4.3 BAYESIAN PROCESSOR OF OUTPUT FOR PROBABILISTIC FORECASTING OF PRECIPITATION OCCURRENCE

Coire J. Maranzano and Roman Krzysztofowicz*
University of Virginia, Charlottesville, Virginia

1 INTRODUCTION

The *Bayesian Processor of Output* (BPO) is a theoretically-based technique for probabilistic forecasting of weather variates (Krzysztofowicz, 2004). It processes output from a numerical weather prediction (NWP) model and optimally fuses it with climatic data in order to quantify uncertainty about a predictand. The first version of the BPO is for binary predictands. It is being tested by producing probability of precipitation (PoP) occurrence forecasts for a set of climatically diverse stations in the contiguous U.S. For each station, the PoPs are produced for the 6-h, 12-h, and 24-h periods up to 84-h ahead. The overall setup for the test is parallel to (but smaller in scope than) the operational setup of the AVN-MOS system. This system was developed by the Meteorological Development Laboratory (MDL) of the National Weather Service by applying the Model Output Statistics (MOS) technique to output fields from the Global Spectral Model run under the code name AVN. The BPO retrieves samples for the estimation of its forecasting equations from the same archive that was utilized in the development of the AVN-MOS; and the performance of the AVN-MOS system is the primary benchmark for evaluation of the performance of the BPO.

This paper presents a tutorial introduction to the BPO for PoP. Modeling and estimation are explained using a single predictor (an estimate of the total precipitation amount during a specified period). Numerical results are presented for one predictand (6-h period, 42–48h after the 0000 UTC model run), one station (Buffalo, New York), and one season (cool season). The performance of this tutorial BPO (which uses one predictor) is evaluated in terms of informativeness and calibration and then compared with the performance of the operational MOS (which uses four predictors). The results highlight the superior efficiency of the BPO in extracting the predictive information from the output of a NWP model. Implications of this fact on further development of the BPO and on its potential advantages are discussed.

*Corresponding author address: Roman Krzysztofowicz, University of Virginia, P.O. Box 400747, Charlottesville, VA 22904-4747; e-mail: rk@virginia.edu.

2 BPO TECHNIQUE

2.1 Variates

Let V be the predictand — a binary variate serving as the indicator of precipitation occurrence at a specified point and during a specified period of the day, with $V = 1$ if and only if precipitation occurs, and $V = 0$ otherwise; its realization is denoted v, where $v \in \{0, 1\}$.

Let X be the predictor — a variate whose realization x is used to forecast V. Here X denotes an estimate of the total precipitation amount during a specified period, output from the NWP model with a specified lead time. Typically, X is a binary-continuous variate: it takes on value zero on some days, and positive values on other days. Thus the sample space of X is the interval $[0, \infty)$, and the probability distribution of X should assign a nonzero probability to the event $X = 0$ and spread the complementary probability over the interval $(0, \infty)$ according to an appropriate density function.

2.2 Inputs

With P denoting a probability and p denoting a generic density function, the inputs into the BPO are defined as follows.

$g = P(V = 1)$ is the prior probability of event $V = 1$; it is to be estimated from a *climatic sample* $\{v\}$ of realizations of the predictand V.

$r_v = P(X = 0|V = v)$ for $v = 0, 1$; it is the probability of the predictor X taking on value zero, conditional on the hypothesis that the event is $V = v$.

$h_v(x) = p(x|X > 0, V = v)$ for $v = 0, 1$; h_v is the density function of the predictor X, conditional on the hypothesis that X takes on a positive value, $X > 0$, and that the event is $V = v$.

The four elements (r_0, r_1, h_0, h_1) — two constants, r_0 and r_1, and two univariate density functions, h_0 and h_1 — comprise the family of the likelihood functions of V. These four elements are to be estimated from a *joint sample* $\{(x, v)\}$ of realizations of the predictor X and the predictand V.

2.3 Forecast

The probabilistic forecast is specified in terms of the posterior probability $\pi = P(V = 1|X = x)$ of precipitation occurrence, $V = 1$, conditional on a realization of the predictor $X = x$ output from the NWP model. The forecasting equation is

$$\pi = \left[1 + \frac{1-g}{g} L(x)\right]^{-1}, \tag{1}$$

where $(1-g)/g$ is the *prior odds* against event $V = 1$, and $L(x)$ is the *likelihood ratio* against event $V = 1$ given by

$$L(x) = \begin{cases} \dfrac{r_0}{r_1} & \text{if } x = 0, \\[2ex] \dfrac{1-r_0}{1-r_1} \dfrac{h_0(x)}{h_1(x)} & \text{if } x > 0. \end{cases} \tag{2}$$

Equations (1)–(2) define the BPO for a binary predictand, using a single binary-continuous predictor.

3 EXAMPLE OF BPO

3.1 Predictand and Predictor

The event to be forecasted is the occurrence of precipitation (accumulation of at least 0.254 mm of water) in Buffalo, New York, during the 6-h period 1800-2400 UTC, beginning 42h after the 0000 UTC model run. The predictor is the estimate of the total precipitation amount during that period output from the AVN model. Forecasts are to be made every day in the cold season (October – March).

3.2 Samples

The joint sample $\{(x, v)\}$ comes from the database that the MDL used to estimate the operational forecasting equations of the AVN-MOS system. It is a 4-year long sample extending from 1 April 1997 through 31 March 2001. Although an additional climatic sample $\{v\}$ of the predictand is available, it is deliberately not used herein. For the objective of the following example is to contrast the ways in which the BPO and the MOS extract information from the same data record.

Table 1: Sample sizes and estimates of the prior probability g of precipitation occurrence; Buffalo, NY.

	Month					
	Oct	Nov	Dec	Jan	Feb	Mar
Size	123	117	123	123	111	121
g	0.098	0.231	0.268	0.285	0.234	0.207

3.3 Input Probabilities

The prior probability g is estimated for each month, from October through March. Table 1 reports the sample sizes and the estimates of g. The family of the likelihood functions is estimated for the cool season (October – March). Table 2 reports the stratification of the sample, the sizes of the subsamples, and the estimates of the probabilities r_0 and r_1.

3.4 Input Distributions

The conditional density functions, h_0 and h_1, have the corresponding conditional distribution functions, H_0 and H_1, defined by

$$H_v(x) = P(X \le x|X > 0, V = v) \quad \text{for } v = 0, 1. \tag{3}$$

These distribution functions are modeled parametrically; the parameters are estimated using an appropriate subsample (370 realizations for H_0 and 158 realizations for H_1). At Buffalo, H_0 is log-logistic (α_0, β_0), and H_1 is log-Weibull (α_1, β_1), where α_v is the scale parameter and β_v is the shape parameter. Figure 1 shows the goodness of fit and the estimates of parameters. Figure 2 shows the corresponding conditional density functions, h_0 and h_1.

Table 2: Sample sizes for the estimation of the family of likelihood functions, and estimates of the probabilities r_0 and r_1; cool season; Buffalo, NY.

	Sample Size			Probability
	$X = 0$	$X > 0$	Sum	r_v
$V = 0$	190	370	560	0.339
$V = 1$	0	158	158	0.000
Sum	190	528	718	

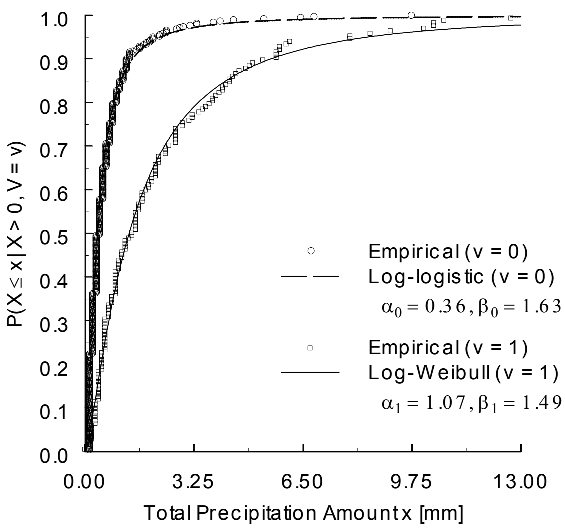

Figure 1: Empirical distribution functions and fitted parametric distribution functions H_v of the total precipitation amount X output from the AVN model, conditional on $X > 0$ and on precipitation event $V = v$ (precipitation nonoccurrence, $v = 0$; and precipitation occurrence, $v = 1$); 6-h forecast period 1800-2400 UTC, beginning 42h after the 0000 UTC model run; cool season; Buffalo, NY.

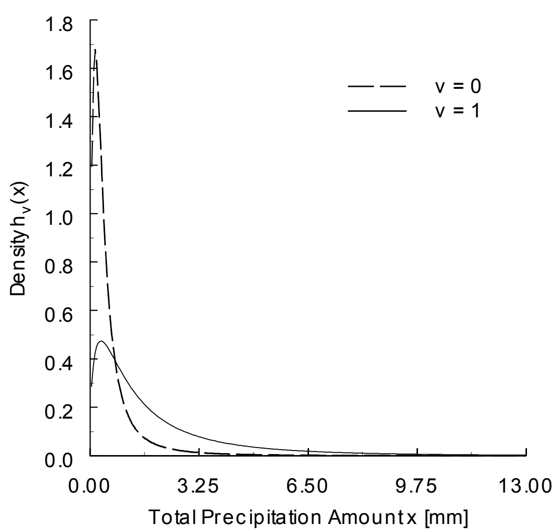

Figure 2: Conditional density functions h_v $(v = 0, 1)$ corresponding to the fitted parametric distribution functions H_v shown in Figure 1.

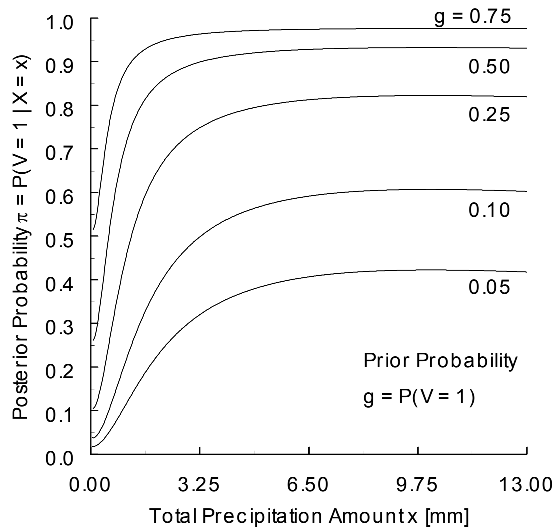

Figure 3: Posterior probability π of precipitation occurrence as a function of the total precipitation amount x output from the AVN model, for five values of the prior probability g; 6-h forecast period 1800-2400 UTC, beginning 42h after the 0000 UTC model run; cool season; Buffalo, NY.

In summary, the family of the likelihood functions for the cool season is specified in terms of six parameters $(r_0, r_1; \alpha_0, \beta_0; \alpha_1, \beta_1)$. These likelihood parameters encapsulate the informativeness of the predictor X for forecasting the predictand V. An intuitive judgment can be based on the following general relationship: the informativeness of the predictor increases with (i) the difference between the conditional probabilities r_0 and r_1 (see Table 2), and (ii) the degree of separation between the conditional density functions h_0 and h_1(see Figure 2).

3.5 Posterior Probability

Given the family of the likelihood functions (r_0, r_1, h_0, h_1) and a fixed prior probability g, the posterior probability π can be plotted as a function of x according to (1)–(2). Figure 3 shows the plots for five values of g.

When $x = 0$, the posterior probability is $\pi = 0$, regardless of the prior probability g, because $r_0 > 0$ and $r_1 = 0$. In other words, every time the AVN model output indicates that the total precipitation amount during the period will be zero, it provides a perfect forecast of precipitation nonoccurrence because such an output is never observed when the precipitation does occur.

When $x > 0$, the posterior probability π is an increasing, non-linear function of x. The shape of this function depends also on g whose value can be easily changed. This has an important practical implication: even though the family of the likelihood functions remains fixed for a season (here the cool season), the forecasting equation changes from month to month as the prior probability changes (see Table 1). Consequently, the posterior probability π is calibrated against the climatic probability for each month rather than for the entire 6-month season.

4 MOS TECHNIQUE

4.1 Forecasting Equation

The benchmark for evaluation of the BPO is the currently used MOS technique (Glahn and Lowry, 1972; Antolik, 2000). For a binary predictand, the MOS forecasting equation has the general form

$$\pi = a_0 + \sum_{i=1}^{I} a_i t_i(x_i), \tag{4}$$

where t_i is some transform determined experientially for each predictor X_i $(i = 1, ..., I)$, and $a_0, a_1, ..., a_I$ are regression coefficients. For the predictand defined in Section 3.1, the MOS utilizes four predictors:

1. Total precipitation amount during 6-h period, 42–48h [mm].

2. Total precipitation amount during 3-h period, 45–48h [mm].

3. Relative humidity at the pressure level of 700 mb at 48h [%].

4. Mean relative humidity of a variable depth layer at 42h [%].

4.2 Grid-Binary Transform

With each predictor being defined at the station (here Buffalo, NY), the transforms can be described as follows. The fourth predictor enters (4) untransformed, i.e., $t_4(x_4) = x_4$. Each of the remaining predictors is subjected to a *grid-binary transformation*, which is specified in terms of an algorithm (Jensenius, 1992). The algorithm takes the gridded field of predictor values and performs on it three operations: (i) mapping of each gridpoint value into "1" or "0", which indicates the exceedance or nonexceedance of a specified cutoff level; (ii) smoothing of the resultant binary field; and (iii) interpolation of the gridpoint values to the value $t_i(x_i)$ at a station. It follows that the transformed predictor value $t_i(x_i)$ at a station depends upon the original predictor values at all grid points in a vicinity. Thus when viewed as a transform of the original predictor X_i into a grid-binary predictor $t_i(X_i)$ at a fixed station, the transform t_i is nonlinear and nonstationary (from one forecast time to the next). The grid-binary predictor $t_i(X_i)$ is dimensionless and its sample space is the closed unit interval [0,1].

4.3 Estimation

The regression coefficients in (4) are estimated from a joint sample $\{(t_1(x_1), ..., t_I(x_I); v)\}$ of realizations of the transformed predictors and the predictand. Like the sample for the BPO, this sample includes all daily realizations in the cool season (October – March) in 4 years. Unlike the sample for the BPO, this sample includes not only the realizations at the Buffalo station, but the realizations at all stations within the region to which Buffalo belongs. The pooling of station samples into a regional sample is needed to ensure a "stable" estimation of the MOS regression coefficients (Antolik, 2000). The estimates obtained by the MDL are:

$$\begin{aligned} a_0 = -0.11234, \quad & a_1 = 0.37495, \quad & a_2 = 0.28693, \\ & a_3 = 0.10625, \quad & a_4 = 0.0029437. \end{aligned}$$

These estimates are assumed to be valid for every station within the region.

5 COMPARISON

5.1 Performance Measures

It is apparent that each system, the BPO and the MOS, processes information in a totally different manner. The objective of the following experiment is to compare the two systems with respect to the efficiency of extracting the predictive information from the same data record — the archive of the AVN model output. Towards this end, each system is used to calculate the forecast probability π based on every one of the 718 realizations of its predictor (BPO) or predictors (MOS) in the data record. Then the joint sample $\{(\pi, v)\}$ of realizations of the forecast probability and the predictand is used to calculate the following performance measures.

The *receiver operating characteristic* (ROC) — a graph of the probability of detection versus the probability of false alarm.

The *calibration function* (CF) — a graph of the conditional probability $\eta(\pi) = P(V = 1|\Pi = \pi)$ versus the forecast probability π.

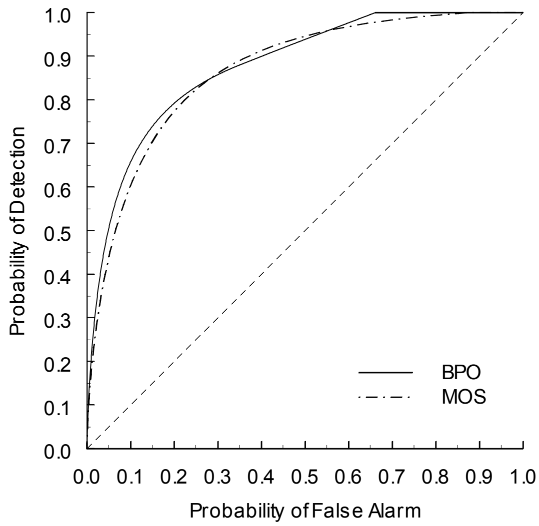

Figure 4: Receiver Operating Characteristics (ROC) of the probability of precipitation occurrence forecasts produced by (i) the BPO using a single predictor output from the AVN model, and (ii) the MOS using four predictors calculated from the output fields of the AVN model.

The *uncertainty score* (US) — the expected reduction of variance of the predictand (the difference between the prior variance and the conditional variance) as a fraction of the prior (climatic) variance:

$$\text{US} = 1 - \frac{E([1-\eta(\Pi)]\eta(\Pi))}{[1-E(V)]E(V)}; \qquad \text{US} \leq 1.$$

The *calibration score* (CS) — the Euclidean distance (the square root of the expected quadratic difference) between the line of perfect calibration and the calibration function:

$$\text{CS} = \left\{ E([\Pi - \eta(\Pi)]^2) \right\}^{\frac{1}{2}}; \qquad 0 \leq \text{CS} \leq 1.$$

Some basic facts pertaining to these performance measures are as follows:

1. System A is more informative than system B (in the sense of Blackwell (1951)) if and only if the ROC of A is superior to the ROC of B.

2. If system A is more informative than system B, then the US of A is not smaller than the US of B (DeGroot and Fienberg, 1983).

3. The quadratic score (also known as the Brier score) is equal to $(1-\text{US})[1-E(V)]E(V)+\text{CS}^2$.

5.2 Comparative Evaluation

The ROCs are shown in Figure 4. Each ROC reaches the ordinate of 1 to the left of the point (1,1)

Table 3: Performance scores of the probability of precipitation occurrence forecasts produced by (i) the BPO using a single predictor output from the AVN model, and (ii) the MOS using four predictors calculated from the output fields of the AVN model.

	System	
	BPO	MOS
Uncertainty Score	0.3506	0.3502
Calibration Score	0.049	0.060

because each system offers the decision maker an upper bound on the probability of false alarm. The bound offered by BPO is $1 - r_0 = 0.661$. The bound offered by MOS is 0.875. The ROC of BPO lies above the ROC of MOS over a larger part of the interval, but the two ROCs cross each other. Thus neither system is more informative than the other. The uncertainty scores, US, reported in Table 3, are nearly identical.

The calibration scores, CS, reported in Table 3, indicate that the BPO system is calibrated somewhat better than the MOS system, by 0.011 on average (on the probability scale).

5.3 Explanations

The total precipitation amount during the 6-h period is the sole predictor utilized by the BPO. The MOS utilizes the same predictor, which it processes through the grid-binary transform. Why is it that MOS needs three additional predictors to barely match the performance of BPO? The explanation is twofold.

First, the laws of probability theory, from which the BPO is derived, ensure the optimal structure of the BPO forecasting equation (1)–(2). The structure of the MOS forecasting equation (4) is different. Thus given any single predictor, the BPO system, if properly operationalized, can never be less informative than the MOS system (or any other non-Bayesian system for that matter). To make up for the non-optimal theoretic structure, a non-Bayesian system needs additional predictors (which are conditionally informative in that system).

Second, the grid-binary transform (Jensenius, 1992) was invented to improve the calibration of the MOS system. But by mapping the original predictor (which is binary-continuous or continuous) into a binary predictor, this transform also removes part of predictive information contained in the original predictor. In the example reported herein, three additional predictors are needed to make up for the lost information and the nonoptimal structure of the MOS forecasting equation.

To dissect the predictive performance of the grid-binary transform, each the MOS and the BPO was estimated and evaluated twice: first, utilizing the original first predictor, and next utilizing the grid-binary transformation of that predictor. There are two findings. (i) The use of the grid-binary transform in the MOS leads to a compromise: the transform improves the CS but deteriorates the US. (ii) The use of the grid-binary transform in the BPO is unnecessary for calibration (because the BPO automatically calibrates the posterior probability against the specified prior probability) and is detrimental for informativeness (because it removes part of the predictive information contained in the original predictor).

6 CLOSURE

6.1 Preliminary Conclusions

1. The BPO utilizing a single predictor performs, in terms of both informativeness and calibration, at least as well as the MOS utilizing four predictors. This shows that BPO is **more efficient** than MOS in extracting predictive information from the output of a NWP model.

2. The single predictor in the BPO is a direct model output (interpolated to the station), whereas three out of four predictors in the MOS are grid-binary predictors whose definitions require subjective experimentation (to set the cutoff levels and smoothing parameters) and algorithmic processing of the entire output fields (to calculate the predictor values). Thus in terms of the definitions of the predictors, the BPO is **more parsimonious** than the MOS.

6.2 Potential Implications

1. There exists a potential for increasing the performance of the BPO by including other predictors that are informative, conditionally on the first predictor already utilized. For instance, almost all applications of MOS and other regression-type techniques in the U.S., Canada, and Europe utilize relative humidity at a fixed pressure level as one of the predictors. This and other predictors will be tried in the BPO as well.

2. Inasmuch as the grid-binary predictors can be dispensed with because only the basic and derived predictors need be considered by the BPO, the set of candidate predictors for the BPO is about 50% smaller than the set of candidate predictors for the MOS. For instance, in one case we counted 67 candidate predictors were offered to the MOS screening regression process — the number that could be reduced to 28 for the BPO. In general, the overall effort needed to select the most informative subset of predictors can be reduced.

3. With fewer number of predictors (say between one and four for BPO, instead of between four and fifteen for MOS), an extension of the BPO to processing an ensemble of the NWP model output will present a less demanding task (in terms of data storage requirements and computing requirements) than it would be if an extension of the MOS technique were attempted.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant No. ATM-0135940, "New Statistical Techniques for Probabilistic Weather Forecasting". The Meteorological Development Laboratory of the National Weather Service provided the AVN-MOS database. The advice of Mark S. Antolik on accessing and interpreting the data is gratefully acknowledged.

References

Antolik, M.S., 2000: An overview of the National Weather Service's centralized statistical quantitative precipitation forecasts. *Journal of Hydrology,* **239**(1–4), 306–337.

Blackwell, D., 1951: Comparison of experiments. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J. Neyman (ed.), University of California Press, Berkeley, pp. 93–102.

DeGroot, M.H., and S.E. Fienberg, 1983: The comparison and evaluation of forecasters. *The Statistician*, **32**, 12–22.

Glahn, H.R., and D.A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, **11**(8), 1203–1211.

Jensenius, J.S.,Jr., 1992: The use of grid-binary variables as predictors for statistical weather forecasting. Preprints of the 12th Conference on Probability and Statistics in the Atmospheric Sciences, Toronto, Ontario, American Meteorological Society, pp. 225–230.

Krzysztofowicz, R., 2004: Bayesian processor of output: A new technique for probabilistic weather forecasting. Preprints of the 17th Conference on Probability and Statistics in the Atmospheric Sciences, Seattle, Washington, American Meteorological Society; paper number 4.2.