BAYESIAN PROCESSOR OF OUTPUT:
PROBABILISTIC QUANTITATIVE PRECIPITATION FORECAST

By

Roman Krzysztofowicz
University of Virginia
Charlottesville, Virginia

and

Coi re J. Maranzano
Johns Hopkins University
Baltimore, Maryland

Research Paper RK–0602
http://www.faculty.virginia.edu/rk/

March 2006

Copyright © 2006 by R. Krzysztofowicz and C.J. Maranzano

Corresponding author address: Professor Roman Krzysztofowicz, University of Virginia, P.O. Box 400747, Charlottesville, VA 22904–4747. E-mail: rk@virginia.edu
ABSTRACT

The Bayesian Processor of Output (BPO) is a theoretically-based technique for probabilistic forecasting of weather variates. The second version of the BPO described herein is for continuous predictands; it is tested by producing the conditional probabilistic quantitative precipitation forecast (conditional PQPF). This is the forecast of the precipitation amount to be accumulated during a specified period at a specified location, conditional on precipitation occurrence. The forecast is in the form of a (continuous) posterior distribution function, which is obtained through Bayesian fusion of a prior (climatic) distribution function and a realization of predictors output from a numerical weather prediction (NWP) model. The strength of the BPO derives from (i) the theoretic structure of the forecasting equation (which is Bayes theorem), (ii) the flexibility of the meta-Gaussian family of likelihood functions (which allows any form of the marginal distribution functions, and a non-linear and heteroscedastic dependence structure), (iii) the simplicity of estimation, and (iv) the effective use of asymmetric samples (typically, a long climatic sample of the predictand and a short operational sample of the NWP model output).

Modeling and estimation of the BPO are explained in a setup parallel to that of the Model Output Statistics (MOS) technique used operationally by the National Weather Service. The BPO is compared with the MOS in terms of the structure of the forecasting equations and examples of forecasts. The results highlight the advantages of the BPO in terms of (i) completeness of characterization of uncertainty for a decision maker, (ii) efficiency of extracting predictive information from the NWP model output (fewer predictors needed), and (iii) parsimony of the predictors (no need for experimentation to find suitable transformations of the NWP model output). Potential implications for operational forecasting and ensemble processing are discussed.
# TABLE OF CONTENTS

ABSTRACT .................................................................................................................. ii

1. INTRODUCTION ..................................................................................................... 1
   1.1 Towards Bayesian Forecasting Techniques ......................................................... 1
   1.2 BPO for Continuous Predictands .................................................................... 2

2. BAYESIAN THEORY ................................................................................................. 3
   2.1 Variates ........................................................................................................... 3
   2.2 Samples .......................................................................................................... 3
   2.3 Input Elements .............................................................................................. 4
   2.4 Theoretic Structure ....................................................................................... 5

3. META-GAUSSIAN MODEL ..................................................................................... 6
   3.1 Input Elements .............................................................................................. 6
   3.2 Forecasting Equations .................................................................................. 9
   3.3 Basic Properties .......................................................................................... 10
   3.4 Numerical Aspects ...................................................................................... 11

4. STRUCTURE OF PQPF .......................................................................................... 12

5. EXAMPLE WITH ONE PREDICTOR ..................................................................... 13
   5.1 Forecasting Equations ................................................................................. 13
   5.2 Prior Distribution Function ..................................................................... 13
   5.3 Marginal Distribution Function .................................................................. 14
   5.4 Likelihood Dependence Structure .............................................................. 15
   5.5 Informativeness of Predictor ....................................................................... 17
   5.6 Posterior Distribution Functions .................................................................. 19
   5.7 Second Predictor ......................................................................................... 20
   5.8 Third Predictor ............................................................................................. 21
   5.9 Comparison of Predictors .......................................................................... 21

6. EXAMPLE WITH THREE PREDICTORS ................................................................ 22
   6.1 Dependence Parameters ............................................................................. 22
   6.2 Predictors Selection ...................................................................................... 23
7. MOS TECHNIQUE .................................................................................. 24
   7.1 Forecasting Equations ................................................................. 24
   7.2 Grid-Binary Transform ................................................................. 25
   7.3 Estimation ................................................................................. 26
   7.4 Predictors Selection ................................................................. 26

8. COMPARISON OF BPO WITH MOS ............................................... 27
   8.1 Forecast Format ......................................................................... 27
   8.2 Information Content ................................................................. 28
   8.3 Comparative Verifications .......................................................... 29

9. CLOSURE ......................................................................................... 31
   9.1 Bayesian Technique .................................................................... 31
   9.2 Structural Advantages ............................................................... 32
   9.3 Potential Implications ............................................................... 32

ACKNOWLEDGMENTS ........................................................................... 34
REFERENCES .......................................................................................... 35
TABLES ................................................................................................... 37
FIGURES ................................................................................................. 39
1. INTRODUCTION

1.1 Towards Bayesian Forecasting Techniques

Rational decision making by industries, agencies, and the public in anticipation of heavy precipitation, snow storm, flood, or other disruptive weather phenomenon, requires information about the degree of confidence that the user can place in a weather forecast. It is vital, therefore, to advance the meteorologist’s capability of quantifying forecast uncertainty to meet the society’s rising expectations for reliable information.

Our objective is to develop and test a coherent set of theoretically-based techniques for probabilistic forecasting of weather variates. The basic technique, called Bayesian Processor of Output (BPO), 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 extended technique, called Bayesian Processor of Ensemble (BPE), processes an ensemble of the NWP model output (or multiple model outputs). The concept is depicted in Figure 1.

The theoretic structures of the BPO (and the BPE) are derived from the laws of probability theory. In particular, the principles of Bayesian forecasting and fusion are followed (Krzysztofowicz, 1983, 1999; Krzysztofowicz and Long, 1990). There are three structures, for

- binary predictands (e.g., indicator of precipitation occurrence),
- multi-category predictands (e.g., indicator of precipitation type),
- continuous predictands (e.g., precipitation amount conditional on precipitation occurrence, temperature, visibility, ceiling height, wind speed).

As is well known, Bayes theorem provides the optimal theoretical framework for fusing information from different sources and for obtaining the probability distribution of a predictand, conditional on a realization of predictors, or conditional on an ensemble of realizations. The
challenge is to develop and test techniques suitable for operational forecasting.

1.2 BPO for Continuous Predictands

The present article describes the BPO for continuous predictands. This BPO is tested by producing the conditional probabilistic quantitative precipitation forecast (conditional PQPF). The forecast is in the form of a (continuous) distribution function of the precipitation amount accumulated during a specified period at a specified location, conditional on precipitation occurrence. The (unconditional) PQPF may be readily constructed from the distribution of amount (DoA) and the probability of precipitation (PoP) occurrence. The PoP is produced by the BPO for binary predictands (Krzysztofowicz and Maranzano, 2006).

The overall setup for the test is parallel to (but smaller in scope than) the operational setup for the Model Output Statistics (MOS) technique (Glahn and Lowry, 1972) used in operational forecasting by the National Weather Service (NWS). In the currently deployed AVN-MOS system (Antolik, 2000), the predictors for the MOS forecasting equations are based on output fields from the Global Spectral Model run under the code name AVN. The performance of the AVN-MOS system is the primary benchmark for evaluation of the performance of the BPO.

The article is organized as follows. Section 2 presents the gist of the Bayesian theory of forecasting. Section 3 details the input elements, the forecasting equations, and the basic properties of the BPO. Section 4 defines the structure of PQPF. Section 5 presents an example of the BPO for DoA using a single predictor. Section 6 presents an example of the BPO for DoA using three predictors. The BPO is compared and contrasted with the operational MOS in terms of the structure of the forecasting equations in Section 7, and in terms of the forecast formats and contents in Section 8. Section 9 summarizes implications of these comparisons and potential advantages of the BPO.
2. BAYESIAN THEORY

2.1 Variates

Let $W$ be the predictand — a continuous variate whose realization $w$ is to be forecasted. Let $X_i$ be the predictor — a variate whose realization $x_i$ is used to forecast $W$. Let $X = (X_1, ..., X_I)$ be the vector of $I$ predictors; its realization is denoted $x = (x_1, ..., x_I)$. Each $X_i$ ($i = 1, ..., I$) is assumed to be a continuous variate.

2.2 Samples

Suppose the forecasting problem has already been structured, and the task is to develop the forecasting equation in a setup similar to that of the MOS technique (Antolik, 2000). In the examples throughout the article, the predictand $W$ is the precipitation amount, conditional on precipitation occurrence (accumulation of at least 0.254 mm of water) in Quillayute, Washington, during the 24-h period 1200–1200 UTC, beginning 12 h after the run of the AVN model at 0000 UTC; thus $W > 0$. The predictors are the variates whose realizations are output from the AVN model. Forecasts are to be made every day in the cool season (October–March).

Let $\{w\}$ denote the climatic sample of the predictand, with each $w > 0$. The climatic sample comes from the database of the National Climatic Data Center (NCDC). This database contains hourly precipitation observations in Quillayute from over 37 years; however, the record is heterogeneous and must be processed in order to obtain a homogeneous sample. To avoid this task, only observations recorded by the Automated Surface Observing System (ASOS) are included in the prior sample. In effect, it is a 7-year long sample extending from 1 January 1997 through 31 December 2003. Each day provides one realization. The sample size for the cool season is $M = 818$.

Let $\{(x, w)\}$ denote the joint sample of the predictor vector and the predictand, with each $w > 0$. 

0. The joint sample comes from the database that the Meteorological Development Laboratory (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. The sample size for the cool season is \( N = 470 \).

The point of the above example is that typically the joint sample is much shorter than the climatic sample: \( N << M \). Classical statistical methods, such as the MOS technique, deal with this sample asymmetry by simply ignoring the long climatic sample. In effect, these methods throw away vast amount of information about the predictand. In contrast, the BPO will use effectively both samples; it will extract information from each sample and then optimally fuse information according to the laws of probability. (Pooling of samples from different months and stations in order to increase the sample size is a separate issue.)

### 2.3 Input Elements

With \( P \) denoting the probability and \( p \) denoting a generic density function, define the following objects.

\[
g(w) = p(w)\] is the prior density function of the predictand \( W \); it is to be estimated from the climatic sample \( \{w\} \). The density function \( g \) quantifies the uncertainty about the predictand \( W \) that exists before the NWP model output is available. Equivalently, it characterizes the natural variability of the predictand.

\[
f(x|w) = p(x|W = w)\] function \( f(·|w) \) is the \( I \)-variate density function of the predictor vector \( X \), conditional on the hypothesis that the actual realization of the predictand is \( W = w \). The family \( \{f(·|w): \text{all } w\} \) of the density functions of \( X \) is to be estimated from the joint sample \( \{(x, w)\} \). For a fixed realization \( X = x \), object \( f(x|·) \) is the likelihood function of \( W \). Symbol \( f \) will denote the family of likelihood functions. This family quantifies the stochastic dependence
between the predictor vector $X$ and the predictand $W$. Equivalently, it characterizes the informativeness of the predictors with respect to the predictand.

2.4 Theoretic Structure

The density function $g$ and the family of likelihood functions $f$ carry information about the prior uncertainty and the informativeness of the predictors into the Bayesian revision procedure. The expected density function $\kappa$ of the predictor vector $X$ is given by the total probability law:

$$\kappa(x) = \int_{-\infty}^{\infty} f(x|w)g(w) \, dw,$$

and the posterior density function $\phi(w) = p(w|X = x)$ of predictand $W$, conditional on a realization of the predictor vector $X = x$, is given by Bayes theorem:

$$\phi(w) = \frac{f(x|w)g(w)}{\kappa(x)}.$$

The corresponding posterior distribution function $\Phi(w) = P(W \leq w|X = x)$ is given by the equation

$$\Phi(w) = \frac{1}{\kappa(x)} \int_{-\infty}^{w} f(x|u)g(u) \, du.$$

The inverse function $\Phi^{-1}$ is called the posterior quantile function. For any number $p$ such that $0 < p < 1$, the $p$-probability posterior quantile of predictand $W$ is the quantity $w_p$ such that $\Phi(w_p) = p$. Therefore

$$w_p = \Phi^{-1}(p).$$

Equations (1)–(4) define the theoretic structure of the BPO for a continuous predictand.
3. META-GAUSSIAN MODEL

To operationalize the BPO, a flexible and convenient model is needed for the family of likelihood functions \( f \). We employ the meta-Gaussian model developed by Kelly and Krzysztofowicz (1994, 1995, 1997) and applied successfully to probabilistic river stage forecasting (Krzysztofowicz and Herr, 2001; Krzysztofowicz, 2002). The meta-Gaussian \( f \) is constructed from specified marginal distributions, a correlation matrix, and the Gaussian dependence structure. When this \( f \) is inserted into Equations (1)–(3), and all derivations are completed, the meta-Gaussian BPO is obtained. It is described below in terms of input elements, forecasting equations, and basic properties.

3.1 Input Elements

The following algorithm defines the input elements, outlines the estimation procedure, and details the calculation of the posterior parameters (the parameters of the forecasting equations).

**Step 0.** Given are two samples, the climatic sample of the predictand \( W \), and the joint sample of the predictor vector and the predictand \( (X, W) \), respectively:

\[
\{ w(n) : n = 1, ..., M \}, \\
\{(x(n), w(n)) : n = 1, ..., N \},
\]

where \( x(n) = (x_1(n), ..., x_I(n)) \) and \( N \leq M \); all realizations of \( W \) from the joint sample are included in the climatic sample.

**Step 1.** Using the climatic sample, estimate the prior (climatic) distribution function \( G \) of predictand \( W \), such that

\[
G(w) = P(W \leq w);
\]

let \( g \) denote the corresponding prior (climatic) density function of \( W \).
Step 2. Using the marginal sample \( \{x_i(n) : n = 1, \ldots, N\} \) of the joint sample, estimate the marginal distribution function \( \bar{K}_i \) of predictor \( X_i \), such that
\[
\bar{K}_i(x_i) = P(X_i \leq x_i), \quad i = 1, \ldots, I.
\]
(The bar over \( K_i \) signifies that this is only an initial distribution function of \( X_i \), which need not cohere to the specified prior distribution function of \( W \) and the yet-to-be-constructed family of likelihood functions. This detail is accounted for in the derivation of the meta-Gaussian BPO, and thus need not be considered in application.)

Step 3. Define the normal quantile transform (NQT) of the predictand and of every predictor:
\[
V = Q^{-1}(G(W)), \quad (5a)
\]
\[
Z_i = Q^{-1}(\bar{K}_i(X_i)), \quad i = 1, \ldots, I, \quad (5b)
\]
where \( Q \) is the standard normal distribution function, and \( Q^{-1} \) is the inverse of \( Q \). Next, apply the NQT to each realization in the original joint sample; specifically, for \( n = 1, \ldots, N \), calculate
\[
v(n) = Q^{-1}(G(w(n))),
\]
\[
z_i(n) = Q^{-1}(\bar{K}_i(x_i(n))], \quad i = 1, \ldots, I;
\]
then form the transformed joint sample
\[
\{(z(n), v(n)) : n = 1, \ldots, N\},
\]
where \( z(n) = (z_1(n), \ldots, z_I(n)) \).

Step 4. Using the transformed joint sample, estimate the following moments. For the transformed predictand \( V \),
\[
\mu_0 = E(V),
\]
\[
\sigma_0^2 = Var(V).
\]
For every transformed predictor $Z_i$, $i = 1, \ldots, I$,

$$
\mu_i = E(Z_i),
$$

$$
\sigma_i^2 = Var(Z_i),
$$

$$
\sigma_{i0} = Cov(Z_i, V).
$$

For $i = 1, \ldots, I - 1$ and $j = i + 1, \ldots, I$,

$$
\sigma_{ij} = Cov(Z_i, Z_j).
$$

The estimates of variances and covariances should be the maximum likelihood estimates (i.e., they should be calculated using $N$ as the divisor).

**Step 5.** Form two $I$-dimensional column vectors

$$
\mu = (\mu_1, \ldots, \mu_I),
$$

$$
\sigma = (\sigma_{10}, \ldots, \sigma_{I0}),
$$

the transpose of vector $\sigma$, which is denoted $\sigma^T$, and an $I \times I$ symmetric matrix

$$
\Sigma = \{\sigma_{ij}\},
$$

with $\sigma_{ii} = \sigma_i^2$ for $i = 1, \ldots, I$, and $\sigma_{ji} = \sigma_{ij}$ for $i = 1, \ldots, I - 1$ and $j = i + 1, \ldots, I$. Next calculate an $I \times I$ symmetric matrix

$$
M = (\Sigma - \sigma_0^{-2}\sigma\sigma^T)^{-1}.
$$

(6)

**Step 6.** Calculate the values of the posterior parameters as follows:

$$
T = \left(\frac{\sigma_i^4}{\sigma^T M \sigma + \sigma_0^2}\right)^{\frac{1}{2}},
$$

(7)

$$
c^T = \frac{T^2}{\sigma_0^2} \sigma^T M,
$$

(8)
where $c^T = [c_1, ..., c_I]$ is an $I$-dimensional row vector.

3.2 Forecasting Equations

When the family of likelihood functions $f$ is meta-Gaussian, and the input elements are defined and estimated as above, the BPO defined by Equations (1)–(4) takes the following form. Given a prior distribution function $G$ of predictand $W$ and given a realization $x = (x_1, ..., x_I)$ of the predictor vector, the meta-Gaussian posterior distribution function of predictand $W$ is specified by the equation

$$
\Phi(w) = Q \left( \frac{1}{T} \left[ Q^{-1}(G(w)) - \sum_{i=1}^{I} c_i Q^{-1}(\bar{K}_i(x_i)) - c_0 \right] \right) \cdot (10)
$$

For any number $p$ such that $0 < p < 1$, the $p$-probability posterior quantile of predictand $W$ is specified by the equation

$$
w_p = G^{-1} \left( Q \left( \sum_{i=1}^{I} c_i Q^{-1}(\bar{K}_i(x_i)) + c_0 + T Q^{-1}(p) \right) \right) \cdot (11)
$$

Given also a prior density function $g$ of predictand $W$, the meta-Gaussian posterior density function of predictand $W$ is specified by the equation

$$
\phi(w) = \frac{1}{T} \exp \left( \frac{1}{2} \left\{ \left[ Q^{-1}(G(w)) \right]^2 - \left[ Q^{-1}(\Phi(w)) \right]^2 \right\} \right) g(w) \cdot (12)
$$

Equation (10) reveals a special case of the meta-Gaussian BPO: the posterior distribution function $\Phi$ is Gaussian if every marginal distribution function $G, \, \bar{K}_1, ..., \, \bar{K}_I$ is Gaussian; this is seldom true for meteorological variates. Equation (12) conveys explicitly the essence of the Bayesian revision: the prior density function $g$, which may be of any form, is revised into a
posterior density function \( \phi \). The revision is accomplished by multiplying the prior density function and the likelihood function, whose element \( \Phi \) carries information from the given realization \( x = (x_1, ..., x_I) \) of the predictor vector.

### 3.3 Basic Properties

The meta-Gaussian BPO specified by Equations (10)–(12) provides a complete characterization of the posterior uncertainty about a continuous predictand. It offers five properties, which are important for application (and which set it apart from other processors of the NWP model output):

1. The prior (climatic) distribution function \( G \) of predictand \( W \) may take any form. Likewise, the marginal distribution function \( \bar{K}_i \) of each predictor \( X_i (i = 1, ..., I) \) may take any form. These distribution functions may be parametric or nonparametric.

2. The transforms (the NQTs) for the predictand and for each predictor are uniquely specified once the marginal distribution functions \( G, \bar{K}_1, ..., \bar{K}_I \) have been estimated.

3. The dependence structure among all the variates \( W, X_1, ..., X_I \) is pairwise; the degree of dependence is quantified by the covariances of the transformed variates.

4. The dependence structure between any two predictors \( X_i \) and \( X_j, i \neq j \), and between any predictor \( X_i \) and the predictand \( W \), may be non-linear (in the conditional mean) and heteroscedastic (in the conditional variance).

5. The probabilistic forecast of a continuous predictand \( W \) may be specified in terms of a continuous distribution function \( \Phi \), or in terms of a continuous density function \( \phi \), or in terms of any number of quantiles \( w_p \). Each of these elements is specified by an analytic expression which is simple to evaluate.

Properties 1 and 4 imply the flexibility in fitting the model to data. Properties 2 and 3 imply the simplicity of estimation. Property 5 implies the computational efficiency — an important attribute.
for operational forecasting.

3.4 Numerical Aspects

To recapitulate, the BPO using $I$ continuous predictors of a continuous predictand is specified completely by: $I + 1$ marginal distribution functions

$$G, \bar{K}_i \text{ for } i = 1, ..., I;$$

and $I + 2$ dependence parameters

$$T, c_0, c_i \text{ for } i = 1, ..., I.$$

Steps 1 and 2 of the estimation procedure outlined in Section 3.1 can be automated by creating a catalog of parametric models and by developing algorithms for estimation of the parameters and choice of the best model. The catalog should include expressions for the distribution functions and for the density functions.

The numerical calculations of forecasts are speeded up when each marginal distribution function is parametric and has a closed-form expression, and when the prior distribution function $G$ has also closed-form expressions for its inverse $G^{-1}$ and for its density function $g$. Functions $Q$ and $Q^{-1}$ can be readily evaluated via polynomial approximations (Abramowitz and Stegun, 1972). In effect, the execution of the BPO can be simple and fast.
4. STRUCTURE OF PQPF

Let $W$ denote the precipitation amount accumulated during a specified period at a specified location; naturally $W \geq 0$. Let $X$ denote a vector of predictors. Because $W$ is a binary-continuous predictand, forecasting $W$ probabilistically poses an extra challenge. We handle it by decomposing the forecasting task as follows. The BPO for binary predictands, described by Krzysztofowicz and Maranzano (2006), produces the probability of precipitation (PoP) occurrence:

$$\pi = P(W > 0|X = x).$$

The BPO for continuous predictands, described herein, produces the distribution of amount (DoA), conditional on occurrence:

$$\Phi(w) = P(W \leq w|X = x, W > 0), \quad w \geq 0,$$

where $\Phi(w) > 0$ if $w > 0$, and $\Phi(w) = 0$ if $w = 0$. Then the probabilistic quantitative precipitation forecast (PQPF) is constructed from PoP and DoA. Specifically, the posterior distribution function of $W$ is given by

$$P(W \leq w|X = x) = (1 - \pi) + \pi\Phi(w), \quad w \geq 0,$$

and the posterior density function of $W$ is given by

$$p(w|X = x) = (1 - \pi)\delta(w) + \pi\phi(w), \quad w \geq 0,$$

where $\delta$ is the Dirac function. Figure 2 shows an example of this construction. The next section details examples of the DoA forecasts.
5. EXAMPLE WITH ONE PREDICTOR

5.1 Forecasting Equations

When there is only one predictor \((I = 1)\), its subscript is omitted. Thus \(X\) replaces \(X_1\), \(\bar{K}\) replaces \(\bar{K}_1\), and the forecasting equations (10)–(11) can be written

\[
\Phi(w) = Q \left( \frac{1}{T} \left[ Q^{-1} (G(w)) - c_1 Q^{-1} (\bar{K}(x)) - c_0 \right] \right),
\]

\[
w_p = G^{-1} \left( Q \left( c_1 Q^{-1} (\bar{K}(x)) + c_0 + T Q^{-1} (p) \right) \right).
\]

Equation (12) for \(\phi\) remains intact. In effect, five elements are needed for forecasting: two univariate distribution functions, \(G\), \(\bar{K}\), and three parameters, \(T\), \(c_0\), \(c_1\).

5.2 Prior Distribution Function

The prior distribution function \(G\) of precipitation amount \(W\) is conditional on precipitation occurrence:

\[
G(w) = P(W \leq w | W > 0).
\]

It is estimated for the cool season from the climatic sample of size \(M = 818\). Figure 3 shows the empirical distribution function and the estimated parametric \(G\), which is Weibull with the scale parameter \(\alpha = 0.592\) and the shape parameter \(\beta = 0.880\). Despite the large sample, there are only 15 realizations greater than 75 mm. Thus, one of the roles of a parametric model for \(G\) is to interpolate (and to extrapolate) the right tail of the distribution function where only a few extreme amounts have been observed.

Overall, the prior distribution function has five attributes important for application. (i) It may be location-specific and season-specific and thereby may capture the “micro-climate”. (ii) It may be estimated from a large climatic sample. (iii) It interpolates (and extrapolates) the empirical
distribution function in the regions of the sample space where the observations are sparse (or nonexistent). (iv) It is independent of the choice of the predictors and the length of the NWP model output available for estimation (the size of the joint sample). (v) It need not be re-estimated when the NWP model changes; thus it ensures a stable calibration of the forecast probabilities for as long as the climate remains stationary.

5.3 Marginal Distribution Function

The single predictor $X$ is the estimate of the total precipitation amount accumulated in the 24-h period 12–36 h after the 0000 UTC model run (for short, 24-h total precipitation). The marginal distribution function $K$ of $X$ is conditional on precipitation occurrence:

$$K(x) = P(X \leq x | W > 0).$$

It is estimated for the cool season from the joint sample of size $N = 470$. Figure 4 shows the empirical distribution function and the estimated parametric $K$, which is Weibull with parameters $\alpha_1 = 9.603$ and $\beta_1 = 0.910$.

The comparison of $X$ with $W$ reveals two significant distinctions. First, the highest model output $x$ in the joint sample does not exceed 54 mm, whereas 29 observed amounts $w$ in the climatic sample exceed 54 mm. Second, both distribution functions, $K$ and $G$, are Weibull — which is a desirable property when $X$ is an estimator of $W$; however, while the shape parameter values are nearly identical, the scale parameter values are vastly different. Thus, the statistical distinction between $X$ and $W$, in this example, lies in the scaling. Specifically, when the shape parameter values are rounded to the nearest 0.1, so that $\beta = \beta_1 = 0.9$, each of the rescaled variates, $X/9.603$ and $W/0.592$, has the same Weibull distribution function. This distinction in scaling of $X$ and $W$ may have two causes: the inadequacy of the NWP model, and the size of the joint sample, which is about 43% smaller than the size of the climatic sample. Whatever the causes, the BPO
automatically accounts for the inequality $\tilde{K} \neq G$. This is a superior way of “debiasing” $X$, relative to $W$, than through adjustments of the moments of $X$.

5.4 Likelihood Dependence Structure

The case of a single predictor allows us to explain the meta-Gaussian dependence structure and to illustrate its properties. In accordance with Bayes theorem (2), the family of likelihood functions $f$ must capture the stochastic dependence between the predictor $X$ and the predictand $W$. The meta-Gaussian model for $f$ (and hence for $\phi$) results from the following approach. First, each original variate, $W$ and $X$, is mapped through its NQT into a normally distributed variate: $V = Q^{-1}(G(W))$ and $Z = Q^{-1}(\tilde{K}(X))$. Second, the family of the likelihood functions in the transformed space is assumed to be Gaussian; it is therefore easy to model and estimate; it is also easy to validate. Third, the results are transformed into the original space. The theoretical derivations along this approach yield the forecasting equations (12), (15) and (16). The empirical analyses pertaining to the estimation of parameters and the validation of assumptions are described below.

In the meta-Gaussian model, the stochastic dependence between $Z$ and $V$ is characterized by the linear regression

$$E(Z|V = v) = av + b, \quad (17a)$$

$$Var(Z|V = v) = \sigma^2, \quad (17b)$$

whose parameters can be expressed in terms of the estimates of the moments, which are obtained in Step 4 of Section 3.1 as follows:

$$a = \frac{\sigma_{10}}{\sigma_0^2}, \quad b = \mu_1 - \frac{\sigma_{10}}{\sigma_0^2} \mu_0, \quad (18a)$$

$$\sigma^2 = \sigma_1^2 - \frac{\sigma_{10}^2}{\sigma_0^2}. \quad (18b)$$
The posterior distribution parameters can then be expressed in terms of the linear regression parameters:

\[
c_1 = \frac{a}{a^2 + \sigma^2}, \quad c_0 = -\frac{ab}{a^2 + \sigma^2}, \quad (19a)
\]

\[
T = \left(\frac{\sigma^2}{a^2 + \sigma^2}\right)^{1/2}. \quad (19b)
\]

In the present case, \(a = 0.681, b = 0.028, \sigma = 0.829\); as a result, \(c_1 = 0.592, c_0 = -0.016, T = 0.773\). A graphical analysis supporting this estimation procedure follows.

The original joint sample \(\{(x(n), w(n)) : n = 1, \ldots, 470\}\) is plotted in Figure 5a. The scatterplot reveals that the empirical dependence structure between \(X\) and \(W\) is nonlinear and heteroscedastic: the scatter is not elliptic and increases with \(w\). The transformed joint sample \(\{(z(n), v(n)) : n = 1, \ldots, 470\}\) is plotted in Figure 5b. The scatterplot reveals that the NQT essentially eliminated the nonlinearity and the heteroscedasticity: the scatter is nearly elliptic. (The few vertical columns of points, which may be discerned on the left, are the artifact of the precision of the rain gauge measurement: the left-most column corresponds to the smallest possible measurement \(w = 0.254\) mm, the next column corresponds to \(w = 0.508\) mm, and so on.)

Figure 5b shows also the estimated linear regression of \(Z\) on \(V\) and the 80% central credible interval around it. Figure 5a shows the mapping of these three lines into the original space of \(X\) and \(W\). The mapping is performed as follows. For any \(p\) such that \(0 < p < 1\), the \(p\)-probability quantile of \(Z\), conditional on \(V = v\), is readily obtained from Equation (17):

\[
z_p = av + b + \sigma Q^{-1}(p). \quad (20)
\]

Let \(x_p\) denote the corresponding \(p\)-probability quantile of \(X\), conditional on \(W = w\). The two quantiles are related through the NQT, and so are the realizations of the conditioning variates.
Hence,
\[ x_p = \hat{K}^{-1} \left( Q(aQ^{-1}(G(w))) + b + \sigma Q^{-1}(p) \right). \] (21)
For \( p = 0.5 \), Equation (20) gives the conditional median of \( Z \), which is equal to the conditional mean of \( Z \) given by Equation (17a) and graphed as the regression line in Figure 5b. The mapping of this regression line into the original space via Equation (21) gives the median regression of \( X \) on \( W \), graphed in Figure 5a. In addition, each figure shows graphs of two conditional quantiles for \( p = 0.1 \) and \( p = 0.9 \), which define the 80% central credible interval around the conditional median. Altogether, the graphs illustrate how the NQT maps a linear and homoscedastic dependence structure into a nonlinear and heteroscedastic dependence structure.

5.5 Informativeness of Predictor

A predictor \( X \) used in the BPO is characterized in terms of its informativeness. Intuitively, the informativeness of predictor \( X \) may be visualized by judging the degree of dependence between \( X \) and \( W \). Under the meta-Gaussian model for \( f \), the stochastic dependence between \( X \) and \( W \) is characterized by the Pearson’s product-moment correlation coefficient between \( Z \) and \( V \), which is
\[ \gamma = \left( \text{sign of } a \right) \left[ \left( \frac{a}{\sigma} \right)^{-2} + 1 \right]^{-\frac{1}{2}}. \] (22)
It may be transformed into the Spearman’s rank correlation coefficient between \( X \) and \( W \), which is
\[ \rho = (6/\pi) \arcsin (\gamma/2). \] (23)
The values of \( \gamma \) and \( \rho \) are reported in Figure 5.

The Bayesian measure of informativeness of predictor \( X \) with respect to predictand \( W \) is the signal-to-noise ratio \( |a|/\sigma \). In the linear regression of \( Z \) on \( V \), as defined by Equation (17), the absolute value of the slope coefficient, \( |a| \), is the measure of signal, and the standard deviation of
the residual, $\sigma$, is the measure of noise. If $|a|/\sigma = 0$, then $X$ is uninformative for predicting $W$. If $|a|/\sigma = \infty$, then $X$ is a perfect predictor of $W$.

An equivalent measure is the informativeness score: $IS = |\gamma|$, or more explicitly,

$$ IS = \left[ \left( \frac{a}{\sigma} \right)^{-2} + 1 \right]^{-\frac{1}{2}}. $$  \hspace{1cm} (24)

The score is bounded, $0 \leq IS \leq 1$, with $IS = 0$ for an uninformative predictor, and $IS = 1$ for a perfect predictor. (The informativeness score was called the Bayesian correlation score in the original publication by Krzysztofowicz (1992).)

When there are two or more alternative predictors, they may be compared, and ranked, in terms of a binary relation of informativeness. This relation derives from the Bayesian theory of sufficient comparisons, the essence of which is as follows (Blackwell, 1951; Krzysztofowicz, 1987, 1992). Let $X_i$ and $X_j$ be two alternative predictors of $W$, having informativeness scores $IS_i$ and $IS_j$, respectively. Suppose a rational decision maker will use the probabilistic forecast of $W$ from the BPO in a Bayesian decision procedure with a prior density function $g$ and a loss function $l$. Let $VA_i(g,l)$ denote the (ex-ante) value of a probabilistic forecast generated by predictor $X_i$.

**Definition.** Predictor $X_i$ is said to be more informative than predictor $X_j$ if and only if the value of a forecast generated by predictor $X_i$ is at least as high as the value of a forecast generated by predictor $X_j$, for every prior density function $g$ and every loss function $l$; formally, if and only if

$$ VA_i(g,l) \geq VA_j(g,l), \quad \text{for every } g,l. $$

Inasmuch as any two rational decision makers may employ different prior density functions and different loss functions, the condition “for every $g,l$” is synonymous with the statement “for every rational decision maker”.

18
Theorem. If \( IS_i > IS_j \), then predictor \( X_i \) is more informative than predictor \( X_j \).

If \( IS_i = IS_j \), then the two predictors are equivalent, and one should be indifferent between selecting either \( X_i \) or \( X_j \) for the BPO. Such predictors are called equally informative. With this addition, the binary relation of informativeness establishes a weak order on a set of predictors \( X_1, ..., X_I \). The weak order is transitive and strongly complete.

In summary, an advantage of the BPO is that its meta-Gaussian model for the family of likelihood functions \( f \) enables us to directly characterize the informativeness of a predictor for a given predictand. When two or more predictors are available, they can easily be compared, and ranked, in terms of the informativeness relation.

5.6 Posterior Distribution Functions

Once the five elements \((G, \bar{K}; T, c_0, c_1)\) are specified, the probabilistic forecast of the precipitation amount \( W \), conditional on precipitation occurrence, \( W > 0 \), may be calculated, given any value \( x \) of the 24-h total precipitation output from the AVN model. Figure 6a shows three posterior distribution functions \( \Phi \) calculated from Equation (15), given three different values of the predictor \( X \). Each of these posterior distribution functions results from a revision of the prior (climatic) distribution function \( G \), which is shown also. Figure 6b shows the posterior density functions \( \phi \) calculated from Equation (12), and the prior (climatic) density function \( g \).

Three observations are noteworthy. First, the prior density function \( g \) has an inverted \( J \) shape, but each of the posterior density functions \( \phi \) has a mound shape. Second, as the predictor value \( x \) increases, the posterior density function \( \phi \) becomes less sharp — an implication of the heteroscedastic dependence structure between \( X \) and \( W \). Third, even though the highest value of the predictor in the joint sample is 54 mm, the posterior density function can be produced given any value of the predictor, and the shape of \( \phi \) evolves continuously as \( x \) increases.
5.7 Second Predictor

Different predictors behave differently. That is why each predictor should be modeled individually, and the catalog of parametric models from which the marginal distribution functions are drawn should be large enough to afford flexibility. To underscore this point, let us model another predictor: the relative vorticity on the isobaric surface of 850 mb at 24 h after the 0000 UTC model run (for short, 850 relative vorticity at 24 h). Figure 7 shows the empirical distribution function and the estimated parametric $\bar{K}$, which is log-logistic with the scale parameter $\alpha_2 = 6.212$, the shape parameter $\beta_2 = 4.863$, and the shift parameter $\eta_2 = -5$; thus the sample space of this predictor is $(-5, \infty)$.

Figure 8 illustrates the modeling and the estimation of the likelihood dependence structure. In the space of the original variates (Figure 8a), the scatterplot reveals a nonlinear and heteroscedastic dependence structure between $X$ and $W$. The median regression of $X$ on $W$ is nonlinear and heteroscedastic; and the 80% central credible interval around the median is skewed. The NQT is invaluable here: it works automatically and it works well. In the space of the transformed variates (Figure 8b), the regression of $Z$ on $V$ is plausibly linear and homoscedastic: the scatter of points is nearly elliptic (except for the artifact of the measurement, as explained in Section 5.4).

Figure 9 shows three posterior distribution functions $\Phi$ and the corresponding density functions $\phi$, as well as the prior (climatic) distribution function $G$ and the corresponding density function $g$. 

20
5.8 Third Predictor

The third predictor to be considered herein is the vertical velocity on the isobaric surface of 700 mb at 12 h after the 0000 UTC model run (for short, 700 vertical velocity at 12 h). Its empirical distribution function has a long left tail and a short right tail. For this reason, the sample space of \( X \) is an interval bounded above \((-\infty, 0.4)\), and the parametric distribution function \( \bar{K} \) is obtained by estimating a distribution function of \(-X\) on the sample space \((-0.4, \infty)\); it is log-logistic with the scale parameter \( \alpha_3 = 0.539 \), the shape parameter \( \beta_3 = 4.313 \), and the shift parameter \( \eta_3 = -0.4 \).

5.9 Comparison of Predictors

Table 1 compares the marginal distribution functions; Table 2 compares the likelihood parameter values, the posterior parameter values, and the informativeness scores of each predictor. The comparison reveals that for forecasting the precipitation amount during the 24-h period 12–36 h, the 24-h total precipitation is more informative than the 700 vertical velocity at 12 h, which in turn is more informative than the 850 relative vorticity at 24 h. Would a combination of any two predictors, or a combination of all three predictors, be more informative than the most informative predictor alone? This question is answered at the end of Section 6.
6. EXAMPLE WITH THREE PREDICTORS

6.1 Dependence Parameters

Let \( X_1 \) denote the 24-h total precipitation (the predictor analysed in Section 5.3), let \( X_2 \) denote the 850 relative vorticity at 24 h (the predictor analysed in Section 5.7), and let \( X_3 \) denote the 700 vertical velocity at 12 h (the predictor analysed in Section 5.8). The analyses of individual predictors supply the marginal distribution functions \( \bar{K}_1, \bar{K}_2, \bar{K}_3 \), which are listed in Table 1. In order to obtain the BPO with three predictors \((X_1, X_2, X_3)\), the moments defined in Step 4 of Section 3.1 are estimated:

\[
\begin{align*}
\mu_0 &= -0.097, \\
\sigma_0^2 &= 0.980, \\
\mu &= (-0.038, -0.029, -0.021), \\
\sigma &= (0.667, 0.431, -0.497).
\end{align*}
\]

\[
\Sigma = \begin{bmatrix}
1.140 & 0.153 & -0.219 \\
0.153 & 1.029 & -0.400 \\
-0.219 & -0.400 & 1.069
\end{bmatrix}
\]

Then the elements defined in Step 6 of Section 3.1 are calculated:

\[
\begin{align*}
c_1 &= 0.505, \\
c_2 &= 0.241, \\
c_3 &= -0.275, \\
c_0 &= -0.025, \\
T &= 0.641.
\end{align*}
\]

Thereby all input elements have been obtained. After they are inserted into expressions (10), (11), and (12), the BPO is ready for forecasting. Examples of forecasts are presented in Section 8.
6.2 Predictors Selection

For every predictand, 34 potential predictors are defined by appropriately concatenating five variables (total precipitation amount, mean relative humidity, relative vorticity, relative humidity, and vertical velocity), three lead times, and four isobaric surfaces. From this set, the best combination of no more than five predictors is selected. The selection is accomplished via an algorithm that (i) maximizes $IS$ (an informativeness score which is an extension of the score defined in Section 5.5) subject to the constraint that an additional predictor must increase $IS$ by at least a specified threshold, (ii) employs objective optimization and heuristic search, and (iii) estimates the parameters of the BPO and the informativeness score $IS$ from a given joint sample (an estimation sample — here from 4 years).

In the examples for Quillayute with one, two, and three predictors, the scores are as follows:

\[
IS(X_1) = 0.63, \quad IS(X_2) = 0.43, \quad IS(X_3) = 0.48,
\]
\[
IS(X_1, X_2) = 0.73, \quad IS(X_1, X_3) = 0.73,
\]
\[
IS(X_1, X_2, X_3) = 0.77.
\]

Thus the combinations of two predictors, $(X_1, X_2)$ and $(X_1, X_3)$, are equally informative and each is more informative than the most informative single predictor $X_1$. The combination of three predictors $(X_1, X_2, X_3)$ is the most informative of all six combinations considered herein.
7. MOS TECHNIQUE

7.1 Forecasting Equations

The benchmark for evaluation of the BPO is the currently used MOS technique (Glahn and Lowry, 1972; Antolik, 2000). For a continuous predictand, the MOS technique requires that the sample space be discretized. For instance, to forecast precipitation amount \( W \) in a 24-h period, conditional on precipitation occurrence \( W > 0 \), five cutoff levels are prescribed \( \{w_1, w_2, w_3, w_4, w_5\} = \{2.54, 6.35, 12.7, 25.4, 50.8\} \) mm. Then the MOS forecasting equation is developed for each cutoff level \( w_j \) \( (j = 1, ..., 5) \) to produce the conditional exceedance probability in the general form

\[
1 - \Phi(w_j) = a_{0j} + \sum_{i=1}^{I} a_{ij} t_i(x_i), \quad j = 1, ..., 5, \tag{25}
\]

where \( t_i \) is some transform determined experientially for each predictor \( X_i \) \( (i = 1, ..., I) \), and \( a_{0j}, a_{1j}, ..., a_{Ij} \) are regression coefficients estimated from the joint sample

\[
\{(x, v) : v = 1 \Leftrightarrow w \geq w_j, v = 0 \Leftrightarrow w < w_j\},
\]

which is constructed, as indicated, from the joint sample \( \{(x, w)\} \). The predictand and the predictors are defined at a station. For the predictand defined in Section 2.2, the MOS utilizes 15 predictors, listed below in the order in which they were selected into the forecasting equation:

1. Total precipitation amount during 12-h period, 12–24 h; cutoff 6.35 mm.
2. Total precipitation amount during 12-h period, 12–24 h; cutoff 25.4 mm.
3. Total precipitation amount during 12-h period, 24–36 h; cutoff 0.254 mm.
4. Total precipitation amount during 24-h period, 12–36 h.
5. Total precipitation amount during 12-h period, 12–24 h; cutoff 2.54 mm.
6. Total precipitation amount during 24-h period, 12–36 h; cutoff 12.7 mm.
7. Relative vorticity at the pressure level of 850 mb at 12 h.

8. Longitude of the station.

9. Total precipitation amount during 12-h period, 12–24 h; cutoff 12.7 mm.

10. Elevation of the station.

11. Latitude of the station.

12. Convective precipitation amount during 24-h period, 12–36 h; cutoff 0.254 mm.

13. Relative vorticity at the pressure level of 850 mb at 24 h.

14. Vertical velocity at the pressure level of 500 mb at 24 h; cutoff $-0.9$.

15. Vertical velocity at the pressure level of 500 mb at 12 h; cutoff $-0.5$.

### 7.2 Grid-Binary Transform

Some predictors (e.g., predictors 4, 7, 8, 10, 11, 13 above) enter Equation (25) untransformed, i.e., $t_i(x_i) = x_i$. Each predictor with a cutoff level is subjected to a grid-binary transformation, which is specified in terms of a heuristic 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. If 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 $[0,1]$. 
7.3 Estimation

The regression coefficients in Equation (25) 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 Quillayute station, but the realizations at all stations within the region to which Quillayute 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 are assumed to be valid for every station within the region.

7.4 Predictors Selection

For every predictand, there are about 176 potential predictors. The main reason for this number being about five times larger than 34 in BPO is that MOS employs the grid-binary predictors: for each variable there are several cutoff levels, each of which generates a new predictor. The best predictors are selected sequentially according to the maximum variance reduction criterion of linear regression and the stopping criterion whereby an additional predictor must reduce variance by at least a specified threshold. Up to 15 predictors can be selected.
8. COMPARISON OF BPO WITH MOS

8.1 Forecast Format

It is apparent that each system, BPO and MOS, processes information in a totally different manner and produces forecast in a totally different format (continuous versus discrete). To highlight the differences, we compare forecasts that would be made on 21 February 2002. The BPO described in Section 6 has 15 parameters and uses 3 predictors whose values are

\[ x_1 = 30.2, \quad x_2 = 4.8, \quad x_3 = -0.95. \]

The MOS described in Section 7 has 80 parameters and uses 15 predictors. Both systems share two predictors: 24-h total precipitation is \( x_1 \) in BPO and predictor 4 in MOS; 850 relative vorticity at 24 h is \( x_2 \) in BPO and predictor 13 in MOS. In addition, a vertical velocity is \( x_3 \) in BPO, and two vertical velocities enter MOS as grid-binary predictors 14 and 15. On the whole, the BPO uses \( 3/15 \) (or 20%) of the predictors used by the MOS.

The precipitation amount actually observed was 101.85 mm. The empirical prior (climatic) distribution function shown in Figure 10 reveals that this was an extreme amount: it was exceeded by only 5 out of 818 observations in the climatic sample, which implies the probability of nonexceedance equal to 0.994. The NWP model estimate was 30.2 mm — a gross underestimate, equal to 30% of the amount actually observed.

The MOS forecast, shown in Figure 10, consists of five points \( \{ (w_j, \Phi(w_j)) : j = 1, \ldots, 5 \} \), where the abscissa \( w_j \) is a cutoff level of precipitation amount, and the ordinate \( \Phi(w_j) \) is the conditional probability of event \( W \leq w_j \). The largest cutoff level is 50.8 mm, which is 50% of the amount actually observed, and the highest nonexceedance probability is 0.73. Because three probabilities are zero, only three probabilities (attached to the cutoff levels 12.7, 25.4, 50.8 mm) delineate the distribution function of \( W \). An interpolation (linear as in Figure 10, or otherwise)
between the points may be reasonable, but no objective basis exists for extrapolation of the distribution function beyond the largest cutoff level of 50.8 mm and the highest probability of 0.73. In short, the right tail of the distribution function is missing. Hence, the MOS forecast does not provide information about the risk of an extreme event — the kind of information that is essential for probabilistic forecasting of floods and anticipatory response (Krzysztofowicz, 2002).

The BPO forecast, shown in Figure 10, provides a continuous distribution function $\Phi$, which specifies the posterior probability $\Phi(w)$ of event $W \leq w$, for every $w \geq 0$. For instance, the event $W \leq 30.2$ mm (the amount estimated by the NWP model) has the posterior probability $\Phi(30.2) = 0.15$; the event $W \leq 101.85$ mm (the amount actually observed) has the posterior probability $\Phi(101.85) = 0.93$. In addition, the BPO forecast provides, at no extra cost, a continuous density function $\phi$, which is shown in Figure 11, and which is needed by sophisticated users who employ mathematical models to make optimal decisions under uncertainty.

8.2 Information Content

Inasmuch as 101.85 mm of rainfall in 24 h is an extreme and rare observation, it is advisable to interpret the probabilistic forecast in terms of odds on the exceedance event $W \geq 101.85$ mm (Murphy, 1991).

The prior (climatic) probability of this event is 0.006, which gives the prior odds on the event

$$\frac{1 - G(101.85)}{G(101.85)} = \frac{0.006}{0.994} = \frac{1}{165.7}.$$  

The prior probability of 0.006 is revised by the BPO, based on the current values of the three predictors, into the posterior probability of 0.07; this gives the posterior odds on the event

$$\frac{1 - \Phi(101.85)}{\Phi(101.85)} = \frac{0.07}{0.93} = \frac{1}{13.3}.$$
Thus, the ratio of the posterior odds to the prior odds is 12.5. In other words, the BPO forecast increased the odds on the extreme event, which actually did occur, by the factor of 12.5. This is a substantial increase of the risk on that particular day, relative to the average (climatic) risk. In conclusion, the right tail of the distribution function $\Phi$ provides important information to a rational decision maker concerned with the possibility of an extreme event. The MOS forecast, in its discrete 5-point format, does not provide any information about the risk of an extreme event.

To further demonstrate the limitation of the discrete forecast format provided by MOS, and the advantage of the continuous forecast format provided by BPO, a hypothetical situation is considered. Suppose that all predictors retain the values from 21 February 2002, except for the 24-h total precipitation output from the NWP model which is now $x_1 = 60$ mm. Figure 12 shows the resultant forecasts. Four out of five MOS probabilities are now zero; thus only two probabilities (attached to the cutoff levels 25.4, 50.8 mm) delineate the distribution function of $W$. The highest probability specified is 0.57. In short, the entire right half of the distribution function is missing. On the other hand, the BPO provides a continuous distribution function $\Phi$ on the entire sample space, as before. From it, one may read $\Phi(60) = 0.3$, and $\Phi(101.85) = 0.76$. Calculations analogous to the previous ones for the event $W \geq 101.85$ mm show that the ratio of the posterior odds to the prior odds is now 52.3. In other words, the BPO forecast informs the user that the odds on this extreme event are 52.3 times larger today than they are on average. The MOS forecast for the same event is a blank.

8.3 Comparative Verifications

Complete results of comparative verifications of the BPO forecasts and the MOS forecasts will be reported in another paper. The general conclusion that emerges from these results is twofold. (i) The PQPF produced by the BPO is, on average, better calibrated and more informative than
the PQPF produced by the MOS. (ii) There is a consistent and substantial difference in terms of the number of “optimal predictors” selected for each system during its development: BPO uses 1–4 predictors, which are always extracted directly from the output fields of the AVN model; MOS uses 5–15 predictors, most of which are obtained through grid-binary transformations of the output fields of the AVN model (as described in Section 7.2).
9. CLOSURE

9.1 Bayesian Technique

1. The BPO for continuous predictands described herein is the first technique of its kind for probabilistic forecasting of weather variates: it produces the posterior distribution function of a continuous predictand through Bayesian fusion of a prior (climatic) distribution function and a realization of predictors output from a NWP model; it also produces the posterior density function and the posterior quantile function. Each of these functions is analytic and easy to evaluate.

2. The BPO implements Bayes theorem, which provides the correct theoretic structure of the forecasting equation, and employs the meta-Gaussian family of multivariate density functions, which provides a flexible and convenient parametric model. It can be estimated effectively from asymmetric samples — the climatic sample of the predictand (which is typically long), and the joint sample of the predictor vector and the predictand (which is typically short).

3. The development of the BPO has focused on quality of modeling and simplicity of estimation. The BPO allows (i) the prior (climatic) distribution function of the predictand and the marginal distribution functions of the predictors to be of any form (as typically they are non-Gaussian), (ii) the dependence structure between any predictor and the predictand to be non-linear and heteroscedastic (as is typically the case in meteorology), and (iii) the predictors which may be dependent, conditionally on the predictand, and whose dependence structure is pairwise, non-linear, and heteroscedastic. Despite this flexibility, the implementation of the BPO requires the estimation of only distribution parameters, means, variances, and covariances. And the entire process of selecting predictors, choosing parametric distribution functions, and estimating parameters can be fully automated.
9.2 Structural Advantages

1. The continuous format of the BPO forecast provides a complete characterization of uncertainty, whereas the discrete format of the MOS forecast provides inadequate characterization of uncertainty. To wit, the continuous distribution function provided by the BPO always allows the decision maker to calculate the probability of any extreme event; by contrast, the 5-point approximation provided by the MOS degenerates as the probability of precipitation amount greater than 50.8 mm rises, leaving the right tail of the distribution function blank — just as the right tail becomes essential for assessing the risk of flood.

2. The BPO utilizing 1–4 predictors performs, in terms of both calibration and informativeness, at least as well as the MOS utilizing 4–15 predictors. This shows that BPO is more efficient than MOS in extracting predictive information from the output of a NWP model.

3. Every predictor in the BPO is a direct model output (interpolated to the station), whereas most 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 definitions of the predictors, the BPO is more parsimonious than the MOS.

9.3 Potential Implications

1. 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 potential predictors for the BPO is about 5 times smaller than the set of potential predictors for the MOS. Consequently, the overall effort needed to select the most informative subset of predictors can be reduced substantially.

2. 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 and the MOS forecasting equations for comparative verifications. The collaboration of Drs. Harry R. Glahn and Paul Dallavalle in this regard is much appreciated; the advise of Mark S. Antolik on accessing and interpreting the data is gratefully acknowledged.
REFERENCES


Table 1. Marginal distribution functions of three predictors of the precipitation amount $W$, conditional on precipitation occurrence, $W > 0$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$X_i$</th>
<th>Sample Space</th>
<th>$K_i$</th>
<th>$\alpha_i$</th>
<th>$\beta_i$</th>
<th>$\eta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Total Precip.</td>
<td>$(0, \infty)$</td>
<td>Weibull</td>
<td>9.603</td>
<td>0.910</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Rel. Vorticity</td>
<td>$(-5, \infty)$</td>
<td>Log-logistic</td>
<td>6.212</td>
<td>4.863</td>
<td>-5</td>
</tr>
<tr>
<td>3</td>
<td>Ver. Velocity</td>
<td>$(-\infty, 0.4)$</td>
<td>Log-logistic*</td>
<td>0.539</td>
<td>4.313</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

* Distribution function of $-X_3$. 
Table 2. Dependence parameters for each predictor (when used alone) of the precipitation amount $W$, conditional on precipitation occurrence, $W > 0$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.

<table>
<thead>
<tr>
<th></th>
<th>Likelihood Parameters</th>
<th>Posterior Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a$</td>
<td>$b$</td>
</tr>
<tr>
<td>1</td>
<td>0.681</td>
<td>0.028</td>
</tr>
<tr>
<td>2</td>
<td>0.440</td>
<td>0.014</td>
</tr>
<tr>
<td>3</td>
<td>−0.507</td>
<td>−0.028</td>
</tr>
</tbody>
</table>
FIGURE CAPTIONS

Figure 1. The concept of the Bayesian Processor of Output.

Figure 2. Construction of the PQPF from the probability of precipitation occurrence, $\pi$, and the distribution function of precipitation amount, conditional on occurrence, $\Phi$.

Figure 3. Prior distribution function $G$ of precipitation amount $W$, conditional on precipitation occurrence $W > 0$; 24-h forecast period 1200–1200 UTC; cool season; Quillayute, WA.

Figure 4. Marginal distribution function $\bar{K}$ of the 24-h total precipitation amount $X$ output from the AVN model, conditional on precipitation occurrence, $W > 0$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.

Figure 5. Dependence structure of the likelihood function, conditional on precipitation occurrence, $W > 0$: (a) meta-Gaussian regression of the 24-h total precipitation amount, $X$, on the actual precipitation amount, $W$, and the 80% central credible interval in the original space; (b) linear regression of $Z$ on $V$, and the 80% central credible interval in the transformed space; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.
Figure 6. Examples of probabilistic forecasts of the precipitation amount $W$, conditional on precipitation occurrence, $W > 0$, and based on three different realizations $x = 5, 35, 65$ [mm] of predictor $X$ — the 24-h total precipitation amount output from the AVN model: (a) the prior (climatic) distribution function $G$ and three posterior distribution functions $\Phi$; (b) the prior (climatic) density function $g$ and three posterior density functions $\phi$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.

Figure 7. Marginal distribution function $\bar{K}$ of the 850 relative vorticity at 24 h output from the AVN model, conditional on precipitation occurrence, $W > 0$; 0000 UTC model run; cool season; Quillayute, WA.

Figure 8. Dependence structure of the likelihood function, conditional on precipitation occurrence, $W > 0$: (a) meta-Gaussian regression of the 850 relative vorticity at 24 h, $X$, on the actual precipitation amount, $W$, and the 80% central credible interval in the original space; (b) linear regression of $Z$ on $V$, and the 80% central credible interval in the transformed space; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.

Figure 9. Examples of probabilistic forecasts of the precipitation amount $W$, conditional on precipitation occurrence, $W > 0$, and based on three different realizations $x = -3, 5, 13$ of predictor $X$ — the 850 relative vorticity at 24 h output from the AVN model: (a) the prior (climatic) distribution function $G$ and three posterior distribution functions $\Phi$; (b) the prior (climatic) density function $g$ and three posterior density functions $\phi$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.
Figure 10. Retrospective forecasts of the precipitation amount $W$, conditional on precipitation occurrence, $W > 0$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; 21 February 2002; Quillayute, WA: (i) MOS forecast specifies five probabilities of nonexceedance for fixed cutoff levels of precipitation amount \{2.54, 6.35, 12.7, 25.4, 50.8\} mm; (ii) BPO forecast specifies a continuous distribution function $\Phi$ of precipitation amount; (iii) the NWP model estimate is 30.2 mm; (iv) the actual precipitation amount is 101.85 mm; (v) the empirical prior (climatic) distribution function provides a reference.

Figure 11. The conditional density functions corresponding to the conditional distribution functions shown in Figure 10: (i) the posterior density function $\phi$ output from the BPO; (ii) the prior (climatic) density function $g$.

Figure 12. Hypothetical forecasts in the same situation as the one described in Figure 10, except that the NWP model estimate is now 60 mm.
Figure 1. The concept of the Bayesian Processor of Output.
Figure 2. Construction of the PQPF from the probability of precipitation occurrence, \( \pi \), and the distribution function of precipitation amount, conditional on occurrence, \( \Phi \).
Figure 3. Prior distribution function $G$ of precipitation amount $W$, conditional on precipitation occurrence $W > 0$; 24-h forecast period 1200–1200 UTC; cool season; Quillayute, WA.
Figure 4. Marginal distribution function $\bar{K}$ of the 24-h total precipitation amount $X$ output from the AVN model, conditional on precipitation occurrence, $W > 0$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.
Figure 5. Dependence structure of the likelihood function, conditional on precipitation occurrence, $W > 0$: (a) meta-Gaussian regression of the 24-h total precipitation amount, $X$, on the actual precipitation amount, $W$, and the 80% central credible interval in the original space; (b) linear regression of $Z$ on $V$, and the 80% central credible interval in the transformed space; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.
Figure 6. Examples of probabilistic forecasts of the precipitation amount $W$, conditional on precipitation occurrence, $W > 0$, and based on three different realizations $x = 5, 35, 65$ [mm] of predictor $X$ — the 24-h total precipitation amount output from the AVN model: (a) the prior (climatic) distribution function $G$ and three posterior distribution functions $\Phi$; (b) the prior (climatic) density function $g$ and three posterior density functions $\phi$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.
Figure 7. Marginal distribution function $\bar{K}$ of the 850 relative vorticity at 24 h output from the AVN model, conditional on precipitation occurrence, $W > 0$; 0000 UTC model run; cool season; Quillayute, WA.
Figure 8. Dependence structure of the likelihood function, conditional on precipitation occurrence, $W > 0$: (a) meta-Gaussian regression of the 850 relative vorticity at 24 h, $X$, on the actual precipitation amount, $W$, and the 80% central credible interval in the original space; (b) linear regression of $Z$ on $V$, and the 80% central credible interval in the transformed space; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.
Figure 9. Examples of probabilistic forecasts of the precipitation amount $W$, conditional on precipitation occurrence, $W > 0$, and based on three different realizations $x = -3, 5, 13$ of predictor $X$ — the 850 relative vorticity at 24 h output from the AVN model: (a) the prior (climatic) distribution function $G$ and three posterior distribution functions $\Phi$; (b) the prior (climatic) density function $g$ and three posterior density functions $\phi$; 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; cool season; Quillayute, WA.
Figure 10. Retrospective forecasts of the precipitation amount \( W \), conditional on precipitation occurrence, \( W > 0 \); 24-h forecast period 1200–1200 UTC, beginning 12 h after the 0000 UTC model run; 21 February 2002; Quillayute, WA: (i) MOS forecast specifies five probabilities of nonexceedance for fixed cutoff levels of precipitation amount \{2.54, 6.35, 12.7, 25.4, 50.8\} mm; (ii) BPO forecast specifies a continuous distribution function \( \Phi \) of precipitation amount; (iii) the NWP model estimate is 30.2 mm; (iv) the actual precipitation amount is 101.85 mm; (v) the empirical prior (climatic) distribution function provides a reference.
Figure 11. The conditional density functions corresponding to the conditional distribution functions shown in Figure 10: (i) the posterior density function $\phi$ output from the BPO; (ii) the prior (climatic) density function $g$. 
Figure 12. Hypothetical forecasts in the same situation as the one described in Figure 10, except that the NWP model estimate is now 60 mm.