# Chapter 45
# Reliable Probabilities Through Statistical Post-processing of Ensemble Forecasts

**Bert Van Schaeybroeck and Stéphane Vannitsem**

**Abstract** We develop post-processing approaches based on linear regression that make ensemble forecasts more reliable. First of all we enforce climatological reliability (CR) in the sense that the total variability of the forecast is equal the variability of the observations. Second, we impose ensemble reliability (ER) such that the spread around the ensemble mean of the observation coincides with the one of the ensemble members. Since, generally, different ensembles have different sizes, standard post-processing methods tend to overcorrect ensembles with large spreads. By taking variable values of the error variances, our forecast becomes more reliable at short lead times as reflected by a flatter rank histogram. We illustrate our findings using the Lorenz 1963 model.

## 45.1 Introduction

The atmosphere (and its climate) is a complicated system involving multiple components, each with their own time and spatial scales, in constant interaction with one another. This system displays the property of sensitivity to initial conditions drastically limiting its predictability horizon [3]. This property is also shared by many deterministic detailed atmospheric models, reflecting the chaotic nature of their dynamics.

Operational numerical weather and climate models suffer from the presence of both initial-condition and modeling errors. In particular, even though during last decades the amount of satellite observations has increased tremendously, a substantial lack of observations above the oceans remains allowing room for improvement on the quality of initial conditions. In addition, it is nowadays also realized that model errors (errors due to unresolved scales, ill-tuned model parameters and neglected interactions between different components of the climate system) play a major role in the deterioration of forecasts as a function of lead time [6].

The recognition of these intrinsic uncertainties on the initial conditions and model physics forms the starting point of current-day weather forecast practice,

B. Van Schaeybroeck (✉) · S. Vannitsem
Royal Meteorological Institute, Ringlaan 3, 1180 Brussels, Belgium
e-mail: bertvs@meteo.be

T. Gilbert et al. (eds.), *Proceedings of the European Conference on Complex Systems 2012*, Springer Proceedings in Complexity, DOI 10.1007/978-3-319-00395-5_45,
© Springer International Publishing Switzerland 2013

through the development of an ensemble system in which a set of trajectories are integrated in time starting from different initial conditions, and/or using different model physics. The aim of using ensembles is to produce a probabilistic instead of a deterministic forecast. The spread of the ensemble around its mean can be viewed as an uncertainty measure of the current forecast and provides the possibility to evaluate the different potential scenarios that could be followed by the atmosphere. This uncertainty information is of crucial importance as the reliability of the forecast may be strongly flow dependent. However, despite huge efforts at constructing ensembles which produce a good estimate of uncertainty, all experiments show that ensemble forecasts are consistently under-dispersive or overconfident for longer lead times [5]. This also means that outliers occur more frequently than expected. The reasons of this ensemble feature include the immense complication of the natural system in comparison with the finite dimensional model (inducing the presence of model errors) and the finite number of ensemble members. It requires corrections either by modifying the model itself or by performing post-processing of the forecasts.

It is common to correct forecasts by the simplest post-processing method, the bias correction. More generally post-processing (also called Model Output Statistics, MOS) is an approach used for correcting certain aspects of the new forecasts, based on statistical features of prior comparisons between model outputs and observations. The most widespread approach applies ordinary least-squares regression (LMOS) to relate observations and model predictions, thereby assuming the sole presence of errors in the observations [1]. As the bias correction method, LMOS corrects the mean but it tends to degrade the variability, especially at long lead times. This depletion of the ensemble variance is related to the progressive decorrelation of the forecast, associated with the chaotic nature of its dynamics. As a consequence, when applying LMOS on each member of an ensemble forecast, all members will be mapped to a constant value for long lead times.

Different authors proposed an approach based on linear regression that introduces ensemble and climatological reliability [2, 4]. We outline a general framework based on maximum likelihood estimation and Lagrange multipliers for imposing an arbitrary number of constraints and generalize their set-up for use with arbitrarily many predictors. Moreover we discuss how to use the ensemble spread for estimating the error variances in order to avoid overcorrection of ensembles with large spreads.

## 45.2 Notation

Consider the meteorological variable $X$ for which we have $N$ observations $(X_{O,1}, \ldots, X_{O,N})$. Corresponding to each observation $n$, the m-th member of our ensemble forecast produces values $(V_{1n}^m, \ldots, V_{Pn}^m)$ for the $P$ different meteorological variables or predictors. The first predictor $V_1$ is the one corresponding to the variable $X$. We also define the ensemble mean values as $(\bar{V}_{1n}, \ldots, \bar{V}_{pn})$.

Given a training set of observations and corresponding predictors, we want to improve the bare uncorrected forecast $V_1$ by using all model predictors. More specifically for each member $m$ of an ensemble $n$ we construct a corrected forecast or predictand:

$$X_{C,n}^m = \alpha + \sum_p \beta_p \bar{V}_{pn} + \gamma \epsilon_n^m, \tag{45.1}$$

where we defined the deviation from the ensemble mean $\epsilon_n^m = V_{1n}^m - \bar{V}_{1n}$. This relation implies that a bias correction is realized using $\alpha$, the ensemble mean variability is corrected by adjusting the $\beta$'s and the ensemble spread by adjusting $\gamma$.

## 45.3 Constrained Maximum Likelihood Estimation

Basic climatological constraints for forecasts are the equality of mean and variance of the corrected forecast with the mean and variance of the observations. Note that even for perfect model output the associated variance may differ from the variance of the observations due to the presence of measurement errors and representativity errors. The constraint for *climatological reliability* (CR) is:

$$\sigma_C^2 = \sigma_O^2,$$

where $\sigma_C^2$ is the variance of the corrected forecast and $\sigma_O^2$ the variance of the observations.

A reliable forecast is characterized by the fact that the observation may be considered as a member of the ensemble forecast and hence has the same statistical properties. Defining the error variance of the ensemble $n$ as $\sigma_{\epsilon,n}^2$, this implies that the quadratic error divided by $\sigma_{\epsilon,n}^2$ is equal to one, once averaged over all observations and forecasts. The condition for *ensemble reliability* (ER) is therefore:

$$1 = \frac{1}{N} \sum_{n=1}^{N} \frac{(\bar{X}_{C,n} - X_{O,n})^2}{\sigma_{\epsilon,n}^2}. \tag{45.2}$$

A mathematical tool for introducing the ER and CR constraints is by introducing a constrained likelihood. Assuming the errors on the ensemble mean of the corrected forecast are normally distributed with mean zero and variance $\sigma_{\epsilon,n}^2$, the constrained log-likelihood becomes:

$$\begin{aligned}\ln \mathcal{L}(\alpha, \boldsymbol{\beta}, \gamma, \mu, \lambda) = &-\frac{1}{2} \sum_{n=1}^{N} \ln\left(2\pi \sigma_{\epsilon,n}^2\right) - \frac{1}{2} \sum_{n=1}^{N} \frac{(\bar{X}_{C,n} - X_{O,n})^2}{\sigma_{\epsilon,n}^2} \\ &+ \lambda\left(\sigma_C^2 - \sigma_O^2\right) + \mu\left(1 - \frac{1}{N} \sum_{n=1}^{N} \frac{(\bar{X}_{C,n} - X_{O,n})^2}{\sigma_{\epsilon,n}^2}\right).\end{aligned}$$

Upon maximization of this functional with respect to the different parameters we obtain the maximum likelihood estimate (MLE). The ER forecast is obtained by setting $\lambda = 0$, the CR forecast by setting $\mu = 0$, the unconstrained MLE forecast by assuming $\lambda = \mu = 0$ while the ER+CR assumes both to be nonzero.

### 45.3.1 Discussion

The error variance $\sigma^2_{\epsilon,n}$ of an ensemble $n$ is assumed to be given, but has to be specified. We adopt two approaches: a variable and a constant error variance. We start by assuming a constant error variance in the sense that it is independent on $n$ and the obvious choice is the average ensemble variance. Equation (45.2) then simply states that the mean square error of the ensemble mean equals the average ensemble variance [5]. We find that the CR forecast is equivalent to Errors-in-Variables MOS or EVMOS [7–9] which appropriately takes into account the presence of errors in both the forecasts and the observations.[1] Enforcing only ER and using a constant error variance we find that the CR constraint is automatically satisfied as shown before by different authors for the case of one predictor [2, 4].

For a skillful ensemble forecast, the magnitude of the error, which is the squared difference between ensemble mean and the observation, may be estimated by the ensemble variance. Therefore, for an inhomogeneous chaotic system there is a dependence of the error variance $\sigma^2_{\epsilon,n}$ on the ensemble $n$ itself. Applying in that case the unconstrained MLE, we find that the ER constraint is automatically satisfied.

## 45.4 Verification

We test the usefulness of the calibration methods in the context of the well-known Lorenz 63 model by focussing on ensemble scores at short lead times. The system describes thermal convection and involves three coupled first-order differential equations in time for the variables $x$, $y$ and $z$:

$$\begin{aligned}\dot{x} &= \sigma(-x+y),\\ \dot{y} &= rx - y - xz,\\ \dot{z} &= xy - bz.\end{aligned}$$

We adopt the conventional parameter choice $(\sigma, r, b) = (10, 28, 8/3)$ such that the system exhibits chaotic behavior. For generating the observations we assume a slightly biased parameter $r' = r + 0.001$ from the one used for generating the model data. The system is perfectly reliable at time zero in the sense that the observation is randomly sampled from the forecast ensemble. Our training and verification sets include 50.000 ensembles of each 51 members.

### 45.4.1 Ensemble Forecast Skill

Figure 45.1(a) shows the average continuous ranked probability skill score (CRPSS) of the different forecasts for the $z$-variable against lead time. The CRPS score is the

[1] In fact, it is equivalent to EVMOS applied on the ensemble mean.

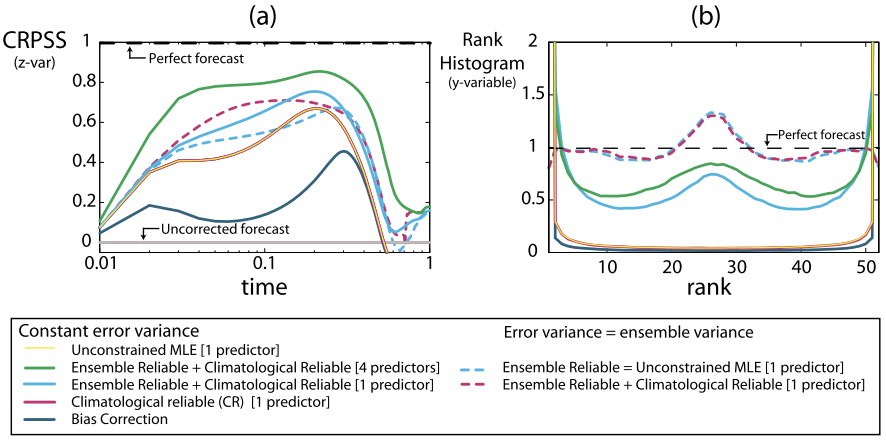


**Fig. 45.1** (**a**) CRPS skill scores for different corrected forecasts applied on the z-variable as a function of the lead time. (**b**) Rank histogram as a function of the 52 ranks for different corrected forecasts applied on the y-variable. Both plots are generated using the Lorenz 63 model with a model error and averaged over 50.000 ensemble forecasts with 51 ensemble members each

integrated squared difference between the cumulative distribution functions of the ensemble forecast and the observation [10]. The CRPS skill score of a forecast is then equal to $\mathrm{CRPSS} = 1 - \mathrm{CRPS}/\mathrm{CRPS}_u$ where $\mathrm{CRPS}_u$ is the value associated with the uncorrected forecast. We distinguish between corrections produced assuming a constant error variance (full lines) and a variable one (dashed lines). For a forecast worse (better) than the uncorrected forecast, the CRPSS is negative (positive) while a perfect forecast has a CRPSS value equal to one. The bias correction (black line) gives rise to the smallest correction while the unconstrained MLE (yellow line) and the CR forecast (red line) improve the forecast but degrade its quality after some time. The ER+CR forecast (blue full line) and all the approaches using a variable error variance (dashed lines) are consistently better. By far the best approach is the ER+CR approach (green full line) with constant error variances but using four predictors as opposed to one.

### *45.4.2 Calibration*

Figure 45.1(b) evaluates the calibration as expressed by the rank or Talagrand diagram for the different forecasts for the $y$-variable integrated over lead times between zero and one.

The rank histogram shows the frequency with which the observation lies at a specific place in the set of ranked ensemble forecasts. Ideally the rank histogram is a straight line at the value one such that the observation can be viewed as a member of the ensemble. Departure from uniformity may, among others, indicate over-dispersion or under-dispersion of the ensemble forecast. It is clearly seen that the

bias-corrected forecast, the CR forecast and the unconstrained MLE forecast (all with a constant error variances) are badly calibrated. The ER+CR forecast with constant error variances performs better but still includes many outliers. These outliers can be attributed to ensembles with small spreads. The best calibrated forecasts are the ones using a variable error variance since these approaches correct small ensembles as well as large ensembles.

## 45.5 Conclusions

In summary, we introduced a post-processing technique based on linear regression enforcing the forecast to become climatologically reliable (CR) and/or to have reliable ensembles (ER). We have tested our methods using a low-order chaotic system with model error. For short lead times our constrained approaches are superior to the uncorrected forecast, the bias-corrected forecast and the forecast obtained by an unconstrained maximum likelihood estimation. Moreover the use of the ensemble spread for estimating the error variance of the ensemble mean leads to a better calibration and avoids the undercorrection of ensembles with small spreads. Note also that, as opposed to most statistical post-processing techniques, our new approaches are computationally cheap.

## References

1. Glahn HR, Lowry DA (1972) The use of model output statistics (MOS) in objective weather forecasting. J Appl Meteorol 11:1203
2. Johnson C, Bowler N (2009) On the reliability and calibration of ensemble forecasts. Mon Weather Rev 137:1717
3. Kalnay E (2002) Atmospheric modeling, data assimilation and predictability. Cambridge University Press, Cambridge
4. Kharin VV, Zwiers FW (2003) Improved seasonal probability forecasts. J Climate 16:1684
5. Leutbecher M, Palmer TN (2008) Ensemble forecasting. J Comput Phys 227:3515
6. Nicolis C, Perdigao RAP, Vannitsem S (2009) Dynamics of prediction errors under the combined effect of initial condition and model errors. J Atmos Sci 66:766
7. Vannitsem S (2009) A unified linear model output statistics scheme for both deterministic and ensemble forecasts. Q J R Meteorol Soc 135:1801
8. Van Schaeybroeck B, Vannitsem S (2011) Post-processing through linear regression. Nonlinear Process Geophys 18:147
9. Van Schaeybroeck B, Vannitsem S (2012) Toward post-processing ensemble forecasts based on hindcasts, Scientific et Technical Publications of Royal Meteorological Institute Belgium, 61
10. Wilks DS (1995) Statistical methods in the atmospheric sciences. Academic Press, San Diego