Hydrol. Earth Syst. Sci. Discuss., 7, C2709–C2711, 2010
www.hydrol-earth-syst-sci-discuss.net/7/C2709/2010/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments" *by* J. A. Velázquez et al.

J. A. Velázquez et al.

juan-alberto.velazquez.1@ulaval.ca

Received and published: 6 October 2010

We thank Referee #2 for his objective review and important comments. Our response follows:

1."The CRPS described by Gneiting and Raftery (2007) is very good in the case the cdf (Ft) is known and has the advantage that it can be calculated very quickly for the optimization purposes. However, in the case of small ensemble numbers, a potentially

large variance of this ensemble could attenuate the CRPS of Gneiting and Raftery (2007). The authors may want to consider using the formulation of Hersbach (2000) instead. (Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and forecasting, 15: 559-570.)"

The reference Hersbach (2000) gives useful guidance on the computation of the score using a discrete representation of the forecast cdf. In our study we worked with ensembles of 17 (and less) members, so we preferred to fit the ensemble to a pdf in order to obtain a more continuous computation. To deal with the variance, we use a large vector in Monte Carlo simulations.

The paragraph will be completed as follows in Page 4030 Line 19: Where X and X' are independent vectors consisting of 1000 random values from a gamma distribution adjusted to the predictive function.

2."There might be misunderstanding how the mean of the CRPS has been calculated – the stations or the simulations. The authors may want to specify more clearly."

For each station, the mean CRPS was calculated with the mean of the scores obtained for each pair ensemble-observation.

The paragraph will be completed as follows in Page 4030, line 21:

However, because the score for a specific forecast-observation pair, at a certain time, cannot be interpreted, we rather consider for each station the average of all individual scores as a measure of the quality of the simulation system, thus comparing mean AE (MAE) and mean CRPS, which values are directly proportional to the magnitude of the observations.

3."According to table 1 the ranges in catchment size are quite large. Obviously the hydrological responses are very different between small size, flashflood type, catchments and the larger catchments. Did the authors look at their results also aggregated to catchment size? This could yield interesting results in particular with regards to the

HESSD

7, C2709–C2711, 2010

extremes of the distributions."

After looking carefully, we have not found evidence of any dependence of the model results to the catchment size.

4."In their analysis the authors considered 10 years time series which will have been dominated largely by average flows. In section 3.2 the authors state that over all flow quantiles the probabilistic system performs better than the deterministic system and "that the system is better at detecting larger events such as quantiles 50 or higher, than low flow events such as quantile 10". What is their explanation for this?"

Our explanation can only be in good part speculative. 1) The models were calibrated on the MSE, which is known to penalize more large simulations errors (from the larger flows). 2) The model structures may also be somewhat biased toward larger flows.

References: Gneiting, T. and Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation, J. Am. Stat. Assoc., 102 (477), 359-378, 2007.

Hersbach, H.: Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, Weather Forecast., 15, 559–570, 2000.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 4023, 2010.