Bayesian Processor of Output:
Estimation, Forecasting, and Verification
Software System

Coiro J. Maranzano and Roman Krzysztofowicz
Draft: 5 July 2006

Copyright 2006 by C. J. Maranzano and R. Krzysztofowicz
Table of Contents

1. Introduction ........................................................................................................ 3
2. Software System Purpose .................................................................................. 3
3. Design .................................................................................................................. 4
4. Computation ......................................................................................................... 6
5. Capabilities .......................................................................................................... 6
   5.1 Special Univariate Distribution Fitting Software ........................................ 7
   5.2 Binary Predictand Verification Software ................................................... 8
   5.3 Continuous Predictand Verification Software ............................................ 8
6. Future Software Systems .................................................................................. 10
7. References ......................................................................................................... 10
8. Acknowledgments .............................................................................................. 10
1. Introduction

The Bayesian Processor of Output (BPO) is a theoretically-based technique for probabilistic forecasting of weather variates. The theoretic structures of the BPO are derived from the laws of probability theory, in particular, from the principles of Bayesian forecasting and fusion. 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).

Extensive theoretical research has led to the complete development and testing of the BPO for a binary predictand (Krzysztofowicz and Maranzano, 2006a) and the BPO for a continuous predictand (Krzysztofowicz and Maranzano, 2006b). The BPO for a multi-category predictand has been developed only in part.

This document describes briefly the Bayesian Processor of Output: Estimation, Forecasting, and Verification (BPO-EFV) software system, which comprises all computational machinery developed in the course of research.

2. Software System Purpose

The purpose of the BPO-EFV software system is to perform all computations needed in the development and testing of the BPO for probabilistic weather forecasting. Within the development and testing paradigm the software has three uses. The primary use of the BPO-EFV software system is to perform an exploratory analysis of the BPO for a binary predictand, or a
categorical predictand, or a continuous predictand. The secondary use of the BPO-EFV software system is to verify (i) the forecasts produced by the BPO and (ii) the forecasts produced by some other technique — in other words, to perform comparative verification. The tertiary use of the BPO-EFV software system is to serve as a prototype of an operational probabilistic forecasting software system.

3. Design

The BPO-EFV software system is designed to have the same operational framework for each predictand type. Figure 1 shows this framework including the inputs, routines, and outputs. The software routines are ordered linearly to support an exploratory analysis. A fully operational forecasting software system would perform the parameter estimation, predictor selection, and verification off-line; only the forecasting routine would be run in real-time.

The following considerations are accounted for in the design of the software. It is designed so that the samples, parameter estimates, forecasts, and verification statistics can be easily transferred to a standard statistical software package for additional analysis and plotting. The software is also designed for reuse. Several subroutines are used in both the binary predictand and the continuous predictand configurations of the BPO-EFV software system. The software is coded in FORTRAN 90 so that subroutines can be easily ported to the National Weather Service for testing or use in an operational forecasting system. The routines are designed so that they can stand alone and be reconfigured as necessary. Finally, a great effort was made to design computationally efficient routines so that forecasts can be produced in real-time with the minimum number of computations.

Coire J. Maranzano and Roman Krzysztofowicz
Figure 1. Operational framework of the BPO-EFV software system.
4. Computation

In order for the BPO-EFV software system to fulfill its purpose, many difficult computational problems had to be overcome. Chief among these problems was the task of automating the estimation and selection of a parametric model for a distribution function of a continuous variate. In overcoming this problem (i) a catalog of parametric distribution functions was created, (ii) a new method of parameter estimation, called the uniform estimation method, was developed, and (iii) an approach to automatically selecting the best fit distribution function was developed. These developments are implemented in the BPO-EFV software system. The uniform estimation method and the approach to selecting a parametric distribution function are described in detail by Maranzano and Krzysztofowicz (2006).

Other challenging theoretical and computational problems included (i) ensuring the monotonicity of the likelihood ratio function, (ii) selecting an approximation for the standard normal quantile function that is a good compromise between computational speed and accuracy, (iii) developing a comprehensive and efficient iterative predictor selection routine, (iv) identifying the minimum mean squared error estimator of a covariance matrix, and (v) optimally binning probabilistic forecasts for the calculation of verification statistics. The problems were overcome, and the solutions were incorporated into the BPO-EFV software system.

5. Capabilities

The BPO-EFV software system is currently capable of performing a comparative verification of probabilistic forecasts made by the BPO and by the Model Output Statistics (MOS) techniques. To facilitate a comprehensive comparative verification study at 28 weather stations across
the continental United States, the software system input file containing the model output and predictand realizations is the formatted output from the Meteorological Development Laboratory’s AVN-MOS archive of output fields from the Global Spectral Model (run under the code name AVN) and predictand realizations. The climatic sample, which comes from the National Climatic Data Center (NCDC), must be reformatted before it is input to the BPO-EFV software system.

The BPO-EFV software system is currently implemented in two parallel configurations. The first implementation is for a binary predictand. The second implementation is for a continuous predictand. While these two configurations share some of the same subroutines, the parameter estimation, forecasting, and verification routines are distinct; the two configurations can be operated independently.

In addition to the two main configurations of the software system, three stand alone software packages have been created. They include a distribution fitting software package, a binary predictand verification software package, and a continuous predictand verification software package. These three packages are described next.

5.1 Special Univariate Distribution Fitting Software

The special univariate distribution fitting software package is an implementation of the uniform estimation method. It is capable of selecting, from a catalog of parametric distribution functions, the parametric distribution function that best fits a given empirical distribution function. This is accomplished by using the uniform estimation method to estimate the parameters of several candidate distribution families and then choosing the distribution family (and its parameter estimates) that minimizes the maximum absolute difference (MAD) between the empirical distribution function and the fitted parametric distribution function. Tables 1–3 list the
distribution families whose parameters are estimated when a particular type of the sample space is specified. The software output includes the empirical distribution function, the best fit parametric distribution function, and the MAD. This software package has wide applicability in fields that require modeling parametric distribution functions, such as risk analysis, decision analysis, signal detection, and operations research.

5.2 Binary Predictand Verification Software

The binary predictand verification software package is capable of producing the verification measures of calibration and informativeness for a probabilistic forecaster or a probabilistic forecast system. This is accomplished by computing the (discrete) calibration function, the (discrete) receiver operating characteristic (ROC), the calibration score, the ROC score, and the uncertainty score. All measures are computed from a joint sample of the forecast and the predictand. The computation of the calibration measures and the uncertainty score requires a climatic (prior) probability. The software package has been used to perform the verification of the probability of precipitation (PoP) occurrence forecasts.

5.3 Continuous Predictand Verification Software

The continuous predictand verification software package is capable of producing the verification measures of calibration and informativeness for a probabilistic forecaster or a probabilistic forecast system. This is accomplished by computing the calibration score and the informativeness score from a joint sample of the forecast and the predictand. The software package has been used to perform the verification of the probabilistic forecasts of precipitation amount, conditional on occurrence.
Table 1. Distribution families on the unbounded interval fit by the software.

<table>
<thead>
<tr>
<th>Distribution Family</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>LG</td>
</tr>
<tr>
<td>Laplace</td>
<td>LP</td>
</tr>
<tr>
<td>Gumbel</td>
<td>GB</td>
</tr>
</tbody>
</table>

Table 2. Distribution families on a bounded-below interval fit by the software.

<table>
<thead>
<tr>
<th>Distribution Family</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>EX</td>
</tr>
<tr>
<td>Weibull</td>
<td>WB</td>
</tr>
<tr>
<td>Inverted Weibull</td>
<td>IW</td>
</tr>
<tr>
<td>Log-Weibull</td>
<td>LW</td>
</tr>
<tr>
<td>Log-Logistic</td>
<td>LL</td>
</tr>
<tr>
<td>Kappa</td>
<td>KA</td>
</tr>
<tr>
<td>Pareto</td>
<td>PA</td>
</tr>
</tbody>
</table>

Table 3. Distribution families on a bounded interval fit by the software.

<table>
<thead>
<tr>
<th>Transformation Class</th>
<th>Distribution Family</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Type I</td>
<td></td>
<td>P1</td>
</tr>
<tr>
<td>Power Type II</td>
<td></td>
<td>P2</td>
</tr>
<tr>
<td>Log-Ratio</td>
<td>Logistic or Laplace</td>
<td>LR1</td>
</tr>
<tr>
<td>Ratio Type I</td>
<td>Each from Table 2</td>
<td>RA1</td>
</tr>
<tr>
<td>Ratio Type II</td>
<td>Each from Table 2</td>
<td>RA2</td>
</tr>
<tr>
<td>Log-Reciprocal Type I</td>
<td>Each from Table 2</td>
<td>LC1</td>
</tr>
<tr>
<td>Log-Reciprocal Type II</td>
<td>Each from Table 2</td>
<td>LC2</td>
</tr>
</tbody>
</table>
6. **Future Software Systems**

Many subroutines in the BPO-EFV software system can be directly reused in an operational forecast system or in a new software system implementing the Bayesian Processor of Ensemble (BPE). In the development of the BPO-EFV software system, many difficult theoretical and computational problems were overcome. By reusing some of the existing subroutines of the BPO-EFV software system, the tasks of designing and developing an operational probabilistic forecasting software system, or a BPE estimation, forecasting, and verification software system, will be simplified.

7. **References**


8. **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 data
for testing. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.