Ensemble model output statistics for temperature forecasts in Veneto

Ensemble model output statistics per la previsione delle temperature in Veneto

Gaetan Carlo, Giummolè Federica, Mameli Valentina, Siad Si Mokrane

Abstract Post-processing methods are nowadays widely used for limiting the impact of errors in ensemble forecast of meteorological variables. Ensemble model output statistics are an easy-to-apply technique for post-processing, based on a linear regression model. In this paper we use an ensemble model output statistic for the forecast of daily maximum temperatures in Veneto. We calculate estimative and calibrated predictive distributions for a time period of three years. We then compare the different predictive distributions by means of the log-score, the continuous ranked probability score and the coverage of the corresponding predictive quantiles. We show that the calibrated approach improves on the estimative ones as regards both mean scores and coverage probabilities.

Abstract *Al giorno d'oggi, metodi di post-processing vengono ampiamente utilizzati per limitare l'impatto di errori nelle previsioni di variabili meteorologiche che utilizzano ensemble. Gli ensemble model output statistics sono una tecnica di post-processing basata sul modello di regressione lineare. Nel nostro lavoro li abbiamo utilizzati per la previsione delle temperature massime giornaliere in Veneto. Abbiamo calcolato diverse distribuzioni predittive estimative e calibrate per un periodo di tre anni e le abbiamo confrontate fra loro utilizzando log-score, continuous ranked probability score e probabilità di copertura dei rispettivi quantili. I risultati mostrano la superiorità delle distribuzioni predittive calibrate rispetto alle corrispondenti distribuzioni estimative, secondo tutti i criteri considerati.*

Key words: Calibration, Coverage probability, Ensemble model output statistics, Post-processing, Predictive distributions, Scoring rules, Weather forecast.

Gaetan Carlo, Giummolè Federica, Siad Si Mokrane
Ca' Foscari University of Venice, Via Torino 155 Mestre - Venezia, e-mail: gaetan@unive.it, giummole@unive.it, simokrane.siad@unive.it

Valentina Mameli
University of Udine, Via Tomadini 30 Udine, e-mail: valentina.mameli@uniud.it

1 Introduction

Weather predictions are a critical component of decision support for a wide variety of applications. As a result, there is a significant demand for accurate weather predictions from a diverse variety of stakeholders, including the general public, the corporate sector, and government agencies that issue weather warnings. Forecasts have gradually improved over the previous decades, partly due to breakthroughs in numerical weather prediction (NWP) ([1]; [8]). The advancement of powerful high-performance computers enables more detailed weather simulations. Despite these advances in NWP, forecasts generated by physics-based models, usually provided as forecast ensembles, exhibit systematic bias and are often underdispersive ([2], [7]). Statistical post-processing methods are nowadays widely used for further refining, improving, and calibrating NWPs. The ensemble model output statistics (EMOS, [5]) is among one of the most popular post-processing techniques used to calibrate the ensemble forecasts. EMOS is based on a simple regression model with parameters depending on the ensemble forecasts. It is able to correct for systematic biases and dispersion errors.

This article proposes an adjustment of EMOS based on a bootstrap calibration procedure proposed by [3] and further extended to the EMOS context in [4]. In order to evaluate the performance of the calibrated EMOS and develop a comparison with the classic EMOS, we consider a real case study dealing with maximum daily temperatures at measurement sites located in the Veneto region, northern Italy. The evaluation and comparison of methods are based on measures of goodness for calibration and sharpness, the most desirable properties that characterise predictive models [6]. In particular, we compare different predictive models by means of the log-score, the continuous ranked probability score (CRPS) and the coverage of the corresponding predictive quantiles. These analyses exhibit the accuracy of the calibrated EMOS, with respect to the classic EMOS, in the presented application, and highlight the great potentiality of this new technique to provide calibrated and sharp predictive models.

The paper is organised as follows. In Section 2 we outline the methodology employed in this research. In Section 3 we introduce the data and we assess the performance of the calibrated EMOS in comparison with the classic EMOS. Finally, in Section 4 we present some concluding remarks.

2 The method

The most popular post-processing techniques used to calibrate the ensemble forecasts are EMOS that allow for probabilistic forecasts of weather variables ([5]) in the form of Gaussian predictive distributions. More formally, it is assumed that a weather variable Z depends on the ensemble forecasts $X_1,\dots,X_m$ in such a way that its mean is equal to $\beta_0+\beta_1X_1+\dots+\beta_mX_m$ and its variance is equal to $\gamma+\delta S^2$, where $S^2=\frac{1}{m-1}\sum_{i=1}^m(X_i-\bar{X})^2$ denotes the ensemble variance and $\beta_1,\dots,\beta_m$, γ and

δ are non-negative unknown coefficients. Therefore, under normality assumptions, the distribution of Z is $N(\beta_0+\beta_1 X_1+\ldots+\beta_m X_m, \gamma+\delta S^2)$.

For fitting the unknown EMOS coefficients, $\beta_0, \beta_1, \ldots, \beta_m, \gamma, \delta$, [5] proposed to use the method of minimum CRPS estimation, which finds the estimates by minimising the CRPS. The CRPS estimates are then substituted to the unknown parameters in the EMOS normal distribution obtaining what is known as an estimative distribution for the future weather quantity Z. In the introduction we have referred to this procedure as the classic EMOS. Here we also consider for comparison the EMOS estimative distribution obtained with the maximum likelihood estimates (MLE), that minimises the log-score. Unfortunately, estimative distributions may perform poorly, especially when the sample size is small with respect to the number of members of the ensemble. In particular, the important requirement of calibration is hardly attained by estimative distributions. As explained in [6], calibration is a sort of consistency between a predictive distribution and future observations and it can be graphically assessed by means of the histogram of the probability integral transformed (PIT) values. Moreover, a well calibrated predictive distribution provides prediction limits with coverage probability close to the nominal value.

In order to obtain calibrated predictive distributions, we consider the approach presented by [3] and recently adapted to the EMOS context by [4] that propose a simple simulation experiment. Suppose that we want to make prediction on the unobservable variable Z with distribution $G(z;\theta)$ depending on an unknown parameter θ. In our case $Z \sim N(\beta_0+\beta_1 X_1+\ldots+\beta_m X_m, \gamma+\delta S^2)$, so that $\theta=(\beta_0, \beta_1, \ldots, \beta_m, \gamma, \delta)$. Let $\hat{\theta}$ be the MLE for θ or an asymptotically equivalent alternative, such as the CRPS estimator, based on an observed sample.

A suitable adjustment of the classic EMOS that yields calibrated predictive distributions can be obtained using the following simple bootstrap procedure:

$$G^{boot}(z;\hat{\theta}) = \frac{1}{B}\sum_{b=1}^{B} G\{z_\alpha(\hat{\theta}^b);\hat{\theta}\}|_{\alpha=G(z;\hat{\theta})}, \tag{1}$$

where $z_\alpha(\theta)$ denotes the α-quantile of Z and $\hat{\theta}^b$, $b=1,\ldots,B$, are the estimates obtained by a parametric bootstrap from the estimative distribution of Z. The goodness of the approximation depends on the efficiency of the bootstrap simulation procedure. We also note that in the case under study it is straightforward to draw random samples and calculate quantiles and probabilities from normal estimative distributions. This makes the bootstrap calibration procedure very easy to apply for improving on EMOS estimative distributions.

3 A case study

For this study, we focus on maximum daily temperature forecasts at surface stations in the Veneto region, northern Italy. Our period of interest spans 3 years from 16 August 2009 to 17 August 2012. Historical maximum daily temperature forecasts were acquired from the website of the Italian national system for the collection, processing and dissemination of climate data, created by ISPRA (http://www.scia.isprambiente.it/). Ensemble predictions were available from World Climate Research Programme. We use the Coupled Model Intercomparison Project Phase 6 system (CMIP6) that consists of over 100 models of which 29 ensemble members cover the Veneto region. The data in the CMIP6 archive are downscaled to station level using elevation as a base for interpolation. In order to assess and compare the performance of different EMOS predictive distributions, we analyse maximum daily temperatures for the station Cavallino-Treporti (Longitude: 12.48642°, Latitude: 45.45805°), located on the Venetian lagoon. After removing missing observations from the selected station, the sample contains 1091 observations of daily temperatures and 26 ensemble members. According to [5], we use a sliding window of 40 observations as training set, with the remaining 1051 days available as test set. In this temporal window the generating process is supposed to be stationary. First, the EMOS parameters are estimated by optimising both the log-score and the CRPS over the sliding training period. Then, for each of the 1051 days available as test set, the performance of the two estimative distributions obtained by estimating with the log-score and the CRPS, and the corresponding calibrated distributions, are evaluated by means of different measures of goodness for calibration and sharpness. Table 1 summarises the mean and standard deviations of the log-score and the CRPS for the four predictive models. The superior performance of the calibrated models is reflected in the smaller values of the two scores.

Table 1 Log-score and CRPS values of the four predictive distributions. Standard errors in brackets. Est log denotes the estimative EMOS with log-score estimates and Est CRPS the estimative EMOS with CRPS estimates, while Cal log and Cal CRPS are the respective calibrated counterparts.

	Est log	Cal log	Est CRPS	Cal CRPS
Log-score	3.22	2.51	3.65	2.58
	(0.09)	(0.03)	(0.14)	(0.06)
CRPS	1.79	1.64	1.82	1.66
	(0.05)	(0.04)	(0.05)	(0.04)

For properly measuring calibration and sharpness (concentration) of the different predictive models, we have also obtained coverage probabilities and mean lengths of central intervals of level 66.7% (Table 2). Moreover, the coverage probability of upper prediction limits of levels 90%, 95% and 99% for the different predictive models have been obtained, being fundamental quantities for the evaluation of environmental risk (Table 3). The effect of calibration can be clearly seen from coverage probabilities for calibrated predictive models, much closer to target values than those for the estimative ones. Of course, shorter central prediction intervals for the estimative models are due to the loss of corresponding coverage.

Finally, Figure 1 shows the PIT histograms for the four considered predictive models. The effect of calibration can be observed in the flat histogram very close to the uniform one for the calibrated predictive models. Instead the U-shaped histograms of the estimative models are the consequence of the excessive under-dispersion.

Table 2 Coverage probabilities and mean lengths of the central prediction interval of level 66.7% for the estimative predictive models with MLE and CRPS estimates, and corresponding calibrated models. Standard errors in brackets. Est log denotes the estimative EMOS with MLE estimates and Est CRPS the estimative EMOS with CRPS estimates, while Cal log and Cal CRPS are the respective calibrated counterparts.

	Est log	Cal log	Est CRPS	Cal CRPS
α=0.667	0.402	0.646	0.375	0.648
	(0.015)	(0.015)	(0.015)	(0.015)
Mean length	3.197	5.419	2.905	5.334
	(0.023)	(0.040)	(0.024)	(0.044)

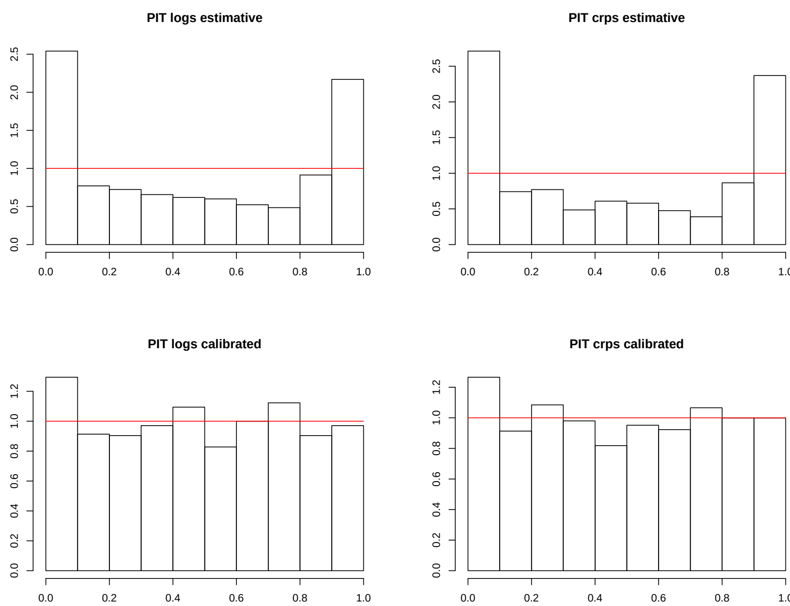

Fig. 1 PIT histograms of the estimative predictive models based on the MLE and the CRPS estimates, and the corresponding calibrated models.

Table 3 Coverage probabilities of upper prediction limits for the estimative predictive models with MLE and CRPS estimates, and the corresponding calibrated models. Standard errors in brackets. Est log denotes the estimative EMOS with MLE estimates and Est CRPS the estimative EMOS with CRPS estimates, while Cal log and Cal CRPS are the respective calibrated counterparts.

α	Est log	Cal log	Est CRPS	Cal CRPS
0.90	0.783	0.903	0.763	0.900
	(0.013)	(0.009)	(0.013)	(0.009)
0.95	0.835	0.952	0.814	0.952
	(0.011)	(0.007)	(0.012)	(0.007)
0.99	0.912	0.991	0.882	0.990
	(0.009)	(0.003)	(0.010)	(0.003)

4 Conclusions

This work proposes a comparison between the classic EMOS and a calibrated EMOS based on a bootstrap procedure, applied to a real case study of temperatures in the Veneto region, Italy. In particular, we have used measurements from a single station, Cavallino-Treporti, in the Venice lagoon. Future work will consider a spatial analysis of temperature data collected all over the Veneto region, including the coast around the Venice lagoon, the low Venetian plan and the Dolomite mountains. Indeed, the Veneto region's height fluctuates from sea level (and also below sea level) up to about 3,300 m, with a corresponding wide variation in the temperatures.

References

1. Bauer, P., Thorpe, A., Brunet, G.: The quiet revolution of numerical weather prediction. Nature, **525**, 47–55 (2015).
2. Buizza, R.: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. Mon. Wea. Rev., **125**, 99–119 (1997).
3. Fonseca, G., Giummolè, F., Vidoni, P.: Calibrating predictive distributions. Journal of Statistical Computation and Simulation, **84**, 373–383 (2014).
4. Giummolè, F., Mameli, V.: Comparing predictive distributions in EMOS. Proceedings of the Italian Statistical Society, 1–6 (2020).
5. Gneiting, T., Raftery, A.E., Westveld III, A.H., Goldman , T.: Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CPRS Estimation. Monthly Weather Review, **133(5)**, 1098–1118 (2005).
6. Gneiting, T., Balabdaoui, F., Raftery, A.E.: Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society: Series B (Statistical Methodology), **69**, 243–268 (2007).
7. Haiden, T., Janousek, M., Vitart, F., Ferranti, L., Prates, F.: Evaluation of ECMWF forecasts, including the 2019 upgrade. ECMWF Tech. Memo. **588**, 56 pp. (2019).
8. Raftery, A., Gneiting, T., Balabdaoui, F., Polakowski, M.: The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. Quart. J. Roy. Meteor. Soc., **145**, 12–24 (2019).