Comparison of calibration techniques for a limited-area ensemble precipitation forecast using reforecasts

Tommaso Diomede, Chiara Marsigli, Andrea Montani, Tiziana Paccagnella
ARPA-SIMC, HydroMeteorological and Climate Service of the Emilia-Romagna Regional Agency for Environmental Protection, Bologna, Italy, tdiomede@arpa.emr.it

Abstract: The calibration of the precipitation forecasted at high resolution is currently a challenge for the ensemble community working with Limited Area Models. Here, the potential of using reforecasts to achieve this goal was investigated. Different calibration techniques were tested. The impact of the application of these techniques to the precipitation forecasts provided by a Limited-area Ensemble Prediction System was verified over the Emilia-Romagna Region (Northern Italy), Switzerland and Germany. The results revealed a beneficial impact of the calibration process for Switzerland and Germany; rather, no significant improvements were obtained for Emilia-Romagna. As the model error is likely to have a systematic dependence on geography, orography and flow direction, weather-regime dependent correction functions should be generated for improving the calibration strategy.

Keywords: calibration, precipitation forecast, ensemble, reforecasts, COSMO-LEPS

1. Introduction

The calibration of the precipitation forecasted at high resolution is currently a challenge for the ensemble community working with Limited Area Models, especially with respect to the improvement of the forecast skill for rare events. The potential of using reforecasts to achieve this goal has been shown in recent studies (Hamill et al., 2008; Fundel et al., 2010). Reforecasts mean a large dataset of retrospective forecasts obtained by the same model that is run operationally. In the present work, thirty years of reforecast of one member of COSMO-LEPS (the Limited-area Ensemble Prediction System based on the non-hydrostatic limited-area model COSMO) were used for the implementation of the calibration strategy over the Emilia-Romagna Region (Northern Italy), Switzerland and Germany. Three calibration techniques were tested: cumulative distribution function based corrections, linear regression and analogs. The choice of these methodologies is due to the need of improving the quantitative precipitation forecasts (QPFs) provided by COSMO-LEPS, especially as an input to hydrological models. Thus, techniques which enable a calibration of QPFs and not only of the probabilities of exceeding a threshold were selected.

2. Materials and Methods

The calibration strategy was based on the availability of historical forecast and observed rainfall data over the areas under investigation. Thirty years of reforecast of one member of COSMO-LEPS (10 km of horizontal resolution, 40 vertical levels) were run

by MeteoSwiss. One reforecast run with a 90-h lead time was available every three days from 1971 to 2000. This model climatology was used to calibrate forecasts of all lead times, without considering the time dependency of model bias (Fundel et al., 2010). According to the model climatology, the observed precipitation data were collected over the period 1971-2000 for Emilia-Romagna and Switzerland; rather, the observed data over Germany were available only for the period 1989-2000. The rainfall data were interpolated on the model grid points which cover the areas under investigation.
The calibration techniques tested in this work provide corrections based on the Cumulative Distribution Function (hereafter, CDF), the Linear Regression (hereafter, LR) and the Analog method. The described methodologies were used to calibrate each member of COSMO-LEPS. Each calibration function was defined by using the historical data forecasted and observed over each grid point for a specific season.
For the CDF method, the calibrated 24-h QPF was determined by comparing the reforecast and observed CDF curves. The value of the observed data which had the same probability of occurrence of the current 24-h QPF was used as the corrected QPF value. For the LR method, the parameters of the regression line estimated on the basis of reforecast and raingauge historical data were used to correct the current 24-h QPF value. The analog-based methodology was applied using two implementations, which differ from each other for the meteorological field used for the analog search. In the first implementation, the analog search was performed in terms of the similarity of the forecasted precipitation field over the area under investigation. In the second implementation, the analog search was performed in terms of the similarity of the forecasted circulation pattern, evaluated in terms of the geopotential at 700 hPa, 12 UTC (hereafter, Z700), over a spatial domain which is significant for the area under investigation to relate the synoptic circulation to the precipitation at ground. In the following of this paper, the first implementation of the analog-based method is referred to as “ANL” and the second implementation as “anlZ”. For both implementations, for each 24-h lead time, the root-mean-square (rms) differences between each member of the current forecast and each reforecast day were computed (the comparison was carried out among fields coming from the same season). The historical date with the smallest rms difference was chosen as the analog day, then the gridded raingauge recordings of that past day were used as the calibrated QPF.
The impact of the calibration process was verified for 24-h QPFs operationally provided by COSMO-LEPS in the years 2003-2007. The probabilistic verification was carried out in terms of the attributes diagram and the Brier Skill Score (BSS).

3. Results

The results obtained by the application of the calibration strategy are here discussed only for the autumn seasons in the years 2003-2007.
Figure 1 shows the attributes diagram for the lead time day 2. The verification was performed for each model grid point with respect to the ninety-fifth percentile of the climatological distribution of observed 24-h precipitation as threshold for the verified events. For Emilia-Romagna, the raw ensemble has no good reliability, providing overconfident forecasts. Only the calibration based on LR allows an increase of reliability. The weakness of the raw forecast system is more evident over Switzerland (i.e. the raw ensemble lies under the no skill line). The ensembles calibrated by the

CDF, LR and ANL methods show an increase of reliability; nevertheless these ensembles are still overconfident. For Germany, a beneficial impact is provided by the calibration based on LR, whereas a slight increase of reliability results for the ensembles calibrated by CDF and rainfall analogs.
Generally, the calibrated ensembles are still overconfident, especially for high probability values. The calibration based on the analogs of geopotential provides bad performance over all the three study areas. This result reveals that the geopotential at 700 hPa is not a good predictor for the precipitation over the selected areas.

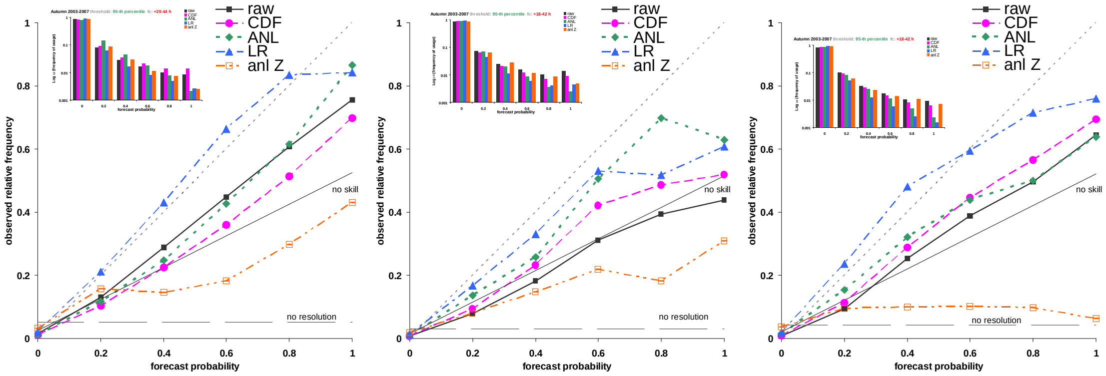

Figure 1: Attributes diagrams for the raw and calibrated ensembles over Emilia-Romagna (left panels), Switzerland (middle panels) and Germany (right panels) in autumn at day 2 lead time, for the 95-th percentile threshold. The inset histograms denote the frequencies of the use of the forecasts for each probability bin.

Figure 2 shows the results obtained in terms of BSS for the autumn season in the period 2003-2007 with respect to the ninety-fifth percentile of the observed climatology as threshold for the verified events. The observed climatology is used as the reference forecast for the computation of the skill score. The calibration process does not provide a beneficial impact on the ensemble QPFs over Emilia-Romagna. Actually, the values of BSS associated to the calibrated ensembles are lower than the BSS of the raw ensemble for all the lead times. The raw ensemble performs worse than climatology over Switzerland, but the calibration process provides the greater amount of skill improvement. With the exception of the anlZ method, the forecasts calibrated by all the methods show a significant increase of BSS values. Even, with respect to climatology, unskillful raw forecasts can be turned into skillful forecasts. In particular, the highest BSS values are provided by the ANL method. For Germany, a beneficial impact is provided by the CDF method for all the lead times; rather, slight improvements are obtained for the ensembles calibrated by LR and rainfall analog only for the longer lead times. Generally, the decay of performance with lead time is evident for the raw and calibrated forecasts.
An additional verification of the calibration process was performed by the coupling of the ensemble precipitation forecasts with an hydrological model. This test was carried out for the Reno river basin, a medium-sized catchment located in the Emilia-Romagna Region. The river hydrograph simulations were carried out for the autumn and spring seasons in the period 2003-2008 by using the distributed rainfall-runoff model TOPKAPI. The results of the coupling were evaluated in terms of missed events and false alarms which would have been issued based on the discharge scenarios driven by the raw and calibrated QPFs, with respect to the exceeding of the warning threshold

defined for the aims of civil protection. The results showed that, on the one hand, a beneficial impact on the reduction of missed events was provided by the calibration performed with the ANL and CDF methods. On the other hand, an increase of false alarms resulted by the application of the two above-mentioned calibration methods, even though this trend is evident for the ANL method only for longer lead times.

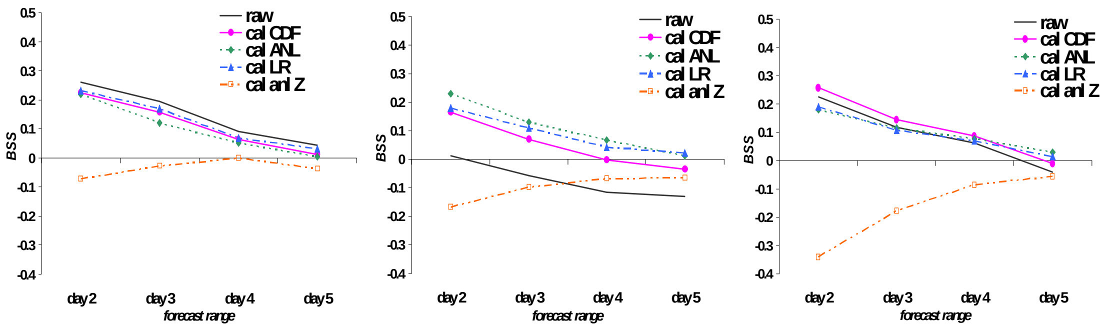

Figure 2: BSS for the raw and calibrated ensembles over Emilia-Romagna (panels on the left), Switzerland (panels in the middle) and Germany (panels on the right) in autumn, as a function of the forecast lead time. Skill at the 95-th percentile threshold.

4. Concluding remarks

The results revealed a beneficial impact of the calibration process over Switzerland and Germany. No significant improvements were obtained over Emilia-Romagna by evaluating the statistical analysis on the calibrated QPFs. The coupling of the QPFs calibrated with the ANL and CDF methods with an hydrological model revealed a beneficial impact of the calibration on the reduction of missed events for a medium-sized catchment (i.e. the Reno river basin) used as a test-bed. The lack of a remarkable improvement, especially over Emilia-Romagna, resulting from the application of the proposed calibration methods suggests the need of defining specific correction functions which should be able to link the model errors to the meteorological situation. Actually, the search for a unique relationship between forecast and observed data hampers to highlight the model errors which are known to have a systematic dependence on geography, orography and flow direction. Therefore, the calibration strategy should be improved by dividing the training sample size in order to pool data which have similar model errors with respect to a given meteorological situation.

References

Fundel F., Walser A., Liniger M.A., Frei C., Appenzeller C. (2010) Calibrated Precipitation Forecasts for a Limited Area Ensemble Forecast System Using Reforecasts, *Mon. Weather Rev.*, 138, 176-189.

Hamill T.M., Hagedorn R., Whitaker J.S. (2008) Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation, *Mon. Wea. Rev.*, 136, 2620–2632.