Hydrol. Earth Syst. Sci. Discuss., 8, C1118–C1129, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C1118/2011/

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.

N. Addor et al.

massimiliano.zappa@wsl.ch

Received and published: 22 April 2011

We would like to thank the referee for her/his valuable comments which clearly contributed to improve both the structure and the content of this manuscript. The pertinence and the constructive dimension of her/his review were greatly appreciated.

Several remarks from the three referees concerned two important aspects of the manuscript: the under-sampling and the assessment of the uncertainty sources. We propose to dedicate two new Sections (2.3 and 3.1) as well as a few new paragraphs to address these concerns. They are presented in first part of the reply to referee H.

Cloke.

We submit in continuation our answers to each point raised by the referee. When necessary, we invite the referee to refer to the aforementioned submitted modifications.

1 - 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.

We propose to reword the 3rd objective as follows:

(3) to analyse forecasts reliability and eventual under- or overforecasting bias, under- or overdispersion of the ensemble members, proneness to false-alarms as well as forecasts ability to capture observed events.

To maintain the concise formulation of this objectives list, we propose not to detail the specific verification metrics, which are developed latter in the manuscript.

2 - 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.

We propose to replace lines 22-27 of page 720 by:

Earlier studies have shown that forecast skill depends on temporal and spatial scales. For instance, the current state of knowledge for larger basins suggests that the skill of ensemble prediction systems improves with increasing catchment size (Renner et al., 2009). Furthermore, forecast uncertainty is reported to decline with increasing catchment size (Jaun et al., 2008). The usefulness of such systems in small mesoscale areas has not yet been investigated and therefore we concentrate on the skill of an operational HEPS for a comparatively small catchment. Concerning the temporal scales considered, while the focus of this study is on high flows (spanning over a few days), we additionnaly assess several aspects of the performance of the model for average

discharge situations (on the basis of the entire 31-month reforecast period).

3 - 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.

This comment was taken into account when formulating the new 3.1 Section (please refer to the first part of the reply to referee H. Cloke) and in the redaction of the answer to referee M. Werner's sixth remark.

4 - 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).

We hope that the answer to referee H. Cloke's second comment brings the necessary information in this respect.

5 - 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 hydrologic uncertainty should be reinforced (see comments below).

We propose to add the following paragraph at the end of the Section 2.2 (line 19, page 724):

The performance of the system was assessed for its current operational setting. Hence, although calibrated COSMO-LEPS precipitation forecasts exist (Fundel et al., 2010), they were not used. No calibration (e.g. Reggiani et al., 2009) or correction for underdispertion or bias was applied to the output of the atmospheric forecasts before their

use in PREVAH. Similarly, no bias correction or post-processing of the hydrologic and hydraulic ensemble forecasts was done.

The new 2.3 Section was written to deal with the necessity to account for hydrologic uncertainty (please refer to the reply to referee H. Cloke).

6 - 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.

The necessity of a longer reforecast is discussed in the new 3.1 Section.

7 - 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".

OK

8 - 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).

We propose to replace the lines 6-13 of page 726 by:

To compare the performance of deterministic and probabilistic forecasts, the Brier Score (BS, Eq. 1) was chosen (e.g. Wilks, 2006). This score can be seen as a mean squared error of probabilities and has the advantage that it can be applied to both deterministic and probabilistic forecasts, without requiring the transformation of a probability forecast into a deterministic one (e.g. by considering the median only). Furthermore, while a continuous ranked probability score would enable to assess the overall quality of the forecast ensembles, the BS permits to focus on specific warnings and thresholds meaningful for this case study, which in our view, leads to a more detailed analysis. The Brier Skill Score (BSS) was used to estimate the skill of each of the forecasts in comparison to a reference forecast. The reference chosen BSref is the probability of occurrence derived from the climatology. Using COSMO-7 as a reference to compute COSMO-LEPS BSS would have been a direct way to assess the added-value of the probabilistic forecast in comparison to the deterministic forecast. This assessment is however also possible with the chosen reference, and permits in addition an individual evaluation of the two models. To analyse how much gain the COSMO-LEPS ensemble brings 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).

For the ensembles based on climatological inputs, as well as for the points 9, 15 and 17, we acknowledge that the proposed investigations would enable a finer diagnosis of the model chain performance. However, as the manuscript already deals with many different aspects of forecast verification, and as we do not expect that the analysis suggested by the referee will modify the main conclusions of this manuscript, we propose not to carry them out. We however propose to mention these points in the text because we will consider them in the future and we hope that it will encourage additional studies to tackle them.

9 - 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).

We propose to cite Efron (1992) and add on line 7 page 727 that:

To improve the estimation of the sampling uncertainty in the metrics, block bootstrapping could be considered to account for temporal dependency (see Lahiri, 2003).

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

OK

11 - 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.

We agree and propose to rename the concerned plots H-FAR curves.

12 - 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.

We propose to remove the reference to consistency to avoid confusion:

Rank diagrams show the rank of OBS (resp. of HREF) within the ensemble members (Anderson, 1996). They highlight whether the ensemble includes OBS (resp. HREF) being predicted as an equiprobable member.

13 - 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).

We propose to add the following paragraph at the beginning of Section 3.2 (page 728):

Uncertainty in probabilistic forecasts is not solely depicted by the spread of the ensemble members, but is also reflected by the persistence of the forecast, i.e. the consistency with which an event is forecast by successive model runs. For instance, a model showing great variability from one run to the next will be interpreted by the end-user as uncertain. This has significant consequences when the forecasts are used for decision-support, for example to decide on flood mitigation actions. In presence of a forecast showing great variability, the end-user might prefer not to base her/his decision on this forecast and to wait for the release of the next model run, delaying thus the decision process and taking the risk to end in an emergency situation, with a limited range of generally sub-optimal actions at choice. Forecast consistency is therefore greatly valued by end-users (Lashley et al., 2008).

For the evaluation of the forecasts for the two most intense events of the study period, a novel way to graphically asses forecast consistency is proposed.

We propose to add the following paragraph on line 12, page 728:

Indices to quantify forecasts consistency (Kay, 2004, Lashley et al., 2008) would constitute a helpful complement to the graphic representations.

14 - 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,. . .”

OK

14 - 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”.

OK

15 - Pages 733 lines 4-5: the authors should mention whether there is any over-estimation of pre- cipitation occurrence (PoP) and very light rain events, as it is common with Numerical Weather Prediction model outputs.

We propose to add this sentence on line 6 page 733 by:

It is also possibly related to an over-estimation of precipitation occurrence (PoP) and very light rain events, as it is common with numerical weather prediction model outputs.

16 - 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.

We took this aspect into account when formulating the new Section 2.3 and also in our reply to point 21.

17 - 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.

We propose to replace the sentence of lines 2-4 of page 735 by:

This tendency of COSMO-LEPS to produce too wet forecasts has also been demonstrated for Switzerland using a single-member reforecast of 30 years (Fundel et al., 2010). The mean amplitude of this bias depends on the intensity of the event considered, the region and the season, but in the large majority of the country (including the Sihl catchment) precipitation amounts are generally overestimated. This bias can however be reduced consistently by post-processing calibration, which leads to more reliable forecasts (Fundel et al., 2010).

We now mention that a longer reforecast would be useful for post-calibration in the new 3.1 Section.

18 - 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.

We mentioned this point in the new 3.1 Section.

19 - 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.

We took this aspect into account in the reply to point 21.

20 - 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”.

OK

21 - Page 741 lines 13-14: Would also add the need to account for the hydrologic uncertainties.

We propose to replace lines 11-19 on page 741 by:

This study focuses on the uncertainty related to atmospheric boundary conditions.

Probable future developments include the integration of modules to account for other uncertainty sources such as the formulation of the atmospheric models, the stations measurements, the interpolation errors, the estimation of the hydropower production and the hydrological and hydraulic modelling. For instance, the combination of ensemble forecasts with deterministic forecasts (seamless predictions) will probably be explored (e.g. Dietrich et al., 2008) to give more weight to uncertainties stemming from the formulation of atmospheric models. COSMO-7 and COSMO2 being more frequently updated than COSMO-LEPS, this would furthermore provide time-lagged ensembles of discharge predictions (e.g. Zappa et al., 2008). In addition, an ensemble radar precipitation (Germann et al., 2009) could be used to assess the measurement errors, and an observational precipitation ensemble (Ahrens et al., 2007) could be implemented to study the interpolation uncertainty. In parallel, using calibrated COSMO-LEPS rainfall forecasts (Fundel et al., 2010) to drive the hydrological and hydraulic model is planned. As this calibration method based on quantile mapping improves the reliability of precipitation forecasts, it is expected to improve the discharge forecasts as well. Note that alternative calibration methods for limited-area ensemble precipitation forecasts are currently investigated (Diomede et al., 2011).

22 - 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.

We propose to add the following text on page 741, line 2:

As presented in Section 3.1, a more robust assessment of the flood forecasting capacity of the system and the further development of an efficient decision-support system implies an enhancement of the reforecast period.

Once again, we would like to thank the referee for her/his inspiring comments. A new linguistic revision of the whole manuscript will be performed.

On behalf of all co-authors | N. Addor

References

Ahrens, B. and Jaun, S.: On evaluation of ensemble precipitation forecasts with observation-based ensembles, Adv. Geosci., 10, 139–144, www.adv-geosci.net/10/139/2007/, 2007.

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

Diomede, T., Marsigli, C., Montani, A., and Paccagnella, T.: Comparison of calibration techniques for a limited-area ensemble precipitation forecast using reforecasts, Geophysical Research Abstracts Vol. 13, EGU2011-7261, 2011 EGU General Assembly 2011.

Efron, B., 1992: Jacknife-after-bootstrap standard errors and influence functions. J. Roy. Stat. Soc., 54B, 83–127.

Germann, U., Berenguer, M., Sempere-Torres, D., and Zappa, M.: REAL - Ensemble radar precipitation for hydrology in a mountanious region, Q. J. Roy. Meteor. Soc., 135, 445–456. doi:/10.1002/qj.375, 2009.

Fundel, F., Walser, A., Liniger, M. A., Frei, C., and Appenzeller, C.: Calibrated precipitation forecasts for a limited-area ensemble forecast system using reforecasts, Mon. Weather Rev., 138, 176–189, doi:10.1175/2009MWR2977.1, 2010.

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

Lashley S.L., L. Fisher, B.J. Simpson, J. Taylor, S. Weisser, and J.A. Logs- don, A.M.

Lammers , 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, 2008.

Reggiani, P., Renner, M., Weerts, A. H., and van Gelder, P. A. H. J. M.: Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, Water Resour. Res., 45, W02428, doi:10.1029/2007WR006758, 2009.

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