evaluation. Statistical Science, <https://doi.org/10.1214/16-STS588>.

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2017-514>, 2017.