

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

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

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). 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 testing the calibration strategy. Three calibration techniques were considered: Cumulative Distribution Function, Linear Regression and Analogs. The analog-based method was implemented in terms of the similarity of precipitation field, as well as in terms of the similarity of upper air fields representing the synoptic pattern.

The impact of the application of these techniques to the ensemble precipitation forecasts operationally provided by COSMO-LEPS (10 km of horizontal resolution) in the years 2003-2007 was verified over the Emilia-Romagna Region (Northern Italy), Switzerland and Germany. The calibration of COSMO-LEPS is a part of the project CONSENS (CONSolidation of COSMO ENSemble), developed within the framework of COSMO (CONsortium for Small-scale MOdeling), which aims at consolidating the COSMO ensemble forecasting systems for the mesoscale.

2. Data and methodology

The calibration strategy was based on historical forecast and observed rainfall data available over the areas under investigation. Thirty years of reforecast of one member of COSMO-LEPS 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 is 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 using the nearest neighbour method for Emilia-Romagna, the SYMAP method (Frei et al. 2006) for Switzerland and the inverse-squared-distance weighting method based on the nearest three gauges for Germany. The 24-h rainfall amounts span the daily time window 08:00-08:00 UTC for Emilia-Romagna and the daily time window 06:00-06:00 UTC for Switzerland and Germany. 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 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. The described methodologies were used to calibrate each member of COSMO-LEPS. The basic implementations of the CDF and LR methods were developed by generating a seasonal correction function for each model grid point. Each function was defined by using the historical data forecasted and observed over each grid point for that specific season.. This approach generates correction functions that are location (i.e. grid point) specific. 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 about 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 fields over the area under investigation. In the second implementation, the analog search was performed in terms of the similarity of the forecasted circulation patterns, 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 “anIZ”. In both the implementation, 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, and averaged over the grid points. Then, the historical date with the smallest rms difference was chosen as the date of the analog day, then the gridded raingauge recordings of that past day were used as the calibrated QPF value. The effect 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). The verification was performed for each grid point with respect to percentiles of the climatological distribution of observed 24-h precipitation as threshold for the verified events.

3. Results

3.1 Comparison of the methodologies over Emilia-Romagna, Switzerland and Germany

The statistical analyses, evaluated for different seasons, thresholds and forecast ranges, reveal an increase of forecast reliability and skill for the calibrated ensembles over Switzerland and Germany, except when the anIZ method is applied. However, the skill improvement resulting from the forecast calibration is small for Germany. No significant improvements of forecast reliability and skill are obtained for the calibrated ensembles over Emilia-Romagna; even, a worsening of the ensemble performance is obtained for the autumn season. Generally, the raw and calibrated ensembles provide overconfident forecasts and the calibrated forecasts show more reliability at lower probabilities. The greatest amount of skill improvement resulting from forecast calibration is obtained over Switzerland. The choice of the best methodology depends on the area under investigation. On the one hand, the ANL method provides the best performance over Switzerland; on the other hand, the CDF and LR methods are to be preferred for Germany. For Emilia-Romagna, the decision about the best

calibration method is not definite, depending on the threshold of the verified event. Generally, forecasts of lower precipitation events are more skilful than forecasts of higher precipitation events. A performance decay with the increase of lead time is evident. Comparing the performance over Emilia-Romagna, Switzerland and Germany, the raw forecasts show more skill in autumn, spring and winter over Emilia-Romagna and Germany than over Switzerland. Rather, the performance for summer is better over Switzerland than over Emilia-Romagna and Germany. The calibrated forecasts has more skill over Switzerland and, with a lesser extent, Germany than over Emilia-Romagna. The Figures 1 and 2 display results only for the autumn and spring seasons, considering the 80-th percentile of the 24-h climatological rainfall amount as verification threshold.

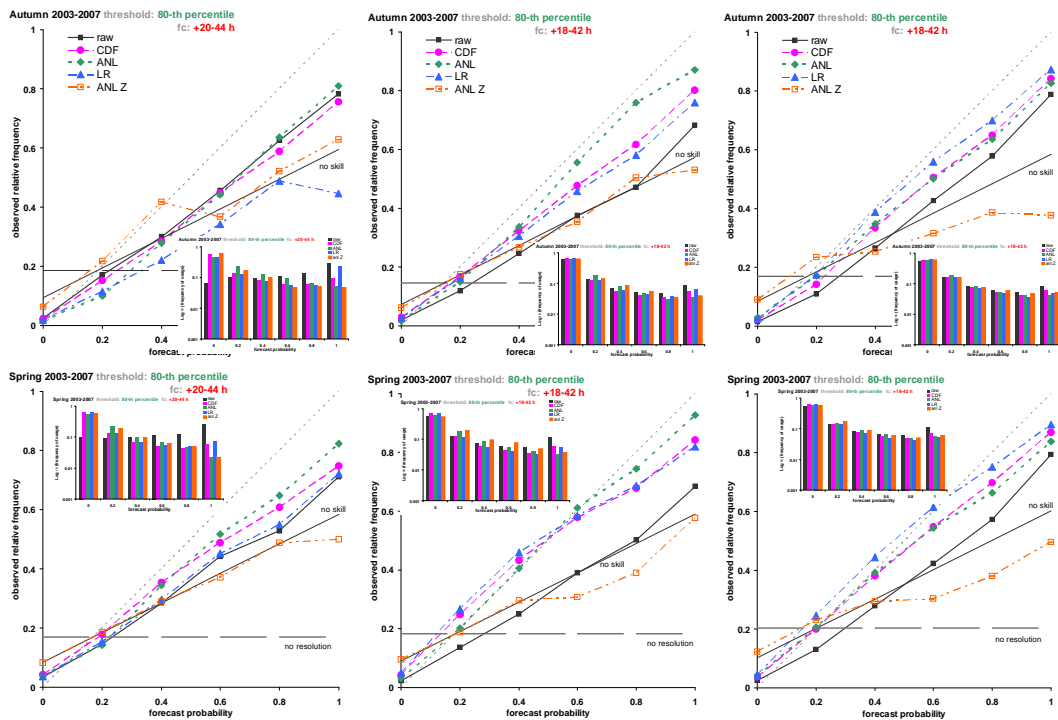


Fig 1. Attributes diagrams for the raw and calibrated ensembles over Emilia-Romagna (left panels), Switzerland (middle panels) and Germany (right panels) in autumn (first row panels) and spring (second row panels) at day 1 lead time, for the 80-th percentile threshold. The solid black line with squares is for the raw ensemble, the dashed fuchsia line with circles is for the ensemble calibrated by CDF, the dotted dark green line with diamonds is for the ensemble calibrated by ANL, the dot-dashed blue line with triangles is for the ensemble calibrated by LR and the double dot-dashed orange line with empty squares is for the ensemble calibrated by anlz. The inset histograms denote the frequencies of the use of the forecasts for each probability bin.

3.2 Testing weather-specific correction functions over Emilia-Romagna

The lack of a remarkable improvement (especially over Emilia-Romagna) resulting from the application of the proposed calibration methods suggested the need of defining specific correction functions which should be able to link the model error to the meteorological situation. Actually, the poor performance of calibration can be ascribed to the lack of a strong relationship between forecast and observed data, which hampers to highlight a model error which has a systematic dependence on geography, orography and flow direction. To overcome this problem, the use of calibration functions which are weather-regime dependent is proposed. In particular, an easy way of doing this has been tested, building calibration functions which were specific for underestimation and overestimation model conditions.

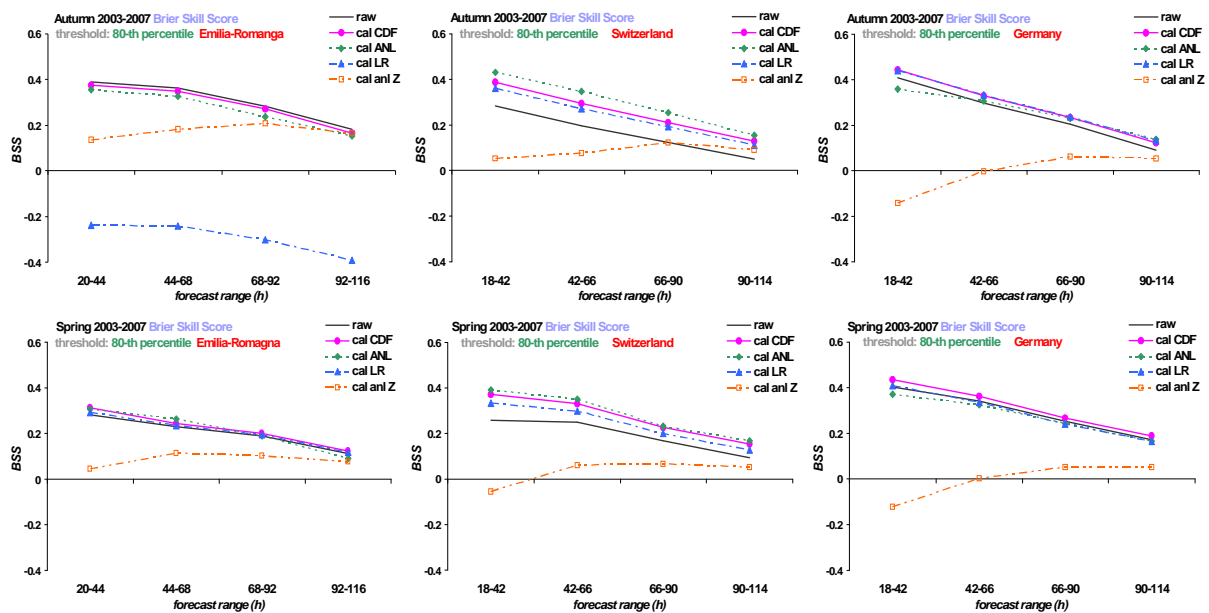


Fig 2. BSS for the raw and calibrated ensembles over Emilia-Romagna (panels on the left), Switzerland (panels in the middle) and Germany (panels on the right) in autumn (first row panels) and spring (second row panels), as a function of the lead time of the forecast. Skill at the 80-th percentile threshold. The solid black line is for the raw ensemble; the solid fuchsia line with circles is for the ensemble calibrated by CDF; the dotted green line with diamonds is for the ensemble calibrated by ANL; the dot-dashed blue line with triangles is for the ensemble calibrated by LR and the double dot-dashed orange line with empty squares is for the ensemble calibrated by anI Z.

The basic idea of this approach was that, for each model grid point and season, the training sample of 24-h rainfall forecast and observation was divided in order to pool data which correspond to, respectively, underestimated and overestimated forecast. Then, a criterion needed for evaluating if the actual forecast corresponds to an underestimation or overestimation model condition. Finally, the corresponding calibration function was applied.

A preliminary test on such a approach was carried out for the Emilia-Romagna case study. The calibration functions used for this test were based on the LR method. Different predictors were tested for the evaluation of the underestimated or overestimated model condition. The selected predictors were: the best analog in terms of rainfall field, Z700, Q700, the combination Z700-Q700, and the mean fields of Z700 and Q700 (specific for each flow regime). For the approach based on analogy, given a certain lead time, the best analog of the actual forecast is searched in terms of the selected predictor. For each model grid point, the forecasted and observed precipitation data of each past analog day were compared to determine if the forecast was overestimated or underestimated. Then, the current forecast is evaluated as an overestimation or underestimation model condition on the basis of what occurred in the selected past analog days. Finally, the specific calibrating function corresponding to the forecasted model condition was used. For the approach based on the mean field of Z700 (or Q700), the current forecast of Z700 (or Q700) was compared, separately, against the mean fields of Z700 (or Q700) corresponding to, respectively, the historical overestimated and underestimated forecasts occurred for a fixed flow regime. Unfortunately, the results are not satisfactory (Fig 3), as the performances of the calibrated ensembles are generally quite worse or similar to the performances of the raw ensemble. Therefore, the selected predictors seem to be not useful within such an approach for evaluating if the meteorological model is currently providing an underestimated or overestimated

24-h rainfall forecast. Nevertheless, the results obtained on the basis of a perfect forecast about the under- or over-estimation model condition are quite encouraging, showing that there is room for the improvement of the forecast by means of such a calibration approach (Fig. 3, lines labelled with “2LR PP” in the graph’s legend).

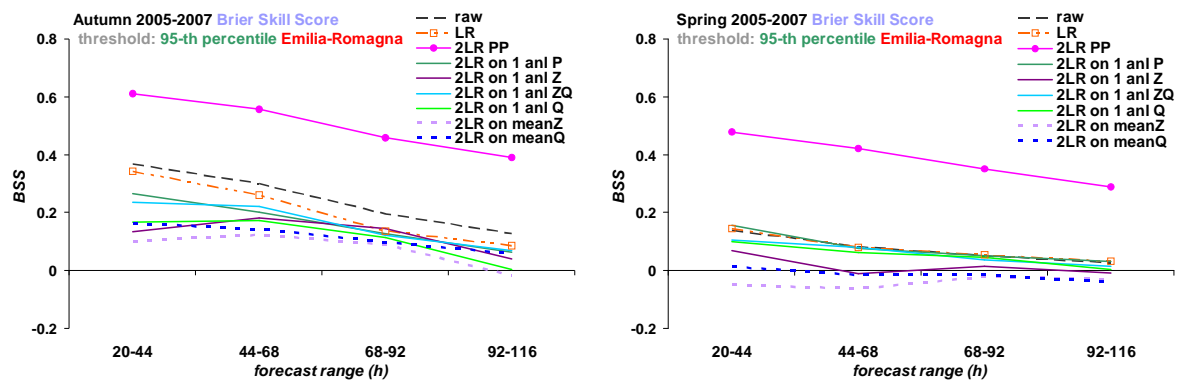


Fig 3. BSS of the raw ensemble and the ensembles calibrated by different weather-regime dependent implementations of the LR method over Emilia-Romagna, as a function of the forecast lead time. Skill at the 95-th percentile threshold. Results refer to the autumn (left panel) and spring (right panel) in the years 2005-2007, as Q700 was available only for such a period in the archive of COSMO-LEPS. The different implementations of the LR method are based on different predictors used for evaluating if the current forecast is providing an underestimated or overestimated 24-h rainfall amount: observations (2LR PP); best analog of rainfall field (2LR on 1 anlP), Z700 (2LR on 1 anlZ), combination Z700-Q700 (2LR on 1 anlZQ), Q700 (2LR on 1 anlQ); similarity of the current Z700 (2LR on meanZ) and Q700 (2LR on meanQ) fields to the corresponding mean field for a specific flow regime.

Conclusions

The statistical analyses revealed a positive impact of the calibration process over Switzerland and Germany; rather, no significant improvements resulted over Emilia-Romagna. The choice of the best calibration method depends on the area under investigation. Results suggested the need of generating correction functions which are weather-regime dependent, as the model error is likely to have a systematic dependence on geography, orography and flow direction. As a future development, a multi-variable approach will be tested for improving this evaluation. This proposed approach should be based on the evaluation of upper air fields at different levels and times by means of a Principal Component Analysis, in order to find a better matching of the space-time evolution of the synoptic pattern.

References

- Frei, C., Schöll, R., Fukutome, S., Schmidli, J., Vidale, P. L., 2006: Future change of precipitation extremes in Europe: Intercomparison of scenarios from regional climate models. *J. Geophys. Res.*, **111**, D06105.
- 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.