Interactive comment on "Evaluation of a probabilistic hydrometeorological forecast system" *by* S. Jaun and B. Ahrens

Anonymous Referee #2

Received and published: 8 April 2009

This paper analyses a system consisting of the limited-area meteorological ensemble prediction system COSMO-LEPS and the semi-distributed hydrological model PRE-VAH. This system is compared with a deterministic system in which the COSMO-7 model provides meteorological forecasts to the same hydrological model. An interesting set of methods for the crossed verification of the probabilistic and deterministic hydrological forecasts is presented. The methodology is then applied to hindcasts corresponding to two years, and to various catchments in Upper Rhine basin.

The chosen methodology is appealing, the paper is prepared with care and the figures invite for a closer inspection. The following comments are related to the spread-skill relationships and refer to different paragraphs across the paper starting p. 1850. The method adopted is simple but not much used so far; the choice could be motivated in

respect to other methods that are more largely documented (e.g. Sherrer et al. 2004 or Houtekamer 1993). A perfect probabilistic forecast could be defined and the qualities addressed with the spread-skill relationship could be explained. The authors are appropriately replying to one possible drawback cited by Lalaurette et al (2005) by testing also their artificial ensemble (HART) which is described p. 1853. The results are mentioned p 1859 but unfortunately not shown. The assumption that the second and third quartiles are symmetrical (Lalaurette et al, 2005) could be commented in regard of Fig. 6. In Fig. 7 and 8, the results are presented together for all the catchments. This includes catchments of the same river at different gauge locations (Table 1). Doesn't such a mixture contribute to the good statistical relationship obtained? A more detailed analysis could take a better profit of this large dataset. Wouldn't the spread-skill relationship of HART be more informative than Fig. 7 (HEPS compared with observed runoff)? Given the large dispersion of the blue circles in such graphs, wouldn't error bars help in assessing the relationship?

The comment about merging the results of all catchments applies also to the RPSS. This latter score has been preferred to the BSS (p. 1851). However, the 0.95 quantile has been added to the quartiles to define the categories. Does this added category resolve runoff peaks as intended?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 6, 1843, 2009.