Hydrol. Earth Syst. Sci. Discuss., 8, C572–C577, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C572/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios" *by* N. Addor et al.

Anonymous Referee #3

Received and published: 21 March 2011

This manuscript addresses an important topic: how the ensemble forecasts produced by the operational hydrological ensemble prediction system for the city of Zurich perform for a range of events and could better support forecasters and users' decision making. This evaluation study based on real-life case is crucial to modelers, forecasters, and end users as it quantifies the different aspects of the ensemble forecast quality and their benefits for risk-based decision making in various situations. It also helps identify areas of future enhancements of the forecasting system. The paper focuses on the evaluation of ensemble reforecasts, including their comparison with deterministic forecasts, the assessment of the error sources in the model chain, and the performance of the output ensembles for two hypothetical extreme discharge scenar-

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

ios. The authors describe the different aspects of the forecast quality through various verification metrics and graphic plots. They underline how the hydrological ensemble prediction system could be used for decision making in flood mitigation and dam management.

The manuscript is well written and need only minor revision to address the following comments.

- Page 720, lines 13-15: the 3rd objective could be reworded to include first the need to describe the different aspects (or attributes) of forecast quality, before listing specific verification metrics.

- Page 720, lines 22-27: the authors could generalize the statement that forecast skill depends on temporal and spatial scales. Also the evaluation of the forecast quality is specific to the basin and application of interest. Here the focus is on flood mitigation, therefore concentrating on high flows.

- Page 721, lines 15-16: the authors could emphasize the importance of reforecast datasets from the atmospheric models to use a fixed model and describe the expected performance of hydrologic ensemble forecasts. This is especially true when flood mitigation measures need to be developed by forecast users as these rules need to focus on rare events. If post-processing of the hydrologic and hydraulic forecasts is developed, a statistical approach will greatly benefit from such reforecasts.

- Page 721, line 22: the authors should describe how 16 ensemble members are selected with clustering technique and give references (e.g., Marsigli et al. 2005; Renner et al. 2009).

- Page 724, lines 2-7: the authors should mention that no bias correction or calibration of the hydrologic and hydraulic ensemble forecasts is done as the current operational ensemble system only quantifies and propagates the atmospheric uncertainty by ingesting atmospheric ensembles from COSMO-LEPS. The need to account for the

HESSD

8, C572–C577, 2011

hydrologic uncertainty should be reinforced (see comments below).

- Page 724 lines 16-20: I think the authors should mention the benefits of having reforecasts for a longer time period; for example, the Q0.99 value considered in this study would have larger sample size and would provide useful information for developing flood mitigation measures.

- Page 725 lines 16-18: suggest adding: “Nevertheless, the ensemble forecasts are usually reduced to their ensemble mean or median value in practice for comparison to deterministic forecasts”.

- Page 726 lines 6-10: the authors should first explain why they pick the Brier Score since the case study focuses on warnings and thresholds for flood mitigation and BS can be computed for both deterministic and probabilistic forecasts (otherwise one could argue that the overall quality of the forecast ensembles would be better estimated with the Continuous Ranked Probability Score). Then they should mention that they used the Brier Skill Score (BSS) to estimate the skill of each of the forecasts in comparison to a reference forecast. They should also explain what reference forecasts they use as the reference forecasts need to be meaningful for the considered case study. To analyze how much gain the COSMO-LEPS ensembles bring to the hydrologic ensembles, one could generate hydrologic ensembles based on climatological forcing inputs using the same hydrologic model chain and same initial conditions (see Demargne et al., 2010 for such analysis).

- Page 727 lines 2-5: the authors should give a reference for estimating the confidence intervals by the bootstrapping technique with replacement. To improve the estimation of the sampling uncertainty in the metrics, the authors could consider block bootstrapping to account for temporal dependency (see Lahiri, 2003).

- Page 727 equation 4: suggest changing the denominator to be “observed non-events”.

HESSD

8, C572–C577, 2011

- Page 728 line28-29: the ROC is by definition the comparison of Hit Rate (H, or Probability of Detection) and False Alarm Rate (F, or Probability of False Detection). It is a discrimination measure conditioned on the observations (H for observed events, F for observed non-events). The measure proposed by the authors that compares Hit Rate and False Alarm Ratio is a mixture of a metric conditioned on the observed events (which measures the forecast discrimination) and a metric conditioned on the forecast events (which measures the forecast reliability). Given these major differences, the proposed measure should not be called ROC.

- Page 728 line13-14: the rank histogram describes the unconditional reliability of the forecast; the term "forecast consistency" is usually mentioned to describe temporal consistency of consecutive forecasts.

- Page 728 lines 19-23: the authors should mention why the temporal consistency or persistence of the ensemble forecasts is meaningful to forecasters and forecast users, especially when focusing on flood mitigation actions that are based on specific thresholds. In future studies, indices of forecast temporal consistency could also be used to complement the visualization plot proposed by the authors. Forecast consistency (also called forecast continuity and forecast convergence) has been discussed by different authors from the atmospheric community and applied to weather forecasts (see discussion in Kay, 2004 and in Lashley et al., 2008).

- Page 730 lines 17-18: would suggest adding "the added value conveyed by the probability information, even when using a single-valued estimate from the probabilistic forecast,. . ."

- Page 731 lines 13-15: would suggest adding "As uncertainty increases with lead time, the gain in using probabilistic forecast (vs. deterministic forecast) is larger". - Pages 733 lines 4-5: the authors should mention whether there is any over-estimation of precipitation occurrence (PoP) and very light rain events, as it is common with Numerical Weather Prediction model outputs.

HESSD

8, C572–C577, 2011

- Page 733 lines 18-21: the authors should clarify that the current system quantifies and propagates only the uncertainty in the atmospheric forcing inputs; for future enhancements, the hydrologic uncertainty should also be quantified.

- Page 735 lines 1-4: the authors could clarify whether the COSMO-LEPS forecasts have an unconditional bias, or conditional bias (e.g., over-forecasting light rain events and under-forecasting very large rain events) since a conditional bias is more difficult to correct. Also suggest rewording the benefits of reforecasts to calibrate precipitation forecasts: the availability of reforecasts for longer period should improve the calibration process, especially in presence of a conditional bias, as large samples are available from a fixed version of the model.

- Page 739 lines 9-12: the authors should emphasize the need for reforecast datasets when developing risk-based decision making rules or when calibrating a decision support system.

- Page 740 lines 4-6: the need for the quantification of the hydrologic uncertainty should be more strongly stated given the hydropower production on the lake and the dam regulations and the need to better support flood mitigation measures.

- Page 741 lines 13-14: the authors should use a stronger statement about the calibration of precipitation forecasts; suggest rewording “As calibration improves the reliability of precipitation forecasts, it is expected to improve the discharge forecasts”. Would also add the need to account for the hydrologic uncertainties.

- Page 741 Lines 20-22: would suggest adding the need for longer atmospheric reforecasts to better support evaluation studies of extreme events and development of decision support rules or system for hydrologic applications.

References:

Kay, M.P., 2004: The design and evaluation of a measure of forecast consistency for the Collaborative Convective Forecast Product.

HESSD

8, C572–C577, 2011

Preprints, 11th Conference on Aviation, Range and Aerospace Meteorology, 4-8 October, Hyannis, MA, Amer. Met. Soc. Available at http://www.esrl.noaa.gov/gsd/ab/fvs/publications/articles/kay_consistency_ARAM2004.pdf

Samuel L. Lashley, L. Fisher, B.J. Simpson, J. Taylor, S. Weisser, and J.A. Logsdon, A.M. Lammers (2008) Observing verification trends and applying a methodology to probabilistic precipitation forecasts at a National Weather Service Forecast Office. Preprints, 19th Conf. on Probability and Statistics, New Orleans, LA, Amer. Meteor. Soc., 9.4. Available at http://ams.confex.com/ams/pdfpapers/134204.pdf

Lahiri S., 2003 : Resampling Methods for Dependent Data. New York : Springer.

Demargne J., Brown J., Liu Y., Seo D.-J., Wu L., Toth Z., and Zhu Y., 2010 : Diagnostic verification of hydrometeorlogical and hydrologic ensembles. Atmospheric Sicence Letters, 11(2), 114-122.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 715, 2011.