Comparing predictive distributions in EMOS

Distribuzioni predittive per modelli EMOS

Giummolè Federica and Mameli Valentina

Abstract EMOS models are widely used post-processing techniques for obtaining predictive distributions from ensembles for future weather variables. A predictive distribution can be easily obtained by substituting the unknown parameters with suitable estimates in the distribution of the future variable, thus obtaining a so called estimative distribution. Nonetheless, these distributions may perform poorly in terms of coverage probability of the corresponding quantiles. In this work we propose the use of calibrated predictive distributions in the context of EMOS models. The proposed calibrated predictive distribution improves on estimative solutions, producing quantiles with exact coverage level. A simulation study assesses the goodness of the calibrated predictive distribution in terms of coverage probabilities and also logarithmic score and CRPS.

Abstract *I modelli EMOS forniscono un metodo per ottenere distribuzioni predittive a partire da un insieme di previsioni per una variabile meteorologica di interesse. Una distribuzione predittiva si può ottenere facilmente sostituendo i parametri non noti con delle stime opportune nella distribuzione della variabile futura. Questa procedura dà origine alle cosiddette distribuzioni estimative che però spesso risultano inadeguate in quanto la probabilità di copertura associata ai loro quantili differisce da quella nominale. In questo lavoro proponiamo, nel contesto dei modelli EMOS, una distribuzione predittiva calibrata che fornisce quantili la cui probabilità di copertura coincide con quella nominale. Uno studio di simulazione evidenzia la bontà della predittiva proposta, sia in termini di probabilità di copertura che rispetto alle funzioni di perdita logaritmica e CRPS.*

Key words: Coverage probability, CRPS, EMOS, logarithmic score, predictive distribution.

Giummolè Federica
Ca' Foscari University of Venice, Via Torino 155 Mestre, e-mail: giummole@unive.it

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

1 Introduction

In modern society, weather conditions have wide-ranging economic impacts in fields as diverse as aviation, shipping, tourism and agriculture, just to name a few. All these important applications require accurate forecasts of future weather conditions. Weather forecasts are usually provided as forecast ensembles obtained from multiple numerical models achieved using different initial conditions and different numerical representations of the atmosphere [8]. However, such ensemble forecasts are able to capture only part of the forecast uncertainty exhibiting dispersion errors and systematic biases [7], [2]. For this reason, ensemble forecasts are often statistically post-processed to produce calibrated predictive distributions. Many statistical post-processing methods have been proposed in the literature. The most popular are the ensemble model output statistics (EMOS) that allow for probabilistic forecasts of continuous weather variables ([4]). EMOS is nothing but a linear regression model with heteroschedastic Gaussian errors. The EMOS mean is a linear combination of the ensemble member forecasts, with unknown coefficients that represent the contributions of each member of the ensemble to the interest weather variable. The EMOS variance is a linear function of the ensemble variance that accounts for spread relationship. More precisely, it is assumed that a weather continuos variable Y 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 $\beta_0,\dots,\beta_m$, $\gamma>0$ and $\delta>0$ are unknown coefficients. Under normality assumptions, the distribution of Y is $N(\beta_0+\beta_1X_1+\dots+\beta_mX_m,\gamma+\delta S^2)$. Suitable estimates are then substituted to the unknown parameters, obtaining what is known as an estimative distribution for the future weather quantity Y:

$$N(\hat{\beta}_0+\hat{\beta}_1X_1+\dots+\hat{\beta}_mX_m,\hat{\gamma}+\hat{\delta}S^2),$$

where $\hat{\beta}_0,\hat{\beta}_1,\dots,\hat{\beta}_m,\hat{\gamma}$, and $\hat{\delta}$ are suitable estimates of $\beta_0,\beta_1,\dots,\beta_m,\gamma$, and δ, respectively.

Unknown parameters can be estimated using the method of maximum likelihood or of minimum Continuous Ranked Probability Score (CRPS), which respectively optimise the logarithmic score and the CRPS, see [5, 6].

Unfortunately, an estimative distribution can lead to poor prediction statements, since it does not take into account for the uncertainty introduced by substituting estimates to the true parameter values. In particular, the coverage of prediction intervals obtained by the estimative distribution does not achieve the nominal coverage level, see [1, 3].

In this work we recommend for the EMOS model the use of a calibrated predictive distribution based on a bootstrap procedure proposed by [3], which improves on the estimative solution. On a simulation study, we compare the Gaussian estimative distributions obtained with minimum CRPS estimates and maximum likelihood estimates with their calibrated counterparts. We show that the calibrated predictive distributions always improve on the estimative ones in terms of coverage of prediction intervals and of logarithmic score. As regards the CRPS, all the considered distributions perform in a similar way.

2 Calibrated predictive distributions

In this section we briefly review, in the context of EMOS models, the calibrating approach proposed by [3], which provides predictive distributions whose quantiles give well-calibrated coverage probability.

Suppose that $\{Y_i\}_{i\geq 1}$ is a sequence of independent continuous random variables with probability distribution specified by the EMOS model:

$$Y_i \sim N(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im}, \gamma + \delta s_i^2), \quad i \geq 1,$$

where x_{ij} is the i-th value of the ensemble variable X_j, $j = 1, \ldots, m$, s_i^2 is the ensemble variance of $x_{i1}, \ldots, x_{im}$, and $\theta = (\beta_0, \beta_1, \ldots, \beta_m, \gamma, \delta)$ is the unknown parameter vector. We assume that $Y = (Y_1, \ldots, Y_n)$, $n > 1$, is observable, while $Z = Y_{n+1}$ is a future or not yet available variable. We indicate with $\Phi(z; \theta)$ the Gaussian cumulative distribution function of Z.

Given the observed sample $y = (y_1, \ldots, y_n)$, an α-prediction limit for Z is a function $c_\alpha(y)$ such that, exactly or approximately,

$$P_{Y,Z}\{Z \leq c_\alpha(Y); \theta\} = \alpha, \quad (1)$$

for every $\theta \in \Theta$ and for every fixed $\alpha \in (0, 1)$. The above probability is called coverage probability and it is calculated with respect to the joint distribution of (Z, Y).

Consider a suitable asymptotically efficient estimator $\hat{\theta} = \hat{\theta}(Y)$ for θ and the estimative prediction limit $z_\alpha(\hat{\theta})$, which is obtained as the α-quantile of the estimative distribution function $\Phi(\cdot; \hat{\theta})$. The associated coverage probability is

$$P_{Y,Z}\{Z \leq z_\alpha(\hat{\theta}(Y)); \theta\} = E_Y[\Phi\{z_\alpha(\hat{\theta}(Y)); \theta\}; \theta] = C(\alpha, \theta) \quad (2)$$

and, although its explicit expression is rarely available, it is well-known that it does not match the target value α even if, asymptotically, $C(\alpha, \theta) = \alpha + O(n^{-1})$, as $n \to +\infty$, see e.g. [1]. As proved in [3], the function

$$G_c(z; \hat{\theta}, \theta) = C\{\Phi(z; \hat{\theta}), \theta\}, \quad (3)$$

which is obtained by substituting α with $\Phi(z; \hat{\theta})$ in $C(\alpha, \theta)$, is a proper predictive distribution function, provided that $C(\cdot, \theta)$ is a sufficiently smooth function. Furthermore, it gives, as quantiles, prediction limits $z_\alpha^c(\hat{\theta}, \theta)$ with coverage probability equal to the target nominal value α, for all $\alpha \in (0, 1)$.

The calibrated predictive distribution (3) is not useful in practice, since it depends on the unknown parameter θ. However, a suitable parametric bootstrap estimator for

$G_c(z;\hat{\theta},\theta)$ may be readily defined. Let y^b, $b=1,\dots,B$, be parametric bootstrap samples generated from the estimative distribution of the data and let $\hat{\theta}^b$, $b=1,\dots,B$, be the corresponding estimates. Since $C(\alpha,\theta)=E_Y[\Phi\{z_\alpha(\hat{\theta}(Y));\theta\};\theta]$, we define the bootstrap calibrated predictive distribution as

$$G_c^{boot}(z;\hat{\theta})=\frac{1}{B}\sum_{b=1}^{B}\Phi\{z_\alpha(\hat{\theta}^b);\hat{\theta}\}|_{\alpha=\Phi(z;\hat{\theta})}. \tag{4}$$

The corresponding α-quantile defines, for each $\alpha\in(0,1)$, a prediction limit having coverage probability equal to the target α, with an error term which depends on the efficiency of the bootstrap simulation procedure.

3 A simulation study

In order to assess and compare the performance of the estimative and the calibrated predictive distributions for the EMOS model we perform several experiments with simulated ensembles. The ensemble members are drawn from a m-variate normal distribution with zero mean and identity covariance matrix, with $m=5,10,15$. The i-th observation is generated from a normal random variable with mean $\beta_0+\sum_{j=1}^{m}\beta_j x_{ij}$ where $\beta_j=(j+1)/\sum_k\beta_k$, $j=0,\dots,m$, and variance $\gamma+\delta s_i^2$, with $\gamma=0$ and $\delta=1$, $i=1,\dots,n$ with $n=20$. The bootstrap procedure is based on 500 bootstrap samples. The estimation is based on 1000 Monte Carlo replications. We evaluate the estimative and calibrated predictive distributions in terms of coverage probabilities and also using the logarithmic score and CRPS as loss functions, as commonly used in the literature, see [4, 9]. It should be noted that the calibration procedure is based on asymptotic considerations. Thus the improvement over estimative results is more evident with small sample sizes. Here we have chosen $n=20$, having in mind to use the 20 most recent daily observations of a meteorological variable as the training period for estimating the EMOS model parameters, see [4].

Table 1 provides the results of a simulation study for comparing coverage probabilities of 66.7% and 90% central prediction intervals obtained from the estimative and the calibrated distributions with minimum CRPS and maximum likelihood estimates. For this aim, we consider prediction limits of levels $\alpha=0.05,0.167,0.833$, and 0.95. It can be noted that the coverage probabilities associated to the calibrated quantiles almost equal the nominal values, showing accurate coverage. Average width of prediction intervals, not shown here, demonstrates that the calibrated predictive distributions yield slightly longer prediction intervals with respect to the estimative ones, but this can be explained by the greater coverage of these prediction intervals.

We assess the improvement of the calibrated predictive distributions over the estimative ones by computing also the logarithmic score and the CRPS, averaged over 1000 replicates, as shown in Table 2. The superior performance of the calibrated distributions is reflected in the values of the logarithmic score. Indeed average values of the logarithmic score for estimative distributions are significantly worse with respect to their calibrated counterparts. In terms of the CRPS, the estimative solutions perform similarly to the calibrated ones.

4 Conclusions

This work proposes a comparison between estimative and calibrated predictive distributions based on a bootstrap resampling procedure. The comparison is carried out on a simulation study, where appropriate verification measures, such as the CRPS, logarithmic score and coverage probabilities, are used for assessing the predictive performance of the considered distributions. From the results one can conclude that the calibrated predictive distributions always improve on the estimative ones, in terms of logarithmic score and coverage. Instead, the considered predictive distributions perform similarly with respect to the CRPS. Future development of the work will explore sliding training periods of constant size in the same way as in the work of [4]. Moreover other verification measures will also be considered.

Acknowledgments
This work was partially supported by the Italian Ministry for University and Research under the PRIN2015 Grant No. 2015EASZFS_003.

References

1. Barndorff-Nielsen, O.E., Cox, D.R.: Prediction and asymptotics. Bernoulli, **2**, 319–340 (1996).
2. Buizza, R., Houtekamer, P.L, Toth, Z., Pellerin, G., Wei, M., Zhu, Y.: A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems. Monthly Weather Review, **133**, 1076–1097 (2005).
3. Fonseca, G., Giummolè, F., Vidoni, P.: Calibrating predictive distributions. Journal of Statistical Computation and Simulation, **84**, 373–383 (2014).
4. Gneiting, T., Raftery, A.E., Westveld III, A.H., Goldman , T.: Calibrated ProbabilisticForecasting Using Ensemble Model Output Statistics and Minimum CPRS Estimation. Monthly Weather Review, **133(5)**, 1098–1118 (2005).
5. Gneiting, T., Raftery, A.E.: Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, **102**, 359–378 (2007).
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. Hamill, T. M., Colucci, S. J.: Verification of Eta-RSM short-range ensemble forecasts. Monthly Weather Review, **125**, 1312–1327 (1997).

$m = 5$				
α	Est log	Cal log	Est crps	Cal crps
0.05	0.125	0.052	0.128	0.050
0.167	0.249	0.181	0.274	0.180
0.833	0.730	0.803	0.722	0.809
0.95	0.865	0.946	0.854	0.935
$m = 10$				
α	Est log	Cal log	Est crps	Cal crps
0.05	0.162	0.050	0.184	0.046
0.167	0.290	0.191	0.304	0.187
0.833	0.719	0.822	0.719	0.823
0.95	0.845	0.939	0.836	0.941
$m = 15$				
α	Est log	Cal log	Est crps	Cal crps
0.05	0.146	0.044	0.151	0.035
0.167	0.265	0.152	0.265	0.134
0.833	0.766	0.864	0.757	0.867
0.95	0.864	0.959	0.852	0.967

Table 1 Coverage probabilities of the four predictive distributions for different nominal levels α. Standard errors are always smaller than 0.015. Est log denotes the estimative distribution with maximum likelihood estimates and Est crps the estimative distribution with CRPS estimates, while Cal log and Cal crps are the respective calibrated counterparts.

$m = 5$			
Est log	Cal log	Est crps	Cal crps
1.672	1.56	1.716	1.582
(0.043)	(0.022)	(0.047)	(0.023)
$m = 10$			
Est log	Cal log	Est crps	Cal crps
2.277	1.825	2.539	1.87
(0.085)	(0.0295)	(0.114)	(0.0368)
$m = 15$			
Est log	Cal log	Est crps	Cal crps
2.493	1.656	3.052	1.687
(0.116)	(0.024)	(0.158)	(0.022)

$m = 5$			
Est log	Cal log	Est crps	Cal crps
0.636	0.636	0.644	0.645
(0.016)	(0.014)	(0.016)	(0.015)
$m = 10$			
Est log	Cal log	Est crps	Cal crps
0.792	0.806	0.809	0.817
(0.020)	(0.0179)	(0.021)	(0.18)
$m = 15$			
Est log	Cal log	Est crps	Cal crps
0.695	0.691	0.716	0.706
(0.018)	(0.015)	(0.019)	(0.015)

Table 2 Logarithmic (left) and CRPS (right) values of the four predictive distributions for different values of m. Est log denotes the estimative distribution with maximum likelihood estimates and Est crps the estimative distribution with CRPS estimates, while Cal log and Cal crps are the respective calibrated counterparts.

8. Palmer, T.N.: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. Quarterly Journal of the Royal Meteorological Society, **128**, 747–774 (2002).
9. Schlosser, L., Hothorn, T., Stauffer, R., Zeileis, A.: Distributional regression forests for probabilistic precipitation forecasting in complex terrain. The Annals of Applied Statistics, **13(3)**, 1564–1589 (2019).