Bayesian Framework Design for Quantifying Model Uncertainty : A Univariate Probabilistic Mixture Model Approach

ECMWF/WWRP Workshop on Model Uncertainty 11-15 April 2016

Husain Najafi[1], Ali Reza Massah Bavani[1]

[1] Department of Water Resources Engineering College of Aburaihan, University of Tehran

On the Importance of Model Uncertainty Quantification

State-of-the-art climate predictions still contain many sources of uncertainty, in particular those arisen from model uncertainty. These uncertainties may obscure the analysis and interpretation of underlying processes at the spatiotemporal scales.

Despite the fact that ensemble forecasts address the major sources of uncertainty, they exhibit biases and dispersion errors and therefore are known to improve by calibration or statistical post-processing.

Subseasonal to interannual Predictability and Prediction Skill

The continuous improvements in seasonal climate forecasting have made it possible to provide more skillful and tailored forecasts. Therefore, any potential predictability can significantly help for operational management decisions for example in hydro energy and agricultural sectors.

With the increasing size of ensemble members provided by operational as well as research canters, now we can understand climate predictability better at subseasonal to interannual time scales.

Aim and Scope

This study introduces a framework to quantify uncertainties associated with sub-seasonal to interannual climate predictions. A framework is provided for Bias correction and Calibration of real-time forecasts based on hindcsast data from seasonal GCM outputs for variable of interest. It is based on postrprocessing ensemble of forecasts from a probabilistic point of view.

From single value - deterministic to ensemble-based probabilistic forecasting, the framework can convey forecast uncertainty in a user relevant form.

Framework Implementation

Building on existing state-of-the-art of climate prediction models that are available,the purpose of this research activity will be to explore the general usefulness of seasonal forecasts for the application side mainly water resources planning and managment.

The outcome of this research is highly valuable for

- Developing early drought alert systems,
- Real time ensemble streamflow prediction
- Components of water balance
- Agricultural insurance

among other applications once has been established .at operational centers.

Concept

Mixture models are widely used as computationally convenient representations for modeling complex probability distributions.

In a fully Bayesian treatment of the mixture modeling problem, conditioning on the observed data leads to a posterior distribution over the number of components. This is mainly important to quantify model uncertainty. Specifically, the skill of climate models may be different in various climatic conditions and this is the case where mixture models might improve the overall result (Probabilistic forecasts of variable of interest) in a multi-model ensemble framework based on Bayesian Model Averaging.

Developing Bayesian Model Average (BMA)

Bayes' theorem provides an approach to update the probability distribution of a variable (e.g. precipitation) based on information newly availabl (NMME hindcast data) by calculating the conditional distribution of the variable given this new information.

The updated (conditional) probability distribution reflects the new level of belief about the variable.

If $\vartheta = (\theta 1, \ldots, \theta K, \eta)$ are unknown parameters that need to be estimated from the data then, as noted earlier, from a Bayesian perspective all information contained in the data **y** about ϑ is summarized in terms of the posterior density $p(\vartheta|y)$, which is derived using Bayes' theorem: $p(\vartheta|y) \propto p(y|\vartheta)p(\vartheta)$

Considering a predicted climatological variable y, the corresponding observational data y_T, and K model forecasts $\{f_1, f_2, \ldots, f_k\}$ of variable y.

In the method of BMA, we assumed that each model forecast, f_k, is associated with a conditional PDF, $g(y|f_k)$, that can be represented by a probability density function. Under the law of total probability, the BMA predictive model, given observational data y_T, can be expressed as

$$p(y|f_1, f_2, \ldots, f_k) = \sum_{k=1}^{K} w_k p(y|f_k) * p(f_k|y_T)$$

where

y:predicted climatological variable

y : the corresponding observational data, and

$\{f_1, f_2, \ldots, f_k\}$: K model forecasts of variable y

p (fk|yT) is the posterior probability of forecast k

w_k : weight of model k

Likelihood Estimation

Expectation– Maximization (EM)

Here we use the Expectation– Maximization (EM) algorithm for estimating the weights w_K based on the performance of each model during the training period and the unknown parameters of the mixture model for likelihood function.

In brief, the EM algorithm casts the maximum likelihood problem as a ''missing data'' problem. The missing data may not be actual data. Rather, it can be a latent variable that needs to be estimated. Two situations will be evaluated :

1-The likelihood is estimated based on the known z_i considering a predefined-categorized climatic conditions (e.g. Below Normal/ Normal/ Above Normal) and each ensemble member corresponded to those defined categories.

2- Finding the optimal z_i based on the EM algorithm

E Step

$$Q(\psi;\psi^{(0)}) = E_{\psi}^{(0)}\{\log L_c(\psi)\,|y\}$$

$$E_{\psi}^{(k)}\{Z_{ij}\,|y\} = \pi_i^{(k)} f_i(y_j;\theta_i^{(k)})/f((y_j;\psi^{k})$$

M Step

$$\hat{\pi}_i = \sum_{j=1}^{n} z_{ij}/n \quad (i = 1,\ldots,g)$$

$$\pi_i^{(k+1)} = \sum_{j=1}^{n} \frac{\tau_i(y_i;\psi^{(k)})}{n} \qquad (i = 1,\ldots,g)$$

Posterior Analysis

Monte Carlo Markov Chain (MCMC)

If we have a conjugate prior (Gaussian for instance), it can be an easy calculation to obtain the posterior, and then we can sample from it.

However, sometimes conjugate or semiconjugate priors are not available or are unsuitable. In such cases, it can be difficult to sample directly from the posterior; Markov Chain Monte Carlo (MCMC) will be applied to sample from posterior and estimate its mean and variance for producing a probabilistic map of variable of interest.

Future Work

The proposed framework will be applied to support water resources planning and drought prediction as a univariate analysis for (precipitation) over west of Iran. Next, the framework will be developed for bivariate analysis of temperature-precipitation analysis to see whether a joint perspective provides added value in different lead times considering better skill of models in temperature prediction at longer lead times if there is any dependence between two variables.