Methods for Post-Processing

Breakout Group Report
Statistical Post-Processing Workshop, 21 Jan 2005
Version 3

Working Group Members

- Kim Elmore
- David John Gagne
- Ping Liu
- David Myrick
- David Rudak
- Zoltan Toth
- John Williams
- One or two additional colleagues?

Postprocessing Requirements

FACT: Statistical Post-Processing (SPP) is a critical step adding significant skill and value to the end-to-end forecast process. SPP needs to (a) combine predictive information from multiple sources; (b) calibrate prognostic variables from numerical forecasts (leading moments of ensemble, including unperturbed forecasts); and (c) derive additional calibrated user relevant variables.

FINDING: There exists, and NOAA uses, a multitude of methods addressing one or more SPP needs.

RECOMMENDATIONS: NOAA should create: (1) a statistical post-processing testbed with (i) sample problems, (ii) datasets and software toolboxes updated with new implementations, (iii) a development and testing environment, and (iv) quantitative assessment metrics for the intercomparison of different methods; (2) a consolidated operational suite systematically addressing SPP needs providing output that are consistent in multiple formats.

SPP Configuration

FACT: Model prognostic variables are not directly suitable for user applications because NWP models: (a) exhibit a lead-time dependent drift; (b) do not carry fine scale user relevant variables.

FINDING: Statistical derivation of calibrated user variables from raw model prognostic variables can be accomplished either (a) by directly connecting raw model prognostic variables to calibrated user variables (1-sweep approach), or (b) by first calibrating raw model prognostic variables (needed for downstream applications such as hydrologic forecasting), and then deriving user variables from them (2-stage approach).

RECOMMENDATIONS: In/for its SPP operations, NOAA should (a) create a calibrated ensemble of model prognostic variables; and (b) determine when user variables are best derived and calibrated from raw (1-sweep approach) vs. calibrated prognostic variables (2-stage approach).

Calibration of Prognostic Variables

FACT: Many users require forecast information on multiple variables. The calibration of covariances across variables is impractical with current methods and achievable sizes of training sample.

FINDING: The univariate calibration of model prognostic variables, followed by an adjustment of ensemble members so they reflect the univariate posterior distributions while retaining the ensemble rank correlation structure, may offer a viable solution.

RECOMMENDATIONS: Evaluate the calibrated/adjusted ensemble as to the potential loss of thermodynamic and other balances across the prognostic variables present in the direct model output, and the quality of forecast covariance information.

Derivation of User Variables / Products

FACT: A multitude of methods exist for the derivation of user variables from model prognostic variables (e.g., Perfect Prog, MOS, Analogs, Decision Tree). In various fields machine learning and Bayesian methods add value over linear methods. Most methods can be applied with raw or calibrated prognostic variables.

FINDING: The methods can be applied in different ways: (a) (i) interpolate prognostic variables to observation sites, (ii) develop derivation method parameters, (iii) apply derivation method at observation sites and interpolate output to fine scale grid (gridded MOS), or interpolate parameters to and run method on fine scale grid (e.g, Weather Company); (b) (i) analyze all observations on fine scale grid, (ii) interpolate prognostic variables to fine scale grid, (iii) apply derivation method (e.g., NAEFS downscaling).

RECOMMENDATIONS: (1) Assess the relative benefits of developing the user variable derivation methods at observation locations vs. on fine scale grid using complex evaluation criteria discussed below; (2) Explore the use of machine learning and Bayesian methods in addition to traditional methods.

Evaluation Criteria

FACT: There exist a host of criteria for evaluating and comparing SPP configurations and methods, including: Quality (statistical reliability and resolution, user relevant metrics); Computation cost; Ease of implementation / maintenance; Scientific Soundness; Extendability for future improvements and extensions.

FINDING: Evidence-based SPP related decisions must rely on a well-defined set of evaluation criteria.

RECOMMENDATIONS: (1) NOAA should define its set of preferably quantifiable SPP evaluation criteria; (2) SPP decisions should be driven by criteria-based evaluation results; (3) When comparing the quality of a method against benchmarks such as persistence or climatological forecasts, or against competing methods, generally permutation or resampling (and not parametric) statistical tests should be used.

Data needs

FACT: SPP applications need a variety of datasets - NWP (re)analyses and (re)forecasts, and corresponding climatology of observations and/or fine scale (re)analyses of user relevant variables - the larger the training samples, the better. The generation of reanalyses / reforecasts incur significant costs.

FINDING: As users require SPP forecasts on fine temporal and spatial scales, there is a need for (a) more frequent (e.g. hourly) analyses & forecasts on fine grids, (b) variability related model output (e.g, hourly max), and (c) more computer and storage capacity.

RECOMMENDATIONS: NOAA should assess (1) the cost of creating, and the benefits from, more reforecasts vs. the benefits of enhancing the real time forecast system at the same cost of generating more reforecasts; and (2) which SPP methods have less dependency on large reforecast datasets.

Background Slides

Approach to Method Development

- Define set of problems
- Provide data / Share tools
  - Create canonical datasets for optional use in development
- Develop and run various methods
- Testing and intercomparison
  - Done on independent (near real time?) data
- Assessment of methods

Postprocessing Requirements

- Combine various sources of predictive info
  - Unperturbed forecasts, ensemble forecasts, latest observations, climatology
  - To provide unified guidance
  - Data needs - climatology of predictand, joint sample of predictors & proxy for truth
- Calibrate model prognostic variables
  - Create calibrated ensemble data
  - For downstream applications - eg, hydrologic ensemble forecasting
  - Data needs - Reanalysis & reforecast for all valid forecast times
- Derive user variables / products
  - From model prognostic variables on fine scales
    - Must be calibrated on resolved spatiotemporal scales
  - For user applications on fine scales
  - Data needs - Depends on 1-sweep vs 2-stage approach (next slide)
    - NWP and observations / obs-analysis requirement
- Output
  - Calibrated univariate cdf/pdf, quantiles, climatological percentiles, joint probabilities, ensemble

Configuration - From model prog. to user variables

- TWO-STAGE APPROACH
  - Calibration of prognostic vars on model grid against model analysis
    - Data needs - Reanalysis, reforecast on model grid
    - Ensemble members adjusted to be consistent with posterior pdf/cdf
    - Is balance retained? Research question
  - Derivation of user variables at arbitrary points from calibrated prog vars
    - Observation locations, or on fine scale grid
    - Data needs - Reanalysis and observations or observationally based fine scale reanalysis
- ONE SWEEP APPROACH
  - From raw prog vars directly to calibrated derived fine scale vars
    - Data needs - Reforecasts and observations or observationally based fine scale reanalysis
    - Is there added value compared to 2-stage configuration? Research question
      - If two stages have nonlinear connection
      - If balance is distorted by calibration on model grid

Methods - Calibration

- Calibration - Univariate, all prognostic vars
  - First 2-3 moments in ensemble adjusted to match corresponding analysis distribution
  - Multiple linear regression
  - BMA, EKDMOS, BPE
- Calibration - Multivariate, for covariance among prog vars
  - Use calibrated univariate distributions & ensemble rank structure
    - Copula methods, multitask methods
  - Calibration of covariance too expensive? Not enough sample data? - Research question
- Combination of all forecast info
  - BMA, Bayesian use of climatology (BPE)...
- Techniques
  - Adaptive estimation of parameters - To follow changes in regimes / DA-model suites
    - Recursive parameter estimation, Kalman filters, etc
  - Avoid overfitting - use regularization

Methods - Derivation of User Variables / Products

- Configuration
  - Most methods can be used either in 1-sweep or 2-stage configuration
- Regression methods
  - Perfect prog
  - Traditional MOS
  - Multiple Linear Regression
  - Logistic regression
  - Ridge, lasso, elastic net regularized methods
  - Bayesian approaches
- Machine learning and other methods
  - Decision Tree ensemble Methods (Random Forests, Gradient Boosting)
  - Analog/Nearest Neighbor methods
  - Neural Networks
- Choice of proxy for truth
  - Observations or observationally based fine scale analysis?
    - Depends on observation density, user needs, etc

Evaluation Criteria

- Statistical quality
  - Reliability
  - Stat. resolution / info content
  - User relevant metrics - Economic value, etc
  - Use permutation / resampling non-parametric tests when comparing to benchmarks
  - Compare against random / climate baselines and competing methods
- Computational cost
- Ease of implementation / maintenance
- Scientific soundness
- Extensibility - path to future improvements / extensions

Summary of data needs

- Prognostic variables from real time and re-analysis/forecast
    - More frequent output for more prognostic variables
    - E.g., hourly precip accumulations
    - Additional parameters on variability (e.g., hourly max, mean, standard deviation)
- Observations and fine scale reanalysis of user variables
- Canonical sample datasets for testing/comparing methods
    - Keep datasets updated as forecast suites change
- Reanalysis / reforecast
    - Reanalysis needed for predictand climatology
    - Reforecasts needed for joint sample with reanalysis
    - Bayesian methods may need smaller sample - Research question
    - Cost of running, and benefit from reforecasts -
    - Compared to enhanced real time forecast system? - Research question
    - Real time vs. batch generation of reforecasts?
    - Logistical issues

Breakout Group

FACT: Kim Elmore, David John Gagne, Ping Liu, David Rudack, Zoltan Toth, John Williams and another colleague participated in the “Methods” Breakout Group (BG)

FINDING: The group had candid and productive discussions

RECOMMENDATIONS: NOAA can request the expert opinion of this BG when developing Statistical Post-Processing (SPP) related plans

Following slides to be deleted

Postprocessing Requirements

FACT: Post-processing requires data from deterministic and/or ensemble NWP model output paired with observations and climatology

FINDING: Analyses/Observations are not available at all forecast times of interest

RECOMMENDATION: There should be greater investment in generating more frequent analyses paired with each operational forecast model

New Data Needs

FACT: Weather variables relevant to users are desired at finer space and time scales

FINDING: Generating more variables at finer scales requires more computing and storage power

RECOMMENDATION: NOAA should determine the cost of running and benefit from more reanalysis/reforecast data, compared to benefits from enhanced real time forecast system.

Research should determine which methods are more or less dependent reforecast data.

NOAA should provide canonical test datasets for major weather variables in conjunction with major model updates for post-processing method comparisons

Calibration Methods

FACT: Univariate calibration of methods corrects systematic biases in first 3 moments of ensemble distributions. This calibration can be done with multiple linear regression, BMA, BPE, EKDMOS

FINDING: Univariate calibration may affect correlations among related variables and decrease the effectiveness of calibrating user-defined variables

RECOMMENDATION: NOAA should evaluate multivariate calibration methods (Copula, Schaake Shuffle, multitask methods) against univariate methods to determine the effects on correlations across calibrated prognostic variables.

Derivation of User Variables/Downstream Variables

FACT: Statistical derivation of downstream and user-relevant weather variables provides added values in many domains

FINDING: Machine learning and post-processing methods have added predictive performance and value over linear methods in many domains both within and outside the weather community

RECOMMENDATION: NOAA needs to be open to nonlinear machine learning and Bayesian methods in addition to perfect-prog and MOS methods and investigate how they can been implemented on a larger scale.

Evaluation

FACT: Post-processing methods need to be judged based on statistical quality (reliability, resolution, economic value), computational costs, maintainability, scientific soundness, and extensibility

FINDING: Current statistical evaluation metrics do not fully convey all the relevant aspects of implementing current or new post-processing methods

RECOMMENDATION: When NOAA evaluates current and new methods, it should incorporating all aspects of their operation in order to determine which method should be implemented operationally. Any statistical