

Interactive comment on “Seasonal streamflow forecasting by conditioning climatology with precipitation indices” by L. Crochemore et al.

Anonymous Referee #1

Received and published: 24 July 2016

As an outsider to the professional academic world, I feel that I cannot speak with unquestionable credibility to the novelty or scientific soundness of this manuscript – I am simply not familiar enough with the wealth of recent research into seasonal hydrologic forecasting. However, I can supply my overall impression of this work, which may be useful given my background in operational hydrologic forecasting.

The authors reference several studies that utilized approaches similar to the one undertaken here – conditioning historical observation-based ensembles to improve forecasts generated from these ensembles. Thus, the fundamental direction of the current study is not overly original. However, the manner in which the conditioning was applied – using GCM- and climatology-derived precipitation indices to select the most relevant historical ensembles – does appear to be a novel approach.

C1

The potential utility of this approach is presented well in Figure 2, where the precipitation indices generated from the GCM hindcasts (ECMWF Sys4) are compared against those generated from the historical observations. As the authors state, the Sys4 indices perform at least as well as the base indices overall (CRPSS), even outperform at one month lead time, but are consistently sharper (IQRSS). Further, the Sys4 indices have good reliability overall (Figure 3). The reliability of the indices falters when looking at only drier than normal or only wetter than normal conditions, but this seems to be unavoidable with any forecasting approach.

Despite the prefaced potential of using the Sys4 precipitation indices to condition, or subset, historical ensembles, this study's results offer just marginal practical insight:

1) Subsetting the ensembles based on the precipitation indices improve the HistQ performance more than the it does the ESP performance. This result is not very useful, however, since the HistQ approach is rudimentary (and likely rarely used), and the primary benefit of the conditioning is seen during short lead times (which is simply the effect of blending from the last streamflow observation).

2) For ESP, SPI-conditioning appears to outperform SUM-conditioning, but this statement is qualitative at best and neither set of conditioned ensembles provides any notable improvements over the base ensembles. Compared to the base ESP ensembles, the sharpness of the ESP_SPI3 ensembles was improved by up to 10% but the reliability was degraded by up to 40% (Figure 7).

3) The conditioning improved the performance of HistQ ensembles in forecasting low flow events and variables, but the conditioned ensembles were still less skillful than the Sys4 and ESP/ESP_SPI3 ensembles.

4) The authors state that the ESP_SPI3 approach "systematically appears to be one of the best options to forecast deficit volumes." However, this conclusion is very subjective, as it is not authoritatively substantiated by the results presented in Figures 9 and 10.

C2

Although several pages of this manuscript are spent discussing the results in great detail, and the authors walk through the discussion in a relatively clean, scientific manner, much of this discussion is centered around tangential topics. For example, the comparisons between the conditioned ESP/HistQ ensembles to the Sys4 ensembles seem irrelevant given that the conditioning did little to improve, and actually degraded in some cases, the skill compared to the base ensembles. Thus, comparing the conditioned ensembles to the Sys4 ensembles is equivalent to comparing the base ensembles to Sys4, which of course is unnecessary. The results should be restricted to and presented with the stated goal of the study in mind – improving the skill of historical observation-based ensemble forecasting systems.

Unfortunately, because there is little to report on the utility of applying this conditioning method to seasonal streamflow and low flow forecasting, the authors may need to redesign and/or include other experiments before resubmitting this paper. One suggestion, actually offered by the authors, is to examine the utility of using SPEI to condition the ensembles. Although the SPI is likely sufficient to appropriately subset historical precipitation ensembles, it may not be sufficient from a streamflow perspective. It seems likely that the relative magnitude of an individual SPI value may not always be translated into a similar relative magnitude flow or volume value if ET is a major hydrologic control in the watershed of interest (i.e. late season streamflows can be very different following extended dry but mild vs extended dry but hot conditions). Thus, conditioning the ensembles with both precipitation- and temperature-driven indices may provide more robust results.

Lastly, the underlying standard of this manuscript is the stated inherent reliability of historical observation-based ensembles, but this is a bit misleading. In true forecasting (not hindcasting), climatology-driven predictions may not be all that reliable. Several decades worth of historical information is often sought to build an ensemble forecasting system, but the climatic regime of the forecast area may be changing too rapidly for this. Thus, the distribution functions of actual forecasts and their corresponding obser-

C3

ervations may be offset from one another (i.e. not fall on a 1:1 line). Perhaps the authors should frame the goal more along the lines of using the conditioning to sharpen the ensembles, and less along the lines of marrying the reliability of historical ensembles with the sharpness of GCMs.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-285, 2016.

C4