ABSTRACT

DEVINENI, NARESH. Multimodel Ensembles of Streamflow Forecasts: Role of Predictor State in Developing Optimal Combination. (Under the direction of Sankar Arumugam.)

Information on season-ahead streamflow forecasts is beneficial for the operation and management of water supply systems. Developing such long-lead (3-12 months) stream flow forecasts typically depend on exogenous climatic conditions particularly sea surface temperature (SSTs) conditions in tropical oceans. Identification of such conditions that influence the moisture transport into water resources regions is important to develop low-dimensional statistical models and to analyze climatic forecasts from General Circulation Models (GCMs). The main purpose of this study is to develop probabilistic streamflow forecasts for the Falls Lake Reservoir, NC, for the summer season that is critical for developing water management strategies so that the City of Raleigh's water demands could be met through water conservation measures. The study develops two low-dimensional statistical models based on SSTs in the tropical Pacific, tropical Atlantic and over the NC Coast. Given that prediction from any model is bound to have unavoidable error/model uncertainty, the study intends to combine the forecasts from individual models to develop an improved multi-model forecast. For this purpose, the study develops an algorithm for combining forecasts from individual forecasts by evaluating the performance of individual forecasts contingent on climatic (predictor) conditions. The methodology is demonstrated through the development of multi-model ensembles of streamflow forecasts for the Falls Lake reservoir by combining probabilistic streamflow forecasts from two low dimensional statistical models. Using Rank Probability Score (RPS) for evaluating each year's streamflow forecasts for the summer

months (July-August-September) from the two low dimensional models, the methodology proportionately gives higher representation by drawing increased ensembles for a model that has better predictability under similar predictor conditions. The performance of the multi-model forecasts is compared with the individual model's performance using various performance evaluation measures. By developing multi-model ensembles based on leave-one-out cross validation and split sampling, the study shows that evaluating the model's performance based on the predictor state provides a better alternative in developing multi-model ensembles instead of combining the models purely based on their long-term predictability. The method is also extended to combine various GCMs to get improved winter (December-January-February) precipitation forecast for the entire US. Finally the study shows the utility of the multi model precipitation ensembles to develop improved streamflow forecasts for the Falls Lake through statistical downscaling.

MULTIMODEL ENSEMBLES OF STREAMFLOW FORECASTS: ROLE OF PREDICTOR STATE IN DEVELOPING OPTIMAL COMBINATION

by
NARESH DEVINENI

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

CIVIL ENGINEERING

Raleigh, North Carolina

2007

APPROVED BY:

Dr. S. Ranji Ranjithan

Dr. E. Downey Brill

Dr. Sankar Arumugam
Chair of Advisory Committee

DEDICATION

This work is dedicated to my mentor Dr. Sankar Arumugam.

BIOGRAPHY

Naresh Devineni is born on June 23, 1983 in Guntur, a town in Andhra Pradesh, India. He is the younger of the two sons to Mr. Murali Krishna Devineni and Mrs. Jayasree Ramu Devineni. He graduated with a bachelor's degree in Civil Engineering from Osmania University Hyderabad, India. Presently, he is pursuing his Master of Science in Civil Engineering with special focus on Water Resources and Environmental Engineering at North Carolina State University, Raleigh, NC. He plans to continue on a PhD at North Carolina State University.

ACKNOWLEDGEMENTS

I would like to express my profound sense gratitude to Dr. Arumugam for being an excellent mentor and for providing me with necessary guidance and support at every step of my Masters study. I am deeply indebted to him for transforming me not only into a complete engineer but also as a responsible person. I am grateful to Dr. Ranjithan who has given feed back from time to time on my performance and boosted my morale and confidence under difficult situations. I would also like to thank Dr. Downey Brill for his encouraging words on my performance. I am very honored to work under the guidance of these people.

I am always grateful to my parents and grand parents, with whose constant support, encouragement and blessings I was able to come to the United States to pursue my higher degree. I also thank my brother Ravi Chandra Devineni for his constant belief in me and my capabilities as a performer.

I thank Mr. Kurt Golembesky for his valuable feed back and discussions on my thesis. I would like to thank my friends Shravan Chintapatla, Krishi Siravuri and Uday Maripalli for helping me out in difficult situation. A special thanks to Pawan Imanneni with whose guidance I could make it to NC State University. I would like to thank my friends Praveen Gollakota, Sarat Sreepathi, Sridevi Chalasani, Azam Hossain, Jitendra Kumar and Mr.Satish Regonda for their help, support and feedback.

I would like to acknowledge the Water Resources Research Institute (WRRI) of North Carolina for funding the project. Special thanks to Tom Fransen, River basin Management Chief, Division of water resources, DENR, Terry Brown, USACE and Willie Tanner, Public Utilities, City of Raleigh, and the University of East Anglia(UEA) and International Research Institute for Climate and Society(IRI) for providing us with the data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Information on season-ahead streamflow forecasts is beneficial for operation and management of water supply systems and in addressing the issues of droughts and floods. Unless closely monitored using various sector-specific indicators, the impacts of droughts and floods are progressive, persistent and pervasive over a larger area. Prediction of these hydroclimatic extremes well in advance would help local/state water managers and emergency management agencies to develop appropriate contingency measures and alternative water management strategies. For instance, long-lead (3 months to 6 months ahead) prediction of drought will provide vital information in hedging the associated risk as well as in imposing voluntary restrictions for water supply systems.

1.1 Multi Year Drought and Falls Lake Management

Multi-year drought during 1998-2002 caused severe hardship and economic losses across most of North Carolina [Weaver 2005]. Several local and state-wide water supply systems experienced record shortages. Many communities operated under mandatory water conservation plan during 2001-2003 [Yonts W, 2004]. Economic losses in NC for year 2002 were estimated to be $398 million for agriculture and $15-$20 million for municipalities [Hayes MJ et.al, 2004]. Figure 1.1 shows the number of communities that faced voluntary, mandatory and emergency restrictions during 1998-2002 multi-year drought conditions in NC. Though the figure shows the situation only for North Carolina, the pattern was typical for Raleigh and downstream of Neuse River basin during this period. As one can infer from figure 1.1, the severity of droughts is more pronounced during summer months (July – August – September).

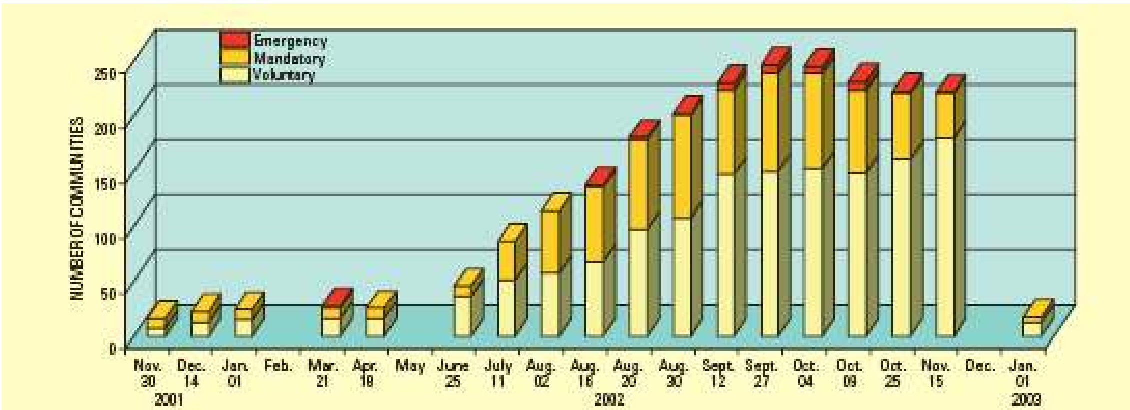


Figure E-1. North Carolina communities operated under mandatory, voluntary, or emergency water conservation from November 2001 through January 2003. (No data were available for February, May, and December 2002; adapted from data provided by Mr. Woody Yonts, North Carolina Department of Environment and Natural Resources, written commun., April 1, 2004.)

Figure 1.1: Recent multi-year drought conditions in NC. Figure shows the number of communities in North Carolina that were under voluntary, mandatory and emergency restrictions from 2001-2003 [Yonts W, 2004].

Falls Lake (location shown in Figure 1.2) is a multipurpose reservoir authorized for flood control, water supply, water quality, recreation and for fish/wildlife protection. The state capital Raleigh gets its supply directly from the dam. Falls Reservoir is an earthen structure having a top elevation of 291.5 msl and extends 28 miles upstream up to the confluence of Eno and Flat Rivers. The top of the conservation pool is 251.5' msl having a storage capacity of 131,395 acre-feet. The flood control pool has a controlled capacity of 221,182 acre-feet and an uncontrolled capacity of 749,010 acre-feet. Given that Falls Lake is the upper most reservoir in the Neuse, releases from the reservoir are crucial for meeting downstream water quality as well as for in stream flow maintenance.

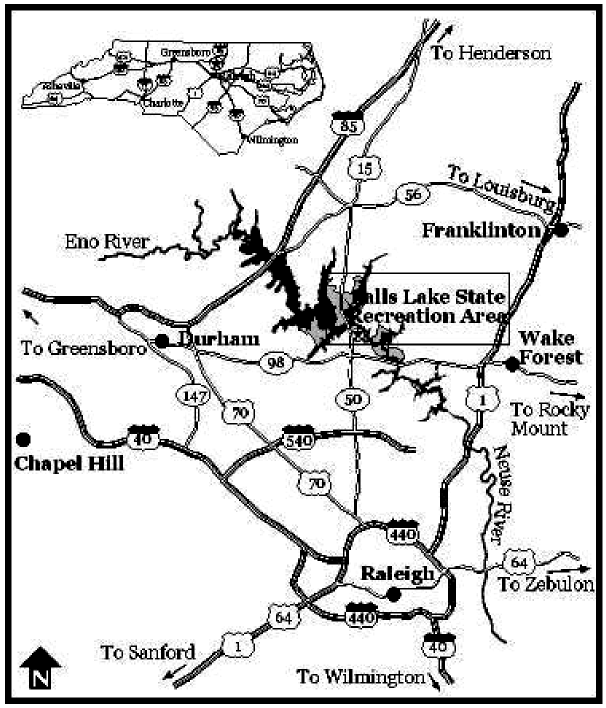


Figure 1.2: Location of the Falls Lake Reservoir in Raleigh NC.

1.2 Climate Forecasting and Water Supply Management

Weather forecasts are useful in predicting significant weather patterns at near- term (1 or 2 days). As the number of days increases, the skill of the forecast decreases. While weather forecasts are essential in terms of meeting the needs for peak hydro power generation and flood management, information on season-ahead streamflow forecasts is beneficial for the operation and management of water supply systems. Developing such long-lead stream flow forecasts depends on exogenous climatic conditions, such as sea surface temperature conditions (SST) which could provide information on the probability of inflows over the upcoming season. Such reservoir inflow forecasts contingent on climatic conditions could be effectively utilized for operating reservoirs, and for invoking restrictions to improve water supply planning and management.

1.3 Streamflow Forecasts- Development Methodologies

Interest in development and application of seasonal to interannual (long-lead) streamflow forecasts has grown tremendously over the last decade primarily due to the improved monitoring of Sea Surface Temperature (SST) in the tropical Pacific as well as due to the issuance of operational climate forecasts from GCMs by various centers and research institutions on a monthly basis. The goal of this study is to develop a seasonal and long-lead streamflow forecasting model that predicts the flow quantile conditioned on exogenous climatic indices. One way to approach this problem is to take the precipitation and temperature from the GCM which is at a resolution of 2.5 by 2.5 degree. Utilizing GCM predicted fields of precipitation and temperature for developing streamflow forecasts require downscaling, since GCM outputs usually are given at large (2.5° X 2.5°) spatial scales. Dynamical downscaling nests a regional climate model (RCM) with GCM outputs as boundary conditions to obtain precipitation and temperature at watershed scale (60 Km X 60 Km). The downscaled precipitation and temperature at watershed scale could be used further as inputs into a watershed model to obtain seasonal streamflow forecasts [Leung et al., 1999; Roads et al., 2003; Seo et al. 2003, Carpenter and Georgakakos, 2001]. An alternative would be to use statistical downscaling, which maps the GCM precipitation and temperature forecasts to observed streamflow forecasts at a given point through a statistical relationship [Robertson et al. 2004, Landman and Goddard, 2002; Gangopathyaya et al., 2005]. A low dimensional statistical model without using GCM outputs by relating the observed streamflow to identified climatic precursors (e.g., El Nino Southern Oscillation (ENSO) indices) that influence the streamflow potential at the given site can also be developed [Hamlet and Lettenmaier, 1999; Souza and Lall, 2003; Sankarasubramanian and Lall, 2003].

1.4 Importance of Multi Model Streamflow Forecasts

Seasonal streamflow forecasts obtained using the above mentioned approaches are better represented probabilistically in the form of ensembles to represent the uncertainty, particularly in quantifying the effects of both changing boundary conditions (SST) and initial conditions (atmospheric and land surface conditions). Apart from these uncertainties resulting from initial and boundary conditions, the model that is employed for developing streamflow forecasts could also introduce uncertainty in prediction. In other words, even if streamflow forecasts obtained by dynamical downscaling are forced with observed boundary and initial conditions (perfect forcings), it is inevitable that the simulated streamflows will have uncertainty in prediction, which is otherwise known as model error/uncertainty. A common approach to reduce model uncertainty is through refinement of parameterizations and process representations in the considered model, which could be either GCMs or RCMs or hydrologic models. Given that developing and running GCMs is time consuming, recent efforts have focused on reducing the model error by combining multiple GCMs to issue operational climate forecasts [Rajagopalan et al., 2002; Robertson et al., 2004; Barnston et al., 2003; Doblas-Reyes et al., 2000; Krishnamurthi et al., 1999]. Similarly, studies have also shown that developing multimodel forecasts by combining different low dimensional streamflow forecasting models show considerable improvement over the performance of individual models [Regonda et al., 2006]. Thus, combining streamflow forecasts from multiple models seems to be a good alternative in improving the overall predictability of seasonal streamflow forecasts and reducing the overall error in prediction. One of the main objectives of this study is to develop and apply a new scheme for combining forecasts from multiple models, which could be either streamflow forecasts from low dimensional models or

GCM forecasts available at large spatial scales, by assessing the model's predictability conditioned on the predictor state. The basic reason leading to better performance of multi-model ensembles is due to the incorporation of realizations from various models, thereby increasing the number of ensembles to represent the conditional distribution of climatic attributes. Recent studies on improving seasonal climate forecasts using optimal multi-model combination techniques assign weights for a particular model based on its ability to predict the climatic variable over the entire period for which the GCM simulations are available [Rajagopalan et al., 2002; Robertson et al., 2004; Barnston et al., 2003]. Given that each model's predictability could also vary depending on the state of the predictor (SSTs for GCMs), a new methodology for multi-model ensembling that assigns weights to each model by assessing, contingent on the predictor state, the skill of the models is developed. The proposed methodology is employed upon two low dimensional seasonal probabilistic streamflow forecasting models that primarily use tropical Pacific and Atlantic SST conditions to develop multi-model ensembles of streamflow forecasts. The methodology is also extended to combine three precipitation forecasts from different GCMs for DJF (ECHAM 4.5 developed by Max Plank Institute, CCMv6 developed by NCAR, National Center for Atmospheric Research and COLA, Center for Ocean-Land-Atmosphere Studies) to develop a new multi model precipitation forecasts that improve the predictability over the entire USA. This is further employed to develop streamflow forecasts for the Falls Lake reservoir by statistical downscaling.

1.5 Outline of the Thesis

Chapter 2 provides a brief background on multi-model ensembling techniques that is currently pursued in the literature for developing operational climate and seasonal streamflow forecasts. Chapter 3 discusses two low dimensional streamflow forecasting models that were employed for developing probabilistic streamflow forecasts for predicting the summer flows (July – August – September, JAS) into Falls Lake, Neuse river basin NC. Chapter 4 presents the proposed multi-model ensembling scheme that assesses the skill of the model contingent on the predictor state. In chapter 5 the proposed multi-model ensembling is employed to develop improved probabilistic streamflow forecasts for predicting JAS inflows into the Falls Lake, NC. Chapter 6 discusses the application of the proposed methodology for combining precipitation from multiple GCMs and applies it for downscaling to streamflow forecasts into Falls Lake during December – January – February, DJF. Finally, chapter 7 summarizes the findings of the study along with potential applications for combining other environmental simulation models.

CHAPTER 2

LITERATURE REVIEW

Hydroclimatic extremes like droughts and floods are generally associated with low frequency climate fluctuations like El Niño Southern Oscillations (ENSO) and decadal and interdecadal climatic modes such as Pacific Decadal Oscillation (PDO) and North Atlantic Oscillation (NAO). ENSO is a quasi oscillatory mode of coupled ocean atmosphere interactions in the tropical Pacific with a characteristic narrow band width of 2 to 7 years. These climate modes govern the interannual variability of climate over most of North America. The following section details the work done in understanding these climatic modes and its teleconnection to the climate variables like precipitation and streamflows.

2.1 Sea Surface Temperature (SST) – Streamflow Teleconnection

Recent progress in understanding ocean-atmosphere interactions show that there are well organized modes of interannual and interdecadal variability in climate that modulate the dominant moisture delivery pathways and has significant projections on the continental and regional scale rainfalls and streamflow patterns. [Trenberth and Guillemot 1996; Cayan et al., 1999; Dettinger et al., 2000b; Guetter and Georgakakos 1996; Piechota and Drucap, 1996]. Efforts in understanding the linkages between exogenous climatic conditions such as tropical sea surface temperature (SST) anomalies to local/regional hydroclimatology over the U.S. have offered the scope of predicting the rainfall/streamflow potential on a season ahead and long-lead (12 to 18 months) basis [Hamlet and Lettenmaier, 1999; Georgakakos, 2003; Wood et al., 2002; Wood et al., 2005]. Interannual modes such as El Nino-Southern Oscillation (ENSO) resulting from anomalous Sea Surface temperature conditions in the tropical Pacific

Ocean primarily determine the interannual variability in precipitation over North and South America [Rasmusson and Carpenter 1982, Ropeleweski and Halpert, 1987]. Studies have shown that ENSO conditions also influence anomalous SST conditions in the tropical Atlantic and Indian Ocean, hence affecting global climate [Enfield 1989]. There are also other dominant decadal and interdecadal climatic modes such as Pacific Decadal Oscillations (PDO) and North Atlantic Oscillations (NOA) that putatively govern the interannual variability in climate over North America [Sankarasubramanian and Lall, 2003].

During the two phases of ENSO, El Nino and La Nina, anomalous SST conditions in the tropical Pacific are communicated to the extra-tropics through ocean atmospheric circulation in the form of upper tropospheric divergence anomalies. These translate into a modulation of storm tracks over the extra tropics and exhibit teleconnections influencing the distribution of temperature and precipitation across the globe [Ropeleweski and Halpert, 1987]. Cayan et al. [1999] showed that the frequency distribution of daily winter precipitation and winter spring daily streamflow in the western United States exhibit strong and systematic responses to the two phases of ENSO. Pizzaro and Lall [2002] showed that the annual maximum peak over the western United States is significantly correlated to the modes of ENSO and PDO. Jain and Lall [2001] identified space time frequency patterns that connect floods at multiple locations in the western United States with concurrent hemispheric SST and sea level pressure patterns. Most of the studies focusing on climatic variability over South Eastern US have shown that warm tropical Pacific conditions lead to below normal precipitation during summer and above-normal precipitation during winter [Schmidt et al., 2001; Lecce, 2000; Hansen et al., 1998; Zorn and Waylen, 1997]. Studies have also reported

ENSO-related teleconnection between precipitation and temperature during both winter and summer seasons over NC [Roswintiarti et al., 1998; Rhome et al., 2000].

Thus, associating seasonal to interannual variations in streamflow variability with low frequency climatic variability will provide useful information in developing season ahead streamflow forecasts contingent on climatic conditions. In this study, we develop season ahead streamflow forecasts for the Falls Lake using two low dimensional models and then combine them to develop multi model ensembling streamflow forecasts for Falls Lake. The next section provides a brief background on multi model ensembling techniques that are currently pursued in the literature for developing operational climate and streamflow forecasts.

2.2 Model uncertainty and Multi-Model combination methods

Efforts to address model uncertainty through combining outputs from multiple models have been investigated in climate and weather forecasting [Doblas-Reyes et al., 2000; Rajagopalan et al., 2002; Krishnamurthi et al., 1999] and in streamflow simulation through calibration [Boyle et al., 2000; Georgakakos et al., 2003; Marshall et al., 2006]. Perhaps the simplest approach to develop multi-model forecasts is to pool the predicted values or the ensembles from all the models, thus giving equal weights for all the models [Palmer et al., 2000]. Recent research from PROVOST (PRediction Of climate Variations On Seasonal to interannual Time-scales) shows that multi-model ensembles of climate forecasts provided improved reliability and resolution than the individual model forecasts [Palmer et al., 2000; Doblas-Reyes et al., 2000]. Though the improved predictability of multi-model ensembles partly arise from increase in the sample size, studies have compared the performance of single models having the same number of ensembles as the pooled multi-model ensembles and have shown that multi-model approach naturally offers better predictability because of the ability to incorporate outcomes from multiple models, thereby encompassing underlying different process parameterizations and schemes [Hagedorn et al., 2005]. Since the advantage gained through multi-model ensembling is a better representation of conditional distribution of climatic attributes, it is important to evaluate probabilistic forecasts developed from multi-model ensembles through various performance evaluation measures and by analyzing the predictability for various geographic regions [Doblas-Reyes et al., 2005]. Recent studies have also considered climatology as one of the forecasts in developing multi-model ensembles [Rajagopalan et al., 2002; Robertson et al., 2004].

Another approach that is currently gaining attention is to develop a strategy for combining multi-model ensembles using either optimization methods [Rajagopalan et al., 2002; Robertson et al., 2004] or by statistical techniques [Krishnamurthi et al., 1999]. Incorporation of multi-model ensembling techniques to develop operational climate forecasts has also been shown to improve the forecast reliability resulting in better correspondence between observed relative frequency and their forecast probability [Barnston et al., 2003]. Under optimal combination approach, weights are obtained for each model as a fraction, such that the chosen skill/performance measure of the multi model ensembles constituted using these fractions is maximized [Rajagopalan et al., 2002; Robertson et al., 2004; Regonda et al., 2006]. The easiest approach to obtain weights for multi-model ensembles is to give a higher weight for a model that has lower forecast error (such as RMSE). Methods that employ statistical methods such as linear regression has also been employed so that the developed multi-model forecasts has better skill than single models [Krishnamurthi et al., 1999]. However, application of optimal combination approach using either statistical or optimization techniques require observed climatic or streamflow attributes at a particular grid point or station. Studies have also used advanced statistical techniques such as canonical variate method [Mason and Mimmack, 2002] and Bayesian hierarchical method [Stephenson et al., 2005] for developing multi-model combinations. Hoeting et al., [1999] show that the mean of the posterior distribution of the predictand obtained by averaging over all the models with its probability of occurrence provides better predictive ability (measured by logarithmic scoring rule) than the mean of the posterior distribution of the predictand obtained from a single model.

The multi-model ensembling method proposed here is motivated by the fact that the skill of the GCM forecasts or downscaled streamflow forecasts depends on the predictor conditions that determines the conditional distribution of the hydroclimatic attributes. Studies focusing on the skill of GCMs show that the overall predictability of GCMs is enhanced during ENSO years over North America [Brankovic and Palmer, 2000; Shukla et al., 2000; Quan et al., 2006]. Similarly, studies have also shown the importance of various oscillations or climatic conditions in influencing the predictability of GCMs over various parts of the globe. For instance, Giannini et al., [2004] show that tropical Atlantic variability (TAV) plays as a preconditioning state in the development of ENSO related teleconnection in determining GCM's ability to predict rainfall over North East Brazil, which is a region shown to have significant skill in seasonal climate prediction. [Moura and Shukla, 1981;

Ropelewski and Halpert, 1987] and references therein. Giannini et al., [2004] show that the predictability of Nordeste rainfall using CCM3 GCM [Kiehl et al., 1998] is poor particularly if the North Atlantic SSTs exhibit opposite anomalous conditions to the tropical Pacific SSTs. More precisely, with positive SST anomalies in tropical Pacific (warm) and negative SST anomalies (cold) in North Atlantic as well as under cold tropical Pacific (negative SST anomalies) and warm North Atlantic conditions (positive SST anomalies), the predictability of Nordeste rainfall by CCM3 is negative. Naturally, under these predictor conditions, one would prefer to use climatology instead of climate forecasts, since they are negatively correlated with the observed rainfall. Several studies show that the predictive ability of GCMs is dependent highly on ENSO conditions [Quan et al., 2006 and Brankovic and Palmer, 2000; Shukla et al., 2000]. Thus, for post-processing of individual model's climate forecasts to develop multi-model ensembles, one needs to assess the skill of the individual model ensembles based on the predictor state. By considering climatology as one of the candidate forecasts, we develop a multi-model ensembling scheme that formally assesses and compares the skill of the competing models under a given predictor conditions so that lower weights are assigned for a model that has poor predictability under such conditions. The next chapter describes the two forecast models that were developed for summer months of July-August-September using both leave-one-out cross validation and adaptive forecasting methods.

CHAPTER 3

SEASONAL STREAMFLOW FORECASTS DEVELOPMENT FOR THE FALLS LAKE

Development of probabilistic seasonal streamflow forecasts from two different models based on climate information for the Falls Lake, Neuse river basin in North Carolina (NC) is the first objective of this study. Streamflow forecasts based on two low dimensional statistical models, one based on parametric regression approach and another using a nonparametric approach based on resampling [Souza and Lall, 2003] were developed. A brief baseline information about the Neuse basin and its importance to the water management of the research triangle area of NC is provided in the next section.

3.1 Hydroclimatology of Neuse Basin

Falls Lake is a multipurpose reservoir authorized for flood control, water supply, water quality, recreation and for fish/wildlife protection. Given that the water demand in the Triangle area has been growing rapidly in the last decade, multi-year droughts (1998-2002) and ensued restrictions has increased the importance of long-lead forecasts towards better management of water supply systems. Observed streamflow information at Falls Lake is available from 1928 to 2002 from United States Army Corps of Engineers (USACE) (http://epec.saw.usace.army.mil/fall05.htm). Figure 3.1 provides the seasonality of inflow into Falls Lake. Typically, 46% of the annual inflow occurs during January – February – March (JFM), and the low flows during July – August – September (JAS) contribute 14% of the annual inflows. From a water management perspective, developing streamflow forecasting models for the low flow season is important since maintaining the operational rule curve of 251.5’ is challenging during those months.

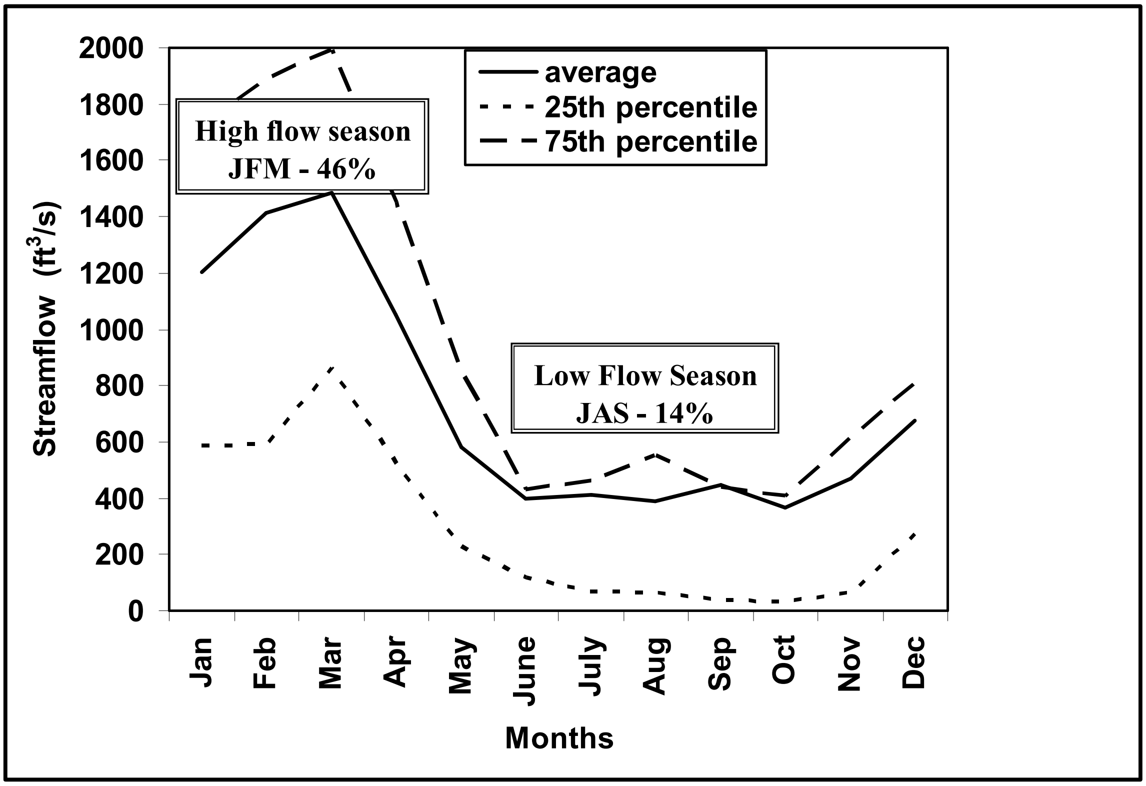


Figure 3.1: Seasonality of Neuse River Basin

3.2 Seasonal Streamflow Forecasts Development – Individual Models

The key objective is to estimate the conditional distribution of streamflows, f(Q_t|$\mathbf{X}_t$), that would occur in the upcoming season based on the climatic conditions $\mathbf{X}_t$ using the chosen statistical model. The estimate of the conditional distribution of streamflow forecasts is $Q_{i,t}^{m}$ with‘t’ denoting the time, ‘i’ representing the ensemble and ‘m’ denoting the model. Based on the observed streamflow, Q_t and the predictors $\mathbf{X}_t = \lfloor x_{1t} \quad x_{2t} \quad \cdots \quad x_{pt} \rfloor$, ($\mathbf{X}_t$ could be SST conditions or principal components of SST over a particular domain such as tropical Pacific) where p is the number of predictors, the conditional distribution of streamflows could be estimated through a parametric approach which explicitly specifies a functional form (e.g., log normal) for the conditional distribution. The other method is to use a data driven approach which estimates the conditional distribution by using nonparametric techniques such as resampling. For the parametric approach, a regression model by assuming the flows follow a lognormal distribution is employed. The estimate of the conditional mean and standard deviation of the lognormal parameters are obtained from the regression estimate and from the point forecast error respectively, which is computed based on the variance of the residuals. Using the lognormal parameters of conditional mean and conditional standard deviation, ensembles from lognormal distribution are generated and are transformed back into the original space to represent the conditional distribution of flows, $Q_{i,t}^{m}$. The other approach is the semi-parametric resampling algorithm reported by Souza and Lall [2003]. The main advantage of this approach is that it does not specify any functional form for estimating the conditional distribution, thus allowing the data to describe the conditional distribution by considering climatic conditions that are similar to current conditions.

3.3 Predictor Identification

Developing season-ahead reservoir inflows requires identification of predictors which could be either SSTs or atmospheric conditions, such as mean sea level pressure. Since long records of streamflow, SSTs and atmospheric conditions are available, statistical approaches that relate "at-site" hydrology to large scale ocean and atmospheric state variables could be developed for forecasting reservoir inflows [Sharma 2000]. Hence, the first step is to identify the location of predictors that influence the streamflow potential into Falls Lake. To identify predictors that influence the streamflow into Falls Lake during JAS, SST conditions during April-June (AMJ) which could be obtained from IRI data library were considered. (http://iridl.ldeo.columbia.edu/expert/ SOURCES/.KAPLAN/.EXTENDED/.ssta).

Predictors are identified using Spearman rank correlation measures which are more powerful in detecting non-linear dependencies between the predictor and the predictand. Figure 3.2 shows the spearman rank correlation between the observed streamflow during JAS at the Falls Lake and the SST conditions during AMJ. From the figure 3.2, we see clearly that SST over ENSO region (170E - 90W and 5S - 5N), the North Atlantic region (80W- 40W and 10N - 20N), and the NC Coast region (75W- 65W and 22.5N- 32.5N) influence the summer flows into Falls Lake. An important note is that SST regions whose correlations are significant and greater than the threshold value of $\pm 1.96/\sqrt{n-3}$ where 'n' is to the total number years (n=75 years for Falls Lake) of observed records used for computing the correlation were considered. Figure 3.2 also shows the 3 month lag correlation between the identified predictors and the streamflow at the Falls Lake reservoir. The negative correlation indicated in figure 3.2 suggests that above normal conditions in the Sea Surface Temperature in the tropical Pacific will influence below normal conditions in the Falls Lake and vice versa.

3.4 Dimension Reduction – Principal Component Analyses on SSTs

Given that SSTs at various grid points exhibit correlation, it is important to identify dominant components so that it explains the maximum variance exhibited by SSTs. The dimension reduction or identifying dominant components is done by performing Principal Component Analysis on the predictor data to get principal components that are a linear combination of the initial values.

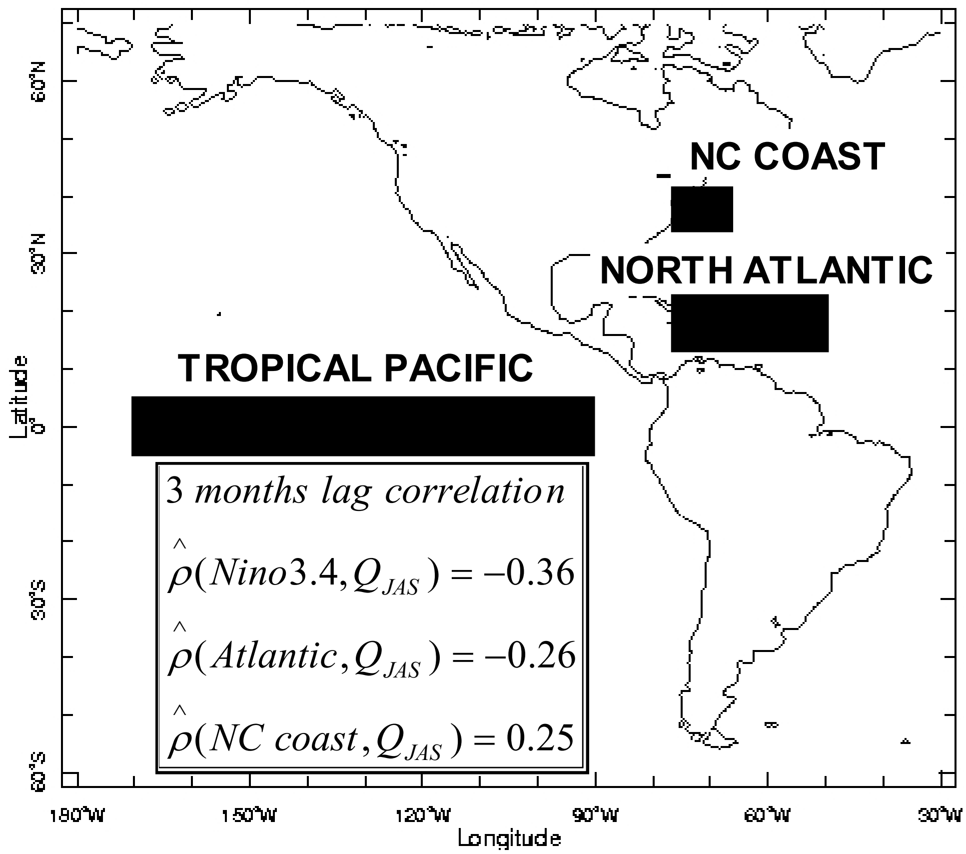


Figure 3.2: Predictor Identification. Figure shows the SST regions that influence the streamflow into the Falls Lake. SST regions that has significant correlation at 95% confidence interval (> 0.22 or < -0.22) are only considered for model development. Also shown in the figure is the 3 month lag correlation of the identified predictors and the streamflows at the Falls Lake.

Principal Components Analysis (PCA) is a multivariate procedure which rotates the data such that maximum variability is projected onto the axes. Essentially, a set of correlated variables are transformed into a set of uncorrelated variables which are ordered by reducing the variability. The uncorrelated variables are linear combinations of the original variables, and the last of these variables can be removed with minimum loss of real data. The first principal component is the combination of variables that explains the greatest amount of variation in the original predictor. The second principal component defines the next largest amount of variation that is remaining and is independent to the first principal component. There can be as many possible principal components as there are variables. In general the m^{th} principal component is the weighted linear combination of the X's or the predictor data set.

Principal component analysis generally gives three outputs, principal components (scores), Eigen vectors (loadings), Eigen values (% variance explained). The first few components explain most of the variance of the original data. The first few eigenvectors will point in the directions where the data jointly exhibits large variations. The remaining eigenvectors will point to directions where the data jointly exhibits less variation. For this reason, it is often possible to capture most of the variation by considering only the first few eigenvectors. The Eigen vectors are useful to locate the source of variability. The variance of the m^{th} principal component is the m^{th} eigenvalue. Therefore, the total variation exhibited by the data is equal to the sum of all eigenvalues. The Eigen values are useful to choose the dominant principal components. A Scree Plot is a simple line segment plot that shows the fraction of total variance in the data as explained by each component. Mathematics of PCA and the issues in selecting the number of principal components using scree plot could be found in Dillon and Goldestein [1984], Wilks [1995] and Von storch and Zweiers [1998].

Given that the SST fields are correlated to each other, Principal Components Analysis (PCA) to identify the dominant modes in the SST field is applied. PCA, also known as empirical orthogonal function (EOF) analysis, on the predictors (SST fields) could also be performed by singular value decomposition (SVD) on the correlation matrix or covariance matrix of the predictors. Since PCA is scale dependent, loadings (Eigen vectors or EOF patterns) obtained from covariance matrix and correlation matrix are different. Importance of each principal component is quantified by the fraction of the variance the principal component represents with reference to the original predictor variance, which is usually summarized by the scree plot. Figure 3.3 shows the percentage of variance explained by each principal component, and the first two components account for 72% of the total variance shown in the predictor field in Figure 3.2. Based on the eigenvectors obtained from PCA, the first component representing the ENSO region has correlation of 0.36 with observed streamflow and the second component representing the Atlantic has a correlation of -0.23 (significance level +/- 0.22 for 75 years of record) with the inflows at Falls Lake. We employ these two principal components to develop seasonal streamflow forecasts for JAS for the Falls Lake.

3.5 Performance of Individual Models

Parametric approach assumes one single model for the entire data, whereas non parametric model is a data driven approach which assumes a particular form locally. For parametric approach, we employ simple regression approach. For non parametric approach, we used a resampling approach. Both models are tested and validated using the leave-one-out cross validation and split sampling validation. Forecasts from both the models are verified using various statistics, such as root mean square error, correlation, average RPS and average RPSS.

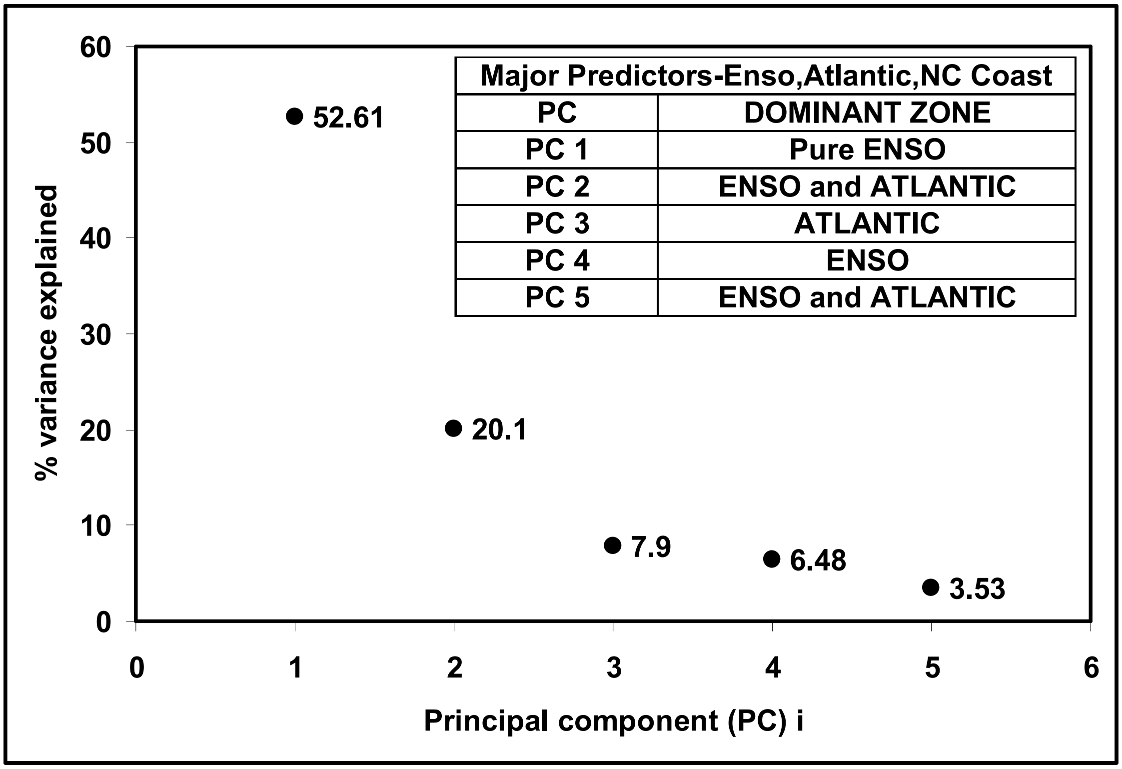


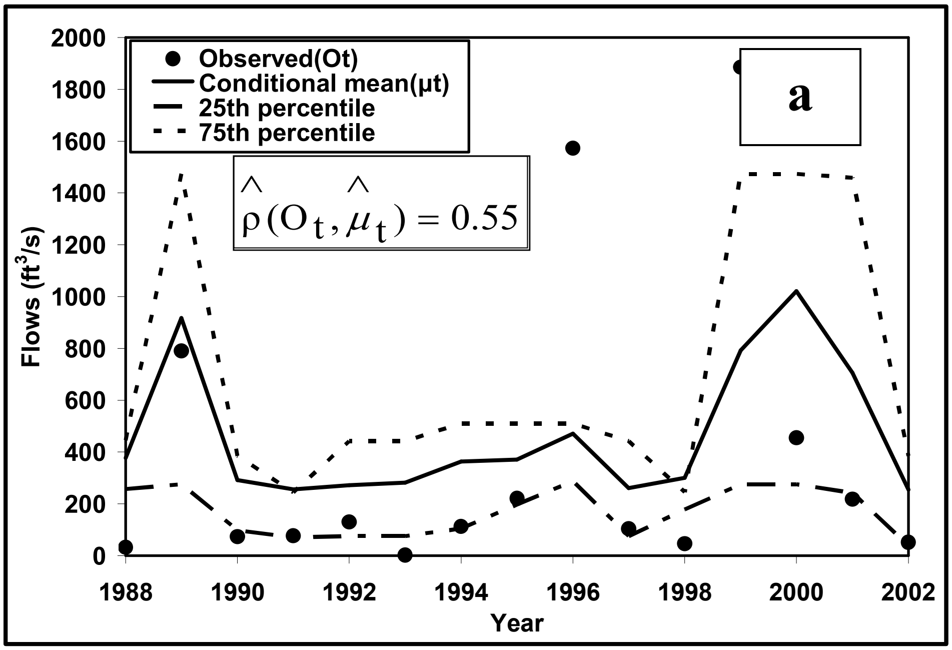
Figure 3.3: Scree plot of the principal components of the SSTs in the three regions (in Figure 3.2) indicating the % variance explained by each component. The dominant zone of each PC obtained based on eigen vectors of the PC is also indicated.

By utilizing the two principal components from PCA, both leave-one-out cross validated retrospective streamflow forecasts and adaptive streamflow forecasts for the season JAS are developed using the two mentioned statistical models. Leave-one-out cross validation is a rigorous model validation procedure that is carried out by leaving out the predictand and predictors from the observed data set (Q_t, $\mathbf{X}_t$, t =1, 2, n) for the validating year and the model is developed using the rest of the (n-1) observations. For instance, to develop retrospective leave-one-out forecasts from parametric regression, a total of ‘n’ regression models are developed by leaving out the observation in each validating year. By employing the developed forecasting model with (n-1) observations, the left out observation (Q_{-t}, with -t denoting the left out year or the validating year) is predicted by using the state of the predictor/principal components ($\mathbf{X}_{-t}$,) in the validating year. To obtain adaptive streamflow forecasts, we develop the forecasting models based on the observed streamflow and the two dominant principal components from 1928-1987 and employ the developed model to predict the streamflow for a 15 years period from 1988-2002.

Table 3.1 gives various performance measures of the probabilistic forecasts from both models. Figure 3.4 shows the adaptive streamflow forecasts for both parametric regression and the semi-parametric models. The correlation between the observed streamflows and the ensemble mean of the cross validated forecasts for resampling and regression approach is 0.40 and 0.35 respectively, which is significant for the 75 years of observed record. From Table 3.1, we infer that the correlation between the observed streamflows and the ensemble mean of the adaptive forecasts is 0.55 and 0.65 for resampling (Figure 3.4a) and regression (Figure 3.4b) approach, respectively. Correlations are significant at 95% confidence level (± 0.51). Table 3.1 also shows other performance evaluation measures such as RPS, RPSS and Root Mean Square Error (RMSE) for adaptive and leave-one-out cross validated forecasts for both models. Since the correlations between observed and ensemble mean is significant for both models under leave-one-out cross validated forecasts and adaptive forecasts, we employ both parametric and nonparametric approaches for developing multi-model ensembles for the Falls Lake system.

Table 3.1: Performance of individual models forecasts under leave-one-out cross validated forecasts and adaptive forecasts. The performance evaluation measures are calculated based on 75 years of data for leave-one-out cross validated forecasts and 15 years for adaptive forecasts from 1987-2002.

	Leave1-out Cross validated (1928-2002)				Adaptive Forecasts (1988-2002)			
	Correlation	RMSE	RPS	RPSS	Correlation	RMSE	RPS	RPSS
Resampling	0.40	423.03	0.43	-0.03	0.55	482.98	0.43	0.00
Regression	0.35	430.93	0.56	-0.30	0.66	477.82	0.61	-0.07



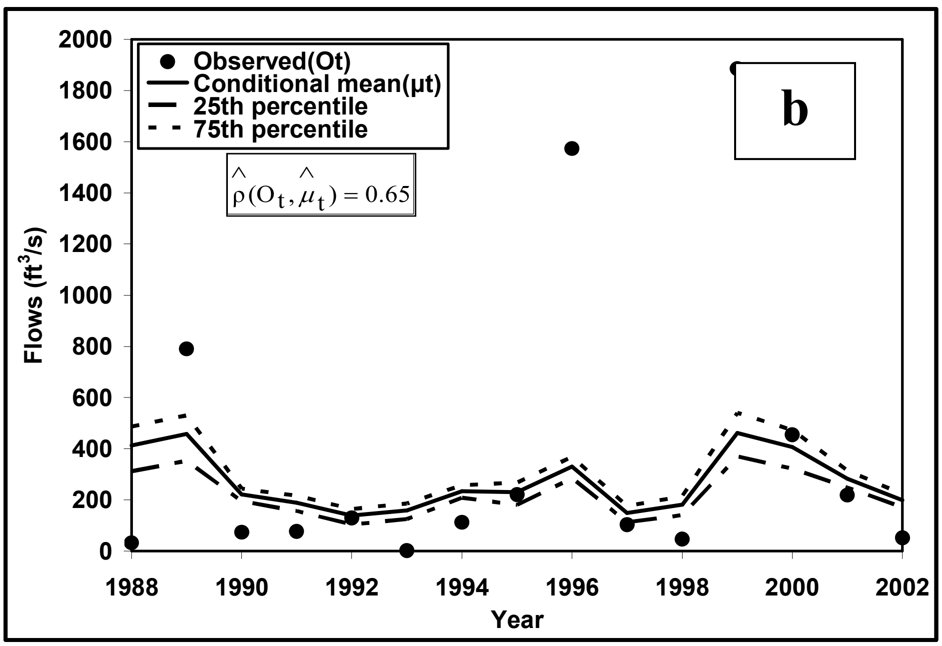


Figure 3.4: Performance of individual models in predicting observed streamflows during 1988-2002 for the Falls Lake. 3.4(a) Semi-parametric resampling model of De Souza and Lall [2003] 4(b) parametric regression. Forecasts from both the models were obtained by using the observed streamflows during JAS and predictors (PC1 and PC2 in figure 3.3) for the period 1928-1987.

CHAPTER 4

MULTI-MODEL ENSEMBLING BASED ON PREDICTOR STATE: METHODOLOGY DEVELOPMENT

Error resulting from climate forecasts is primarily of two types: (a) Uncertainty in initial and boundary conditions and (b) Model error [Hagedorn et al., 2005]. The first source of error is typically resolved by representing the uncertainties in initial and boundary conditions in the form of ensembles. The second source of error arises from process representation, which could be reduced by combining forecasts from multiple models which incorporate various process representation and model physics to develop an array of possible scenarios of outcomes. Developing multi-model ensembles combines these two strategies resulting in reducing both sources of error. However, even after developing multi-model ensembles could result with observations occurring outside the realm of these models (see Figure 8 in Hagedorn et al., 2005). Similarly, the performance of individual models and multi-model ensembles may be poor during certain boundary/SST conditions owing to limited relationship between SST conditions and precipitation/temperature over a particular location/grid [Goddard et al., 2003]. Under these climatic conditions with all models having poor predictability, it may be useful to consider climatology as a forecast.

4.1 Motivation

Figure 4.1 demonstrates the motivation behind the proposed methodology by employing a mixture of regression models that depends on two predictors with the dominant predictor (X1) influencing the predictand only if it crosses a certain threshold ($|X1| > 1.0$). Realizations shown in Figure 4.1a are generated with the predictand y depending on two predictors, $\boldsymbol{X}_t = (\boldsymbol{x}_{1t}, \boldsymbol{x}_{2t})$ with x_1 influencing the predictand only if the absolute value of the predictor x_1 is greater than the threshold value of 1. The underlying model is $y_t = 2x_{1t}+0.5x_{2t} +\varepsilon_t$ if $|x_{1t}| > 1$ and $y_{1t} = 0.25x2_t +\varepsilon_t$ if $|x_{1t}| \leq 1$. The noise term ε_t follows i.i.d with a normal distribution having zero mean and a standard deviation of 2. The predictors follow uniform distribution between -2 to 2. A total of n = 1000 realization is generated from this mixture model setup which could be analogously compared to two predictors as anomalous SST conditions influencing the local hydroclimatology. The correlation between y and x_1 is 0.671 and y and x_2 is 0.134, which would suggest one to give higher importance to predictor x_1. Figure 4.1b shows the skill (correlation) of the fitted regression model between y and x_1 against x_1. To evaluate the correlation between y and the fitted values (y on x_1) against x_1, we consider a bandwidth of 1 on x_1 such that the fitted values of y obtained using the predictor x_1 within that bandwidth are only considered. Note the poor skill between predictand y and the fitted values of y during x_{1t}= -0.5 to 0.5.

Developing a model based on the dominant predictor alone would result in poor prediction particularly when |X1| is below the threshold value (Figure 4.1b). Thus, our approach of multi-model ensembling gives emphasis for assessing the model performance based on the boundary conditions, the predictor state. For instance, if the predictability of all models is really bad during a particular condition, then one would replace model forecasts with climatology by assigning higher weights for climatological ensembles. The following section formally describes the multi-model ensembling procedure that could be employed upon a given set of forecasts from multiple models and the predictors that influence those forecasts.

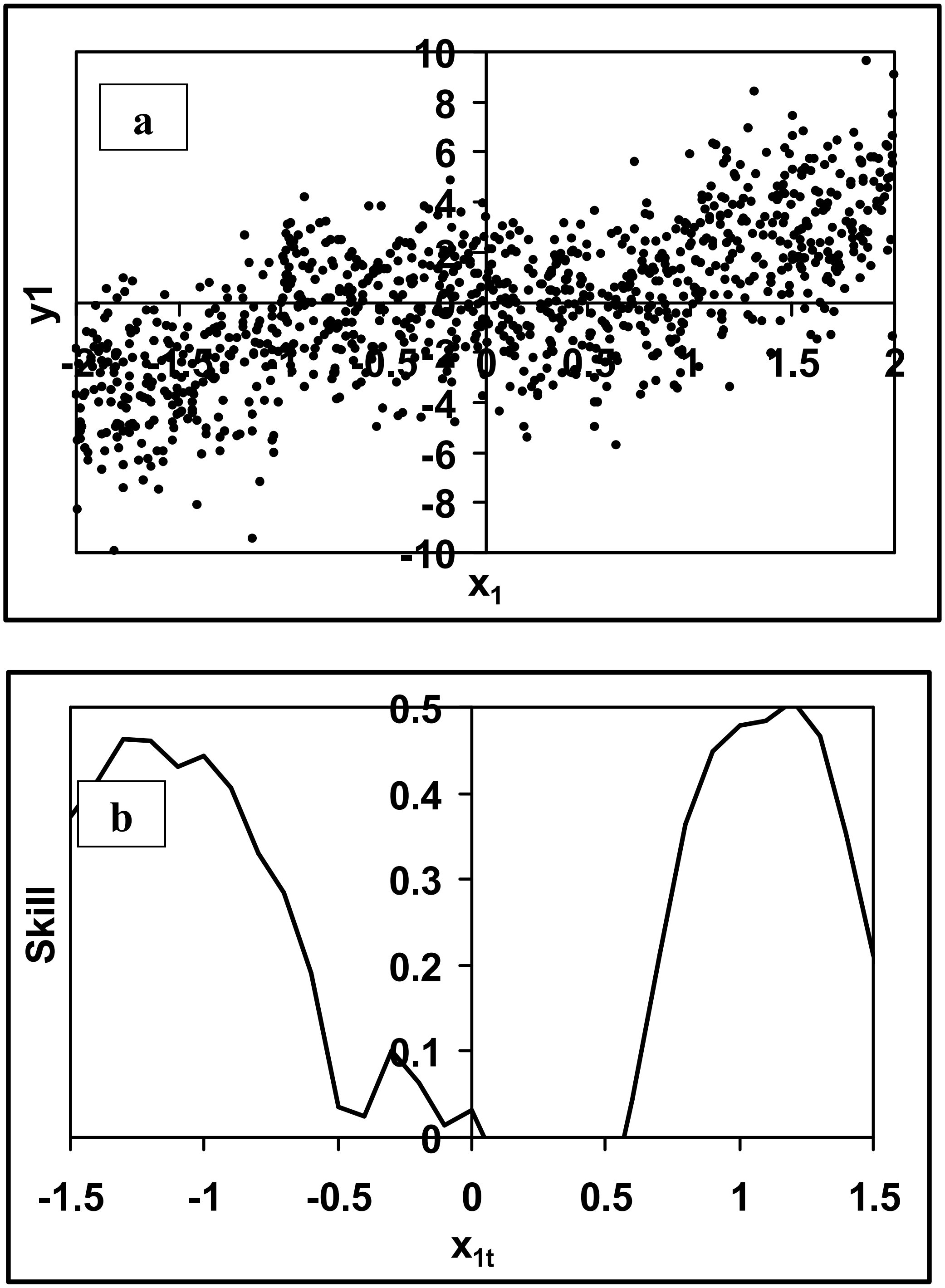


Figure 4.1: Importance of assessing the skill of the model from the predictor space.

4.2 Multi-Model Ensembling based on Predictor State Space – Algorithm Development

Let us suppose that we have streamflow forecasts, $Q_{i,t}^{m}$, where m=1,2..,M denoting the forecasts from 'M' different models, i = 1,2, ..N representing ensembles of the conditional distribution of streamflows with 'N' denoting the total number of ensembles under each model, and 't' denoting the time (season/month) for which the forecast is issued. Assuming that we have a total of t= 1,2,...n years for which the forecasts, $Q_{i,t}^{m}$, are available and the models also have a common predictor vector, $\mathbf{X_t}$, which influences the conditional distribution of hydroclimatic attributes represented using the ensembles. Figure 4.2 provides a flow chart indicating the steps in implementing the proposed multi-model ensembling conditioned on the predictor state. It is important that the proposed approach requires at least one common predictor among the 'M' competing models. Even if the models do not have a common predictor particularly in the context of GCM forecasts, one could use the leading principal component of the underlying boundary conditions (for instance, SSTs) as the common predictor across all the models. As mentioned before, developing multi-model ensembles based on optimal combination method requires the observed climatic/streamflow variables O_t, using which one could assess the skill of the probabilistic forecasts using Rank Probability Score (RPS) [Murphy 1970, Candille and Talagrand 2005, Anderson, 1996] to obtain the weights w_t^m. It is important to note that RPS is evaluated each year using the ensembles (N = 1000) representing the conditional distribution, which is quite different from correlation for which one needs a minimum of two years of forecasts. The Rank Probability Skill Score (RPSS) represents the level of improvement of the RPS in comparison to the reference forecast strategy which is usually assumed to be climatology. Appendix A provides details on obtaining RPS and RPSS for a given probabilistic forecasts.

Let us denote the RPS and RPSS of the probabilistic forecasts, $Q^m_{i,t}$, for each time step as RPS^m_t and $RPSS^m_t$,respectively. Our approach to assess the skill of the model is its ability to predict under similar climatic conditions or the predictor state, which could be identified by choosing a distance metric that computes the distance between the current predictor state, $\mathbf{X_t}$,and the historical predictor vector, $\mathbf{X}$. One could use simple Euclidean distance or a more generalized distance measure such as Mahalonobis distance metric, which is more useful if the predictors' exhibit correlation among them. Compute the distances d_{tl} between the current conditioning state $\mathbf{X_t}$, and the historical predictor vector $\mathbf{X_l}$ as

$$\mathrm{d}_{tl} = \sqrt{(\boldsymbol{X}_t - \boldsymbol{X}_l)^{\boldsymbol{T}} \hat{\Sigma}^{-1} (\boldsymbol{X}_t - \boldsymbol{X}_l)} \qquad \text{... (1)}$$

where $\hat{\Sigma}$ denotes the variance-covariance matrix of the historical predictor vector **X.** One can note that if l=t, the distance metric, d_{tl}, reduces to zero. Using the distance vector **d**, the ordered set of nearest neighbor indices J can be identified**.** Thus, the j^{th} element in the distance vector metric provides the j^{th} closest $\mathbf{X_l}$ to the current state $\mathbf{X_t}$. Using this information, we assess the performance of each model in the predictor state space as

$$\lambda^{\mathrm{m}}_{\mathrm{t,K}} = \frac{1}{K} \sum_{j=1}^{K} RPS^m_{(j)} \qquad \text{... (2)}$$

where $RPS_{(j)}$ denotes the skill of the forecasting model for the year that represents the j^{th} closest condition (obtained from **J**) to the current condition $\mathbf{X_t}$. In other words, $\lambda^{\mathrm{m}}_{\mathrm{t,K}}$ summarizes the average skill of the forecasting model, m, by choosing 'K' years that resemble very similar to the current condition, $\mathbf{X_t}$. Using $\lambda^{\mathrm{m}}_{\mathrm{t,K}}$ obtained for each model at each time step, we obtain the weights for multi-model ensembling so that the models with better performance during a particular climatic conditions needs to be represented with more number of ensembles in comparison to a model with lower predictability under those conditions. It is important to note that RPS is a measure of error in predicting the probabilities and it is evaluated based on the entire ensembles that represent the conditional distribution of streamflows.

$$w_{t,K}^{m} = \frac{1/\lambda_{t,K}^{m}}{\sum_{m=1}^{M} 1/\lambda_{t,K}^{m}} \qquad \text{... (3)}$$

If $\lambda_{t,K}^{m}$ is zero for a subset of models $M_1 \leq M$, then the weights $w_{t,K}^{m}$ are distributed equally between the models for which $\lambda_{t,K}^{m}$ is zero with the rest of models' weights being equal to zero. The multi-model forecasts for each time step could be developed by drawing $w_{t,K}^{m} * N$ ensembles from each model to constitute the multi-model ensembles. Thus, one has to specify the number of neighbors 'K' to implement this approach. It is also important to note that choosing fewer 'K' relates to evaluating the model performance over few years of similar conditions, which does not imply that the forecasts are developed from the predictands and predictors based on the identified similar conditions. In fact, $Q_{i,t}^{m}$ are forecasts developed based on the observed values of the predictor and predictand over a particular training period (for leave-one-out cross validated forecasts, we use 'n-1' years of record as training period; for adaptive forecasts, we use 60 years of observed record from 1928-1987 as the training period).

Thus, we use the weights, $w_{t,K}^{m}$, only to draw the ensembles from $Q_{i,t}^{m}$, which is in fact developed based on the chosen training period in developing the forecasts. The simplest approach for selecting the number of neighbors is to find a fixed 'K' that provides improved predictability using multi-model ensembles over 'n' years of forecasts. We evaluate two different methods in choosing the number of neighbors 'K' to develop multi-model ensembles. The performance of multi-model ensembles is also compared with individual model's predictability using various verification measures such as average RPS, average RPSS, anomaly correlation and root mean square error (RMSE). To apply the same algorithm for k_t that gives the minimum RPS from the multi-model ensembles, compute $RPS_{t,k}^{MM}$ for k=1, 2, n-1 and choose k that corresponds to minimum $RPS_{t,k}^{MM}$. Thus, by computing $RPS_{t,k}^{MM}$ for all the data points, we choose the number of neighbors, k_t that has the minimum $RPS_{t,k}^{MM}$.

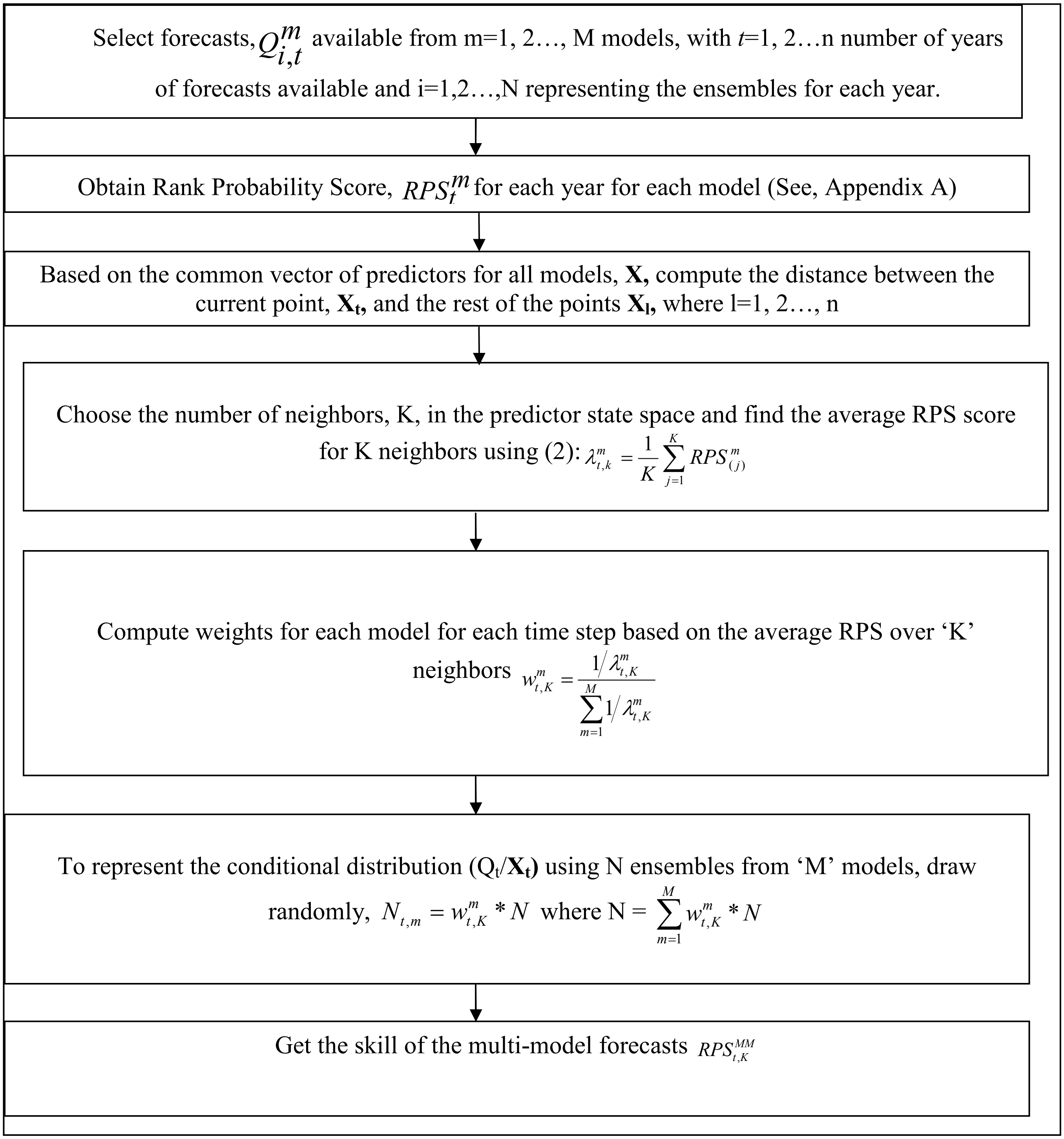


Figure 4.2: Flowchart of the multi-model ensembling algorithm described in section 4.2 for fixed number of neighbors 'K' in evaluating the model skill from the predictor state space.

CHAPTER 5

MULTIMODEL ENSEMBLES OF STREAMFLOW FORECASTS FOR THE FALLS LAKE

In this chapter, we apply the multi-model ensembling algorithm discussed in section 4.2 to combine the forecasts from individual models along with climatological ensembles. The motivation in considering climatology as one of the candidates is upon the presumption that if the observation falls outside the scope of all the models under certain predictor conditions, then climatology should be preferred over individual model forecasts. Recent studies have also shown that a two step procedure of combining first each individual model forecasts separately with climatology and then combining the resulting 'M' combinations at the second step to develop the final, single multi-model ensembles [Robertson et al., 2003; Goddard et al., 2003]. Combining individual models with climatology at one step results with one model getting all the weight (equal to one) leaving the rest of the models' weights to zero [Rajagopalan et al., 2002; Robertson et al., 2004]. We also perform a two step procedure in developing multi-model ensembles by first combining the probabilistic forecasts from resampling model (MM1 in Table 5.1) and regression model (MM2 in Table 5.1) separately with climatology and then the resulting forecasts from two combinations are combined to develop the final multi-model forecasts (MM3 in Table 5.1). Further, we also choose the number of neighbors K in equation (2) by two different methods to identify the relevant predictor conditions: (a) by selecting a fixed 'K' that corresponds to improved multimodel forecasts over the validating period, and (b) varying K_t each year such that the selected 'K_t' corresponds to the minimum RPS that could be obtained from multi-model forecasts. The first method of choosing fixed 'K' is one of the most commonly followed procedure in developing semiparametric and nonparametric models [Sankarasubramanian and Lall, 2003]. By varying K_t, we plan to investigate the role of choosing different K_t in developing multi-model ensembles and their relation to predictor conditions.

Table 5.1: Performance of individual model forecasts and various multi-model schemes under leave-one-out cross validated forecasts and adaptive forecasts for two different strategies of choosing the number of neighbors K (fixed K and varying K_t). All the performance evaluation measures are calculated based on 75 years of data for leave-one-out cross validated forecasts and 15 years for the adaptive forecasts from 1987-2002.

	Leave1-out Cross validated (1928-2002)				Adaptive Forecasts (1988-2002)			
	Correlation	RMSE	RPS	RPSS	Correlation	RMSE	RPS	RPSS
Resampling	0.40	423.03	0.43	-0.03	0.55	482.98	0.43	0.00
Regression	0.35	430.93	0.56	-0.30	0.66	477.82	0.61	-0.07
MM1(k=10)	0.43	422.07	0.42	0.03	0.55	512.83	0.45	0.01
MM1(varying k)	0.42	420.71	0.37	0.15	0.54	506.45	0.40	0.03
MM2 (k=10)	0.31	439.63	0.43	0.03	0.65	523.89	0.48	0.00
MM2 (varying k)	0.36	432.60	0.36	0.21	0.66	516.13	0.43	0.02
MM3 (k=10)	0.44	425.44	0.41	0.06	0.63	511.26	0.45	0.01
MM3 (varying k)	0.43	422.78	0.34	0.23	0.61	510.52	0.40	0.03

MM1 - Resampling+Climatology MM2 - Regression+Climatology MM3 - MM1+MM2

5.1 Skill of Individual Models from Predictor state space

The primary motivation in the proposed approach for multi-model ensembling is to evaluate competing models' predictability in the neighborhood of the predictor state and give appropriate weights based on equation (3) for all the models to develop multi-model ensembles. By analyzing the predictability of two candidate streamflow forecasts shown in figure 5.1, we show the predictability of both models every year using two performance measures, correlation and average RPS (using equation 2), which are computed by choosing a fixed K = 10 based on the dominant predictor PC1. From figure 5.1a, one may prefer to choose forecasts from resampling model instead of forecasts from parametric regression particularly when the dominant principal component, PC1, is less than -2, since the predictive ability of regression model is negative during those conditions. This is seen in Figure 5.1b with the RPS of resampling being lesser than that of RPS of regression. Figures 5.1c and 5.1d show the relative performance of both models against each other. From figure 5.1c, we can see that one would prefer climatological ensembles particularly when correlations estimated from the neighborhood on both models are negative. From 5.1d, we can also identify conditions during which the RPS of regression model being higher than that of RPS of resampling. RPS is computed from the leave-one-out cross validated forecasts given in Table 5.1 for both candidate models and by assuming K=10 in equation (2). Correlation is computed between the observed streamflows and ensemble mean of the leave-one-out cross validated forecasts by considering 10 neighbors from the current state. Note the consistent poor performance of both the models in Figure 5c as well as for high negative values of PC1. Thus, the multi-model ensembling algorithm in section 4.2 identifies these conditions based on RPS using equation (2) and develops a general procedure for multi-model ensembling.

The next two sections show that the performance of multi-model forecasts based on two different strategies of choosing the number of neighbors.

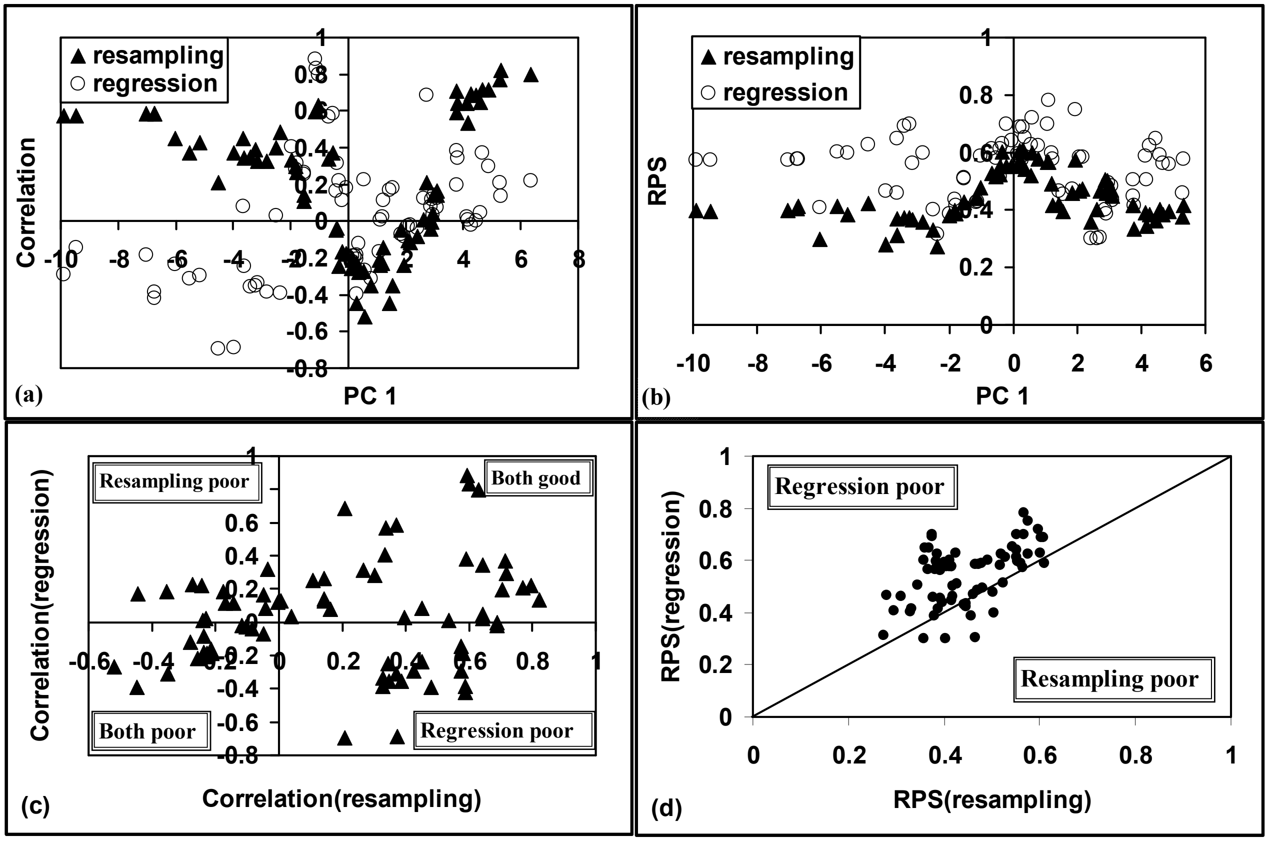


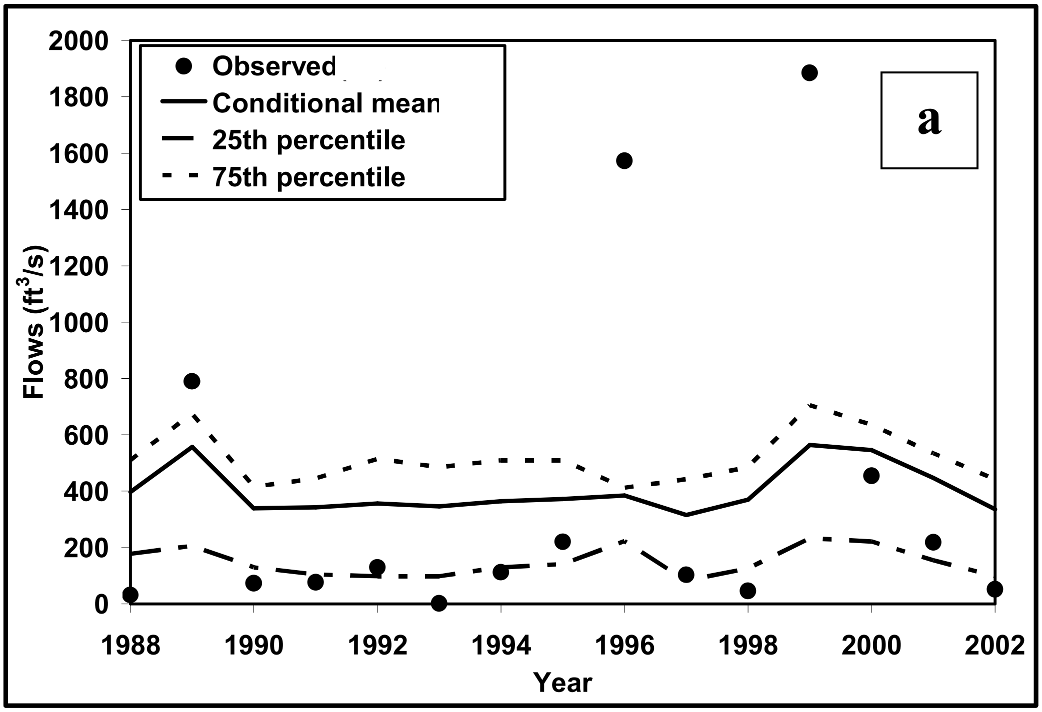
Figure 5.1: Performance of individual models from the predictor state space by considering K=10 neighbors: (5a) Correlation Vs PC1; (5b) RPS Vs PC1; (5c) Correlation of regression Vs Correlation of resampling; (5d) RPS of regression Vs RPS of resampling.

5.2 Performance of Multi-Model Forecasts

As mentioned earlier, the multi-model combination is carried out in two steps: first combining individual model ensembles with climatological ensembles and then the resulting probabilistic forecasts from two combinations will be combined to develop final multi-model ensembles. To generate ensembles that represent climatology, we simply bootstrap the observed streamflows into Falls Lake assuming each year has equal probability of occurrence, which is a reasonable assumption given that there is no year to year correlation between the time series of summer flows. Figure 5.2a gives the multi-model adaptive forecasts by choosing a fixed K=10 in equation (2) for identifying similar conditions in the predictor state space. The fixed number of neighbors K=10 is chosen since it provided the lowest average RPS from multi-model ensembles for the period over which the forecast is developed. For leave-one-out cross validated forecasts, the average RPS is computed from 74 years of forecasts; For adaptive forecasts, average RPS is computed from 15 years of forecasts from 1988-2002. Thus, we chose the fixed 'K' by plotting the average RPS obtained for each neighbor from K = 1 to the maximum of the data length used for model fitting (for leave-one-out cross validated forecasts, it is 74 years; for adaptive forecasts, the maximum K = 60 years of record from 1928-1987) and choosing the value of 'K' that produced the lowest average RPS. Figure 5.2b provides adaptive forecasts developed from multi-model ensembles by choosing a varying K_t each year such that the chosen K_t for that year corresponds to the minimum RPS that could be obtained from multi-model ensembles. By choosing K using any of the above strategy (fixed K or varying K_t), we assess the skill of individual models over 'K' neighbors using equation (2) and obtain weights for each model using equation (3). Based on the weights, we draw proportionately equivalent number of ensembles from each model to constitute multi-model ensembles. The constituted multi-model ensembles in Figures 5.2a and 5.2b have N=1000 ensembles which has been developed through a two step procedure of first combining individual models with climatology and then obtaining the final multi-model from the resulting combination of individual models with climatology. Table 5.1 provides the comparison between individual models and multi-model ensembles using various performance evaluation measures for both leave-one-out cross validated forecasts and adaptive forecasts for fixed K=10 and varying K_t

each year. Under varying K_t, the algorithm in section (4.2) is applied for $1 \leq K \leq 60$ $_t$ K (60 years of training data from 1928-1987) and K_t that corresponds to minimum RPS of the multi-model ensembles is chosen for each year. Both figure 5.2 and Table 5.1 show very clearly that both strategies of choosing the number of neighbors result in significant improvements in predictability from multi-model ensembles compared to the probabilistic forecasts from individual models. It is important to note that the improved performance of multi-model ensembles is seen in almost all evaluation measures.

Even with fixed number of neighbors, the multi-model ensembling algorithm based on predictor state space provides improved predictability than the individual model forecasts. Ideally, one would like to have the number of neighbors varying each year so that the chosen K_t relates to the conditioning predictor state. For instance, under very high values of |PC1|, very few years could be chosen as similar to the conditioning state. To understand whether we see any relationship between the chosen K_t for every year that corresponds to the minimum RPS of the multi-model ensembles for that year, we plot the varying K_t with PC1 in figure 5.3. From figure 5.3, we see in general that smaller number of neighbors is chosen particularly if the PC1 corresponds to above normal or below normal values. Figure 5.3 also shows the distance between the conditioning state, $PC1_t$ and the chosen K_t in the predictor PC1 space. It is important to note that PC1 primarily denotes ENSO conditions (correlation between PC1 and Nino3.4 = 0.36), thus positive (negative) PC1 denotes the El Nino (La Nino) conditions. Further, under varying K_t strategy, we may consider K_t =1 in evaluating the skill of the model in predictor state space using equation (2), which does not imply that the multi-model forecasts is developed from that identified similar conditions' observed



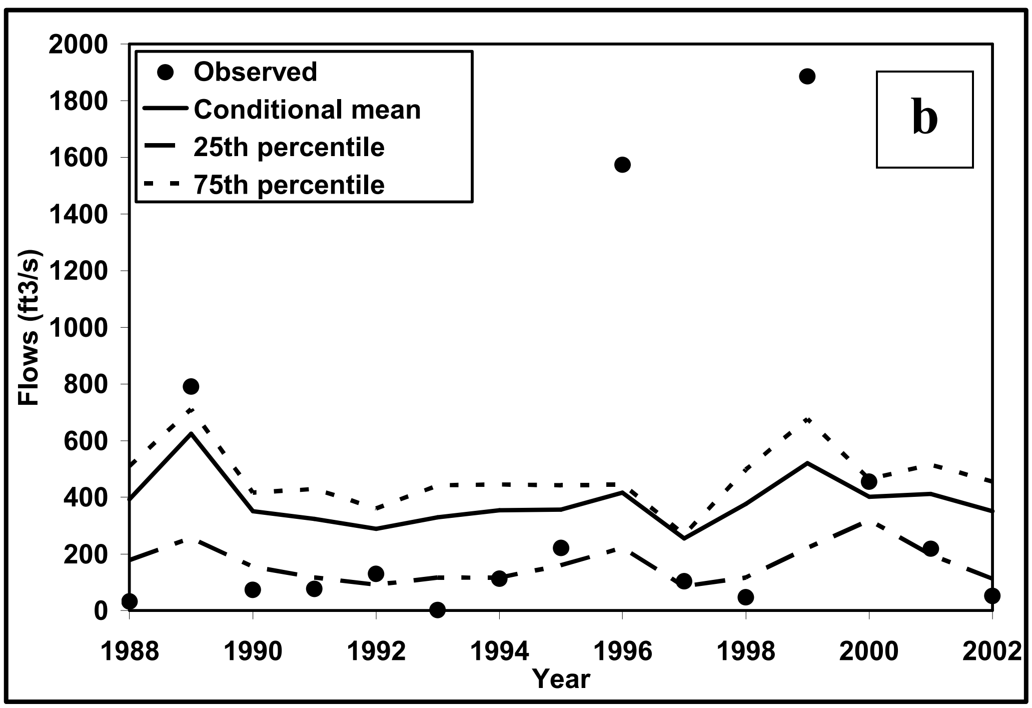


Figure 5.2: Performance of multi-model forecasts developed using the algorithm in (4.2). (5.2a) Fixed K=10 (5.2b) Varying K_t.

predictand and predictors alone. Instead, we identify similar conditions only to evaluate the performance of the individual models so that smaller weights to the model that has higher RPS under those conditions. Thus, the weights obtained by assessing the model skill in the predictors' state space are only used to proportionately draw ensembles from candidate model's probabilistic forecasts which have been actually developed based on the training data used for model fitting.

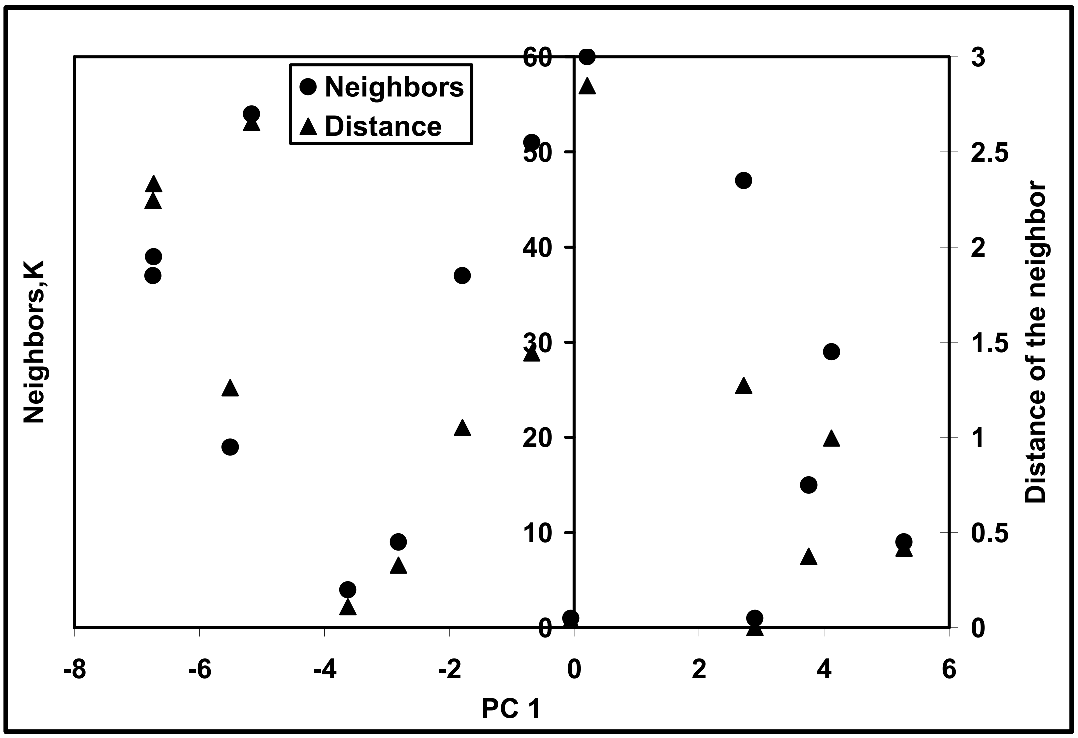


Figure 5.3: Relationship in choosing different neighbors (varying K) according to the predictor conditions.

5.3 Role of Multi-Model Forecasts in improving the forecast reliability

Figures 5.4a and 5.4b compare the reliability of multi-model forecasts with the reliability of individual model forecasts for below normal and above normal categories of forecasts, respectively. Reliability diagrams provide information on the correspondence between the forecasted probabilities for a particular category (above-normal, normal and below-normal) and how often (frequency) that category is being observed under that forecasted probability. For instance, if we forecast the occurrence of below-normal category as 0.9 over n1 years (n1 $\leq$ n), then over the long-term (n years) we expect the actual outcome to fall under below-normal category for 0.9*n1 times. To construct figure 5.4, we utilized leave-one-out cross validated forecasts and divided the forecasted probability for each category into percentiles. Figures 5.4a and 5.4b also show the diagonal perfect reliability line with one to one correspondence between forecasted probability and its observed relative frequency. Figures 5.4a and 5.4b also provide the sum of absolute deviation from the perfect reliability line for regression, resampling and multi-model ensembles. From both figures, we can clearly see that there is a better correspondence between perfect reliability line and the multi-model forecasts with the sum of absolute deviation from the perfect line is small for multimodel forecasts. Of the three forecasts, regression seems to have poor reliability because it employs a parametric log-normal model for estimating the conditional distribution using conditional mean and variance.

Resampling, being a data driven approach without prescribing any functional form, estimates the conditional distribution fairly well and it corresponds better to the perfect reliability line. However, multi-model ensembles have lesser error with the sum of absolute deviation from the perfect reliability line being smaller, more noticeably for above-normal category. Previous studies have also shown that the main advantage of using multi-model ensemble forecasts is in improving the reliability of forecasts [Goddard et al., 2003; Barnston et al., 2003]. Thus, our approach of multi-model ensembling not only improves the aggregate performance measures shown in Table 5.1, but also provides better correspondence between forecasted probabilities and its relative frequency of occurrence under a particular forecast category.

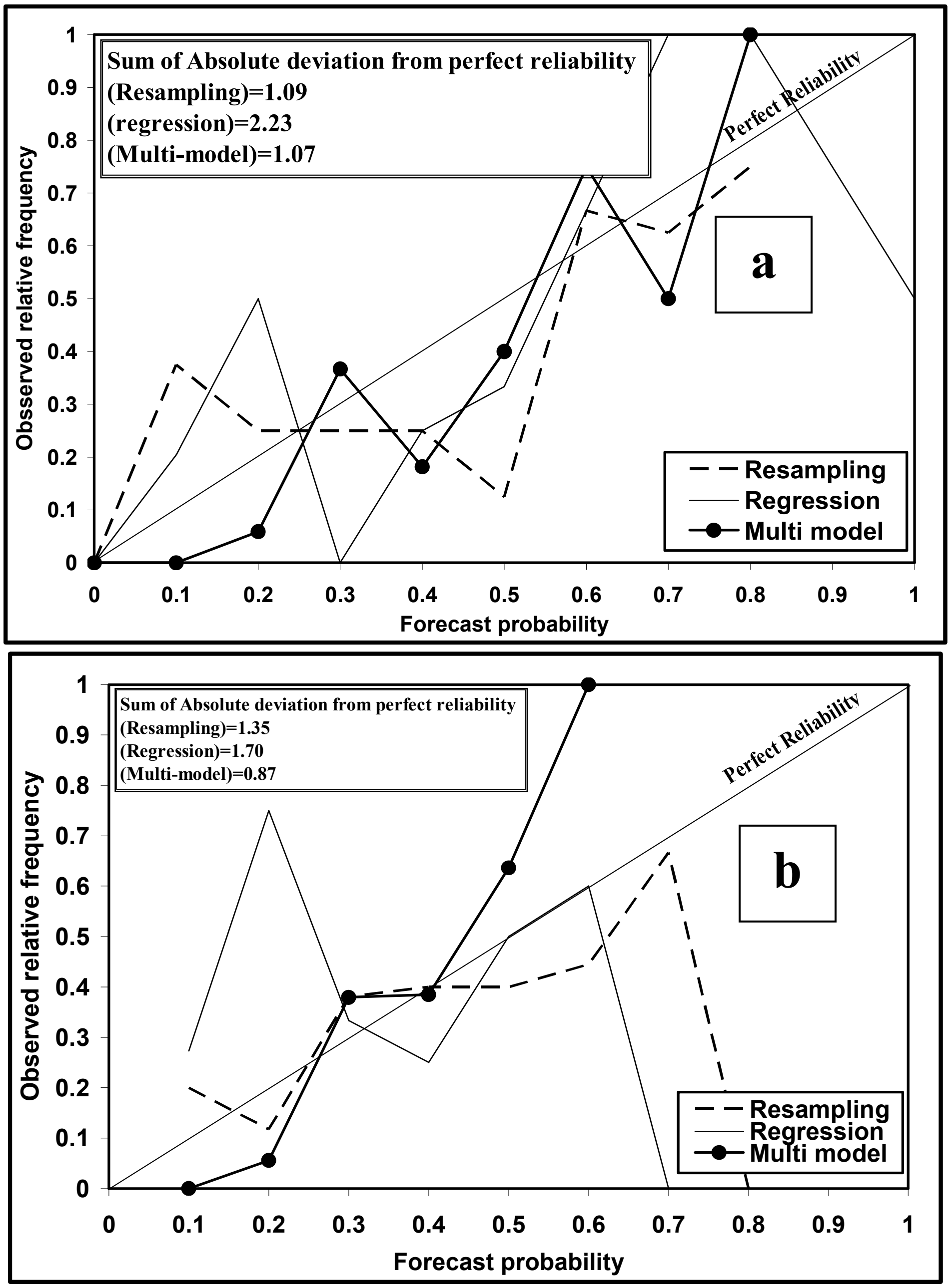


Figure 5.4 Comparison of reliability of leave-one-out retrospective cross validated forecasts from regression, resampling and multimodel forecasts for below-normal (Figures 5.4a) and above-normal (5.4b). Figure also shows the perfect reliability line along with the sum of absolute deviation from the perfect reliability line for each model. Note the sum of absolute deviation from perfect reliability line is smallest for multi-model ensembles under both above-normal and below-normal categories.

CHAPTER 6

MULTI-MODEL CLIMATE FORECASTS FOR THE UNITED STATES

Improved monitoring of SSTs, particularly in the tropical oceans, has led to significant interest using forecasted SSTs to force GCMs to develop operational climate forecasts. The GCMs predict various states and fluxes of the atmosphere over the globe, such as precipitation, temperature geopotential height, surface pressure, winds, moisture based on the given initial atmospheric states and boundary SST conditions. Thus developing climate forecasts is a two tiered process: (a) forecast the SSTs, and (b) force it with GCMs to develop ensembles of climate forecasts over the globe. Model uncertainty along with errors in representing initial and boundary conditions result in errors in model predictions. Efforts to reduce model uncertainty have been addressed primarily by multi-model combination of GCM outputs.

In this section, we apply the proposed algorithm in section 4.2 to improve winter (December-January-February) precipitation forecasts in the US by combining precipitation from different GCMs. The developed multi-model precipitation forecasts were also statistically downscaled to develop forecasts of streamflow into the Falls Lake.

6.1 General Circulation Models and their role in climate prediction

GCMs are commonly employed by various national/international research institutions for developing long lead seasonal climate forecasts. As mentioned earlier, climate forecasts are developed from a two tiered process in which the forecasted SSTs are forced with GCM initial conditions to develop climate forecasts.

The GCM outputs are typically available at large spatial scale (2.5° X 2.5°). The outputs of the GCMs can be used as boundary conditions for regional models whose outputs are obtained at watershed scale (60Km by 60 Km). To obtain streamflow forecasts, one could use the dynamically downscaled precipitation and temperature forecasts with a hydrological model to develop streamflow forecasts.

Since SST anomalies, particularly tropical SSTs are known to be a fundamental driver of atmospheric climate anomalies, forecasting the SST anomalies for the target season is the first step in the climate forecasting task. At International Research Institute for Climate and Society (IRI), climate forecasts are developed by forcing multiple GCMs with three different SST forecasts, i.e. the NCEP coupled model, the Lamont-Doherty Earth Observatory (LDEO) simple coupled model and the NCEP/Climate Prediction Center's constructed analogue statistical model [Goddard et al., 2003]. Prediction of SSTs for Indian Ocean and Atlantic Ocean are carried out using statistical techniques. The GCMs are then forced with the forecasted SSTs for the target period under different atmospheric initial conditions to develop ensembles of forecast. The ensembles provide an idea of the probability distribution of outcomes, as well as the mean outcome which is regarded as a best guess for the forecast. Beginning October 1997, IRI has been issuing global climate forecasts and this process is getting refined due to the continuous development of innovative approaches such as multi-model combination of GCM forecasts [Goddard et al., 2003; Barnston et al., 2003].

6.2 Source of uncertainty in model prediction

Error resulting from climate forecasts is primarily of two types, uncertainty in initial and boundary conditions and model error [Hagedorn et al., 2005]. The first source of error is typically resolved by representing the uncertainties in initial (atmospheric states) and boundary (SST forecasts) conditions in the form of ensembles. The second source of error is inevitable with a particular model, since the model error occurs even if the forecasts are obtained from observed initial and boundary conditions (perfect forcings). A common approach to reduce model uncertainty is through refinement of parameterizations and process representations in the considered model which could be either GCMs or Regional Climate Models (RCMs) or hydrologic models. Given that developing and running GCMs is time consuming, recent efforts have focused in reducing the model error by combining multiple GCMs to issue operational climate forecasts [Rajagopalan et al., 2002; Robertson et al., 2004; Barnston et al., 2003; Doblas-Reyes et al., 2000; Krishnamurthi et al., 1999].

The objectives of this study are to apply the developed algorithm in section 4.2 for combining forecasts from multiple GCMs, available at large spatial scales, by assessing the model's predictability conditioned on the predictor state. Recent studies on improving seasonal climate forecasts using optimal multi-model combination techniques basically assign weights for a particular model based on its ability to predict the climatic variable over the entire period for which the GCM simulations are available [Rajagopalan et al., 2002; Robertson et al., 2004; Barnston et al., 2003]. Given that each model's predictability could also vary depending on the state of the predictor (SSTs for GCMs), it would be appropriate to apply the proposed multi-model ensembling methodology in section 4.2 that assigns weights to each model by assessing the skill of the models based on the predictor state space.

6.3 Multi-Model Ensembles of GCMs : Motivation

The multi-model ensembling method proposed here is motivated by the fact that the skill of the GCM forecasts depends on predictor conditions. Studies focusing on the skill of GCMs show that the overall predictability of GCMs is enhanced during ENSO years over North America [Brankovic and Palmer 2000; Shukla et al., 2000; Quan et al., 2006]. Recent research shows that performance of seasonal forecasts predicted by GCMs depends predominantly on the state of ENSO and local SST conditions [Quan et al., 2006; Giannini et al., 2004]. Similarly, studies have also shown the importance of various oscillations or climatic conditions in influencing the predictability of GCMs over various part of the globe. For instance, Giannini et al., [2004] show that tropical Atlantic variability (TAV) plays as a preconditioning state in the development of ENSO related teleconnection in determining GCM's ability to predict rainfall over North East Brazil, which is a region shown to have significant skill in seasonal climate prediction. Several studies show that the predictive ability of GCMs is dependent highly on ENSO conditions [Brankovic and Palmer 2000; Shukla et al., 2000; Quan et al., 2006].

To understand the performance of GCMs under various predictor states, the study considered the performance of two GCMs under various ENSO states. Figure 6.1 shows the skill of two GCMs for El Nino, La Nina and over the long-term over the US. It gives the correlation between observed precipitation and model forecasted precipitation from two different models. The models taken are ECHAM4.5 by Max Planck Institute and CCM3V6 by NCAR. We can clearly see from the figure that skill of the model depends on SST conditions (ENSO conditions), i.e., there is some signal during ENSO years for the models predictability. Correlations that are significant are only shown (>0.46 for ENSO years and 0.26 for the entire years category). It can also be seen that there is a significant difference in predictability between both the models across space. Hence combining the models based on predictor conditions is a better strategy than combining them based on long term predictability.

Figure 6.1 emphasizes that for post-processing of individual model's climate forecasts to develop multi-model ensembles, one needs to assess the skill of the individual model ensembles based on the predictor state. By considering climatology as one of the candidate forecasts, we develop a multi-model ensembling scheme that formally assesses and compares the skill of the competing models under a given predictor conditions so that lower weights are assigned for a model that has poor predictability under such conditions. The next section describes the GCM precipitation forecasts that are considered in developing the multi-model precipitation forecasts for the entire US.

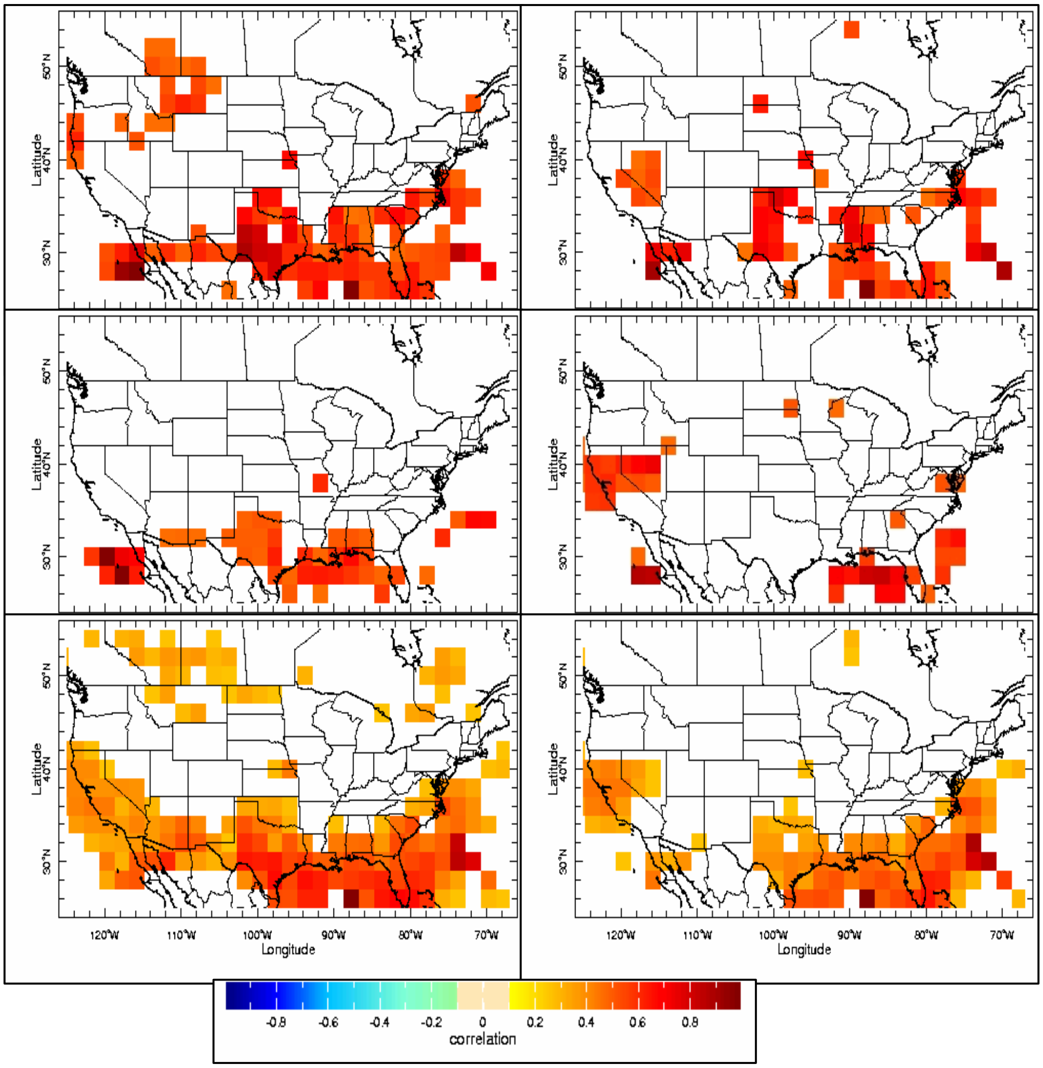


Figure 6.1: Predictability of two GCMs over the US for different climatic states. Panel on the left (right) is for ECHAM4.5 by Max Planck Institute (CCM3V6 by NCAR). Top (Middle) row is the correlation between the observed precipitation and model predicted precipitation during El Nino (La Nina) years. Bottom row is the correlation between observed precipitation and model predicted precipitation for all the years. Correlations that are significant are only shown (>0.46 for ENSO years and 0.26 for the entire years category). Note the significant difference in predictability between both models across space and time.

6.4 Multimodel winter precipitation forecasts for the US

Since operational climate forecasts are available only from October 1997, this study combines models based on the historical simulations with each GCM forced with observed SSTs. Historical simulations of three GCMs are combined to develop multimodel ensembles. The models considered are (a) CCM3 version 6 (developed by NCAR); (b) ECHAM4.5 (developed by Max Plank Institute); and (c) COLA (Center for Ocean-Land-Atmosphere Studies) with all having monthly historical simulations for the period 1950-1996. Historical simulations could be accessed from http://iridl.ldeo.columbia.edu/. Since the chosen individual models differ in the number of ensembles (85 for ECHAM 4.5, 24 for CCM version 6 and 10 for COLA), the ensembles of COLA and CCM version 6 models are increased to 85. Using the proposed multi-modeling algorithm in section 4.2, multimodel ensembles of precipitation forecasts from the above mentioned three combined with climatology to obtain improved climate forecasts. By considering climatology as one of the forecasts, the method ensure that if the skill of all models is poor under certain predictor conditions, climatological ensembles will obviously constitute the most of the multimodel ensembles. Since the method requires observed precipitation to assess the skill of each model, the monthly observed precipitation at 0.5×0.5 grid from University of East Anglia is employed (http://iridl.ldeo.columbia.edu/SOURCES/.UEA/.CRU/.Global/.prcp/). The predictors $\mathbf{X_t}$ are obtained by considering the principal components of global SSTs (http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.v2/).

The multi-model ensembling algorithm described in section 4.2 is applied to combine precipitation forecasts from three GCMs along with the climatological ensembles. To develop multimodel ensembles, we follow the two step procedure of combining individual model ensembles with climatological ensembles and then the resulting ensembles from this step (model+climatology) are combined further to develop one single multimodel ensembles. Previous studies have shown that such a two-step procedure improves the skill of multimodel ensembles [Goddard et al., 2003; Robertson et al., 2004]. Climatological ensembles are developed by just bootstrapping the observed precipitation at the grid point. By identifying similar conditions in relation to the current predictor condition, PC_t, we choose the number of neighbors, K_t, that correspond to the minimum RPS from the multimodel ensembles. Weights, W_t^m ,for each model was obtained using equation (3) corresponding to the identified K_t. These weights, W_t^m, are used to draw N. W_t (N=85) ensembles from each model. Thus if the skill of these models were poor then climatological ensembles will constitute most of the multimodel ensembles. Since climate forecasts are represented in the form of ensembles, the skill of the models are evaluated using RPS and RPSS. A detailed description on computing RPS and RPSS from tercile forecasts is given in Appendix A.

6.5 Results and Analysis

Average RPSS of multimodel forecasts and the individual model forecasts were calculated each year. Performance of multimodel forecasts was compared with individual model forecasting by computing average RPSS over the entire period of verification (1950-1996). Thus, Winter seasonal (DJF) averaged precipitation forecast data is obtained from the mentioned three models for the period 1950-1996. The observed precipitation data from UEA site is used to compare and develop multimodel forecasts. The algorithm described in section 4.2 is employed at each 2.5 degree by 2.5 degree grid point over the US to develop multi-model ensembles of precipitation forecasts. To compare the skill of multi-model forecasts with individual model forecasts over the US, we show the average RPSS maps for each model and multimodel forecasts. Figure 6.2 provides the average RPSS for individual models and the developed multi-model over the US. The figure clearly shows that the multi model forecasts have a higher average RPSS in comparison to the individual model forecasts. In most parts of the region, the multi model forecasts have improved predictability of the individual model forecasts substantially. Average RPSS of the individual models is negative in many regions indicating that the skill of GCMs is poorer than climatology. By combining these poor models with climatology, we improve the resulting multimodel forecasts by analyzing the individual model's predictability conditioned on the predictor state, PC_t.

In some pockets over North Eastern US, we see higher predictability by CCM3v6 GCM in comparison to multimodel forecasts. This could be primarily because of poor relations between the predictor state, (1st EOF of global SSTs) and the precipitation in that region. To understand this, we plot RPSS of CCM3v6 with the correlation between PC1 and observed precipitation. From figure 6.3, we see a slight increase in RPSS from CCMv6 on the regions, for the North Eastern US. This needs to be further investigated by identifying relavent predictors that influence the hydroclimatic potential of the region.

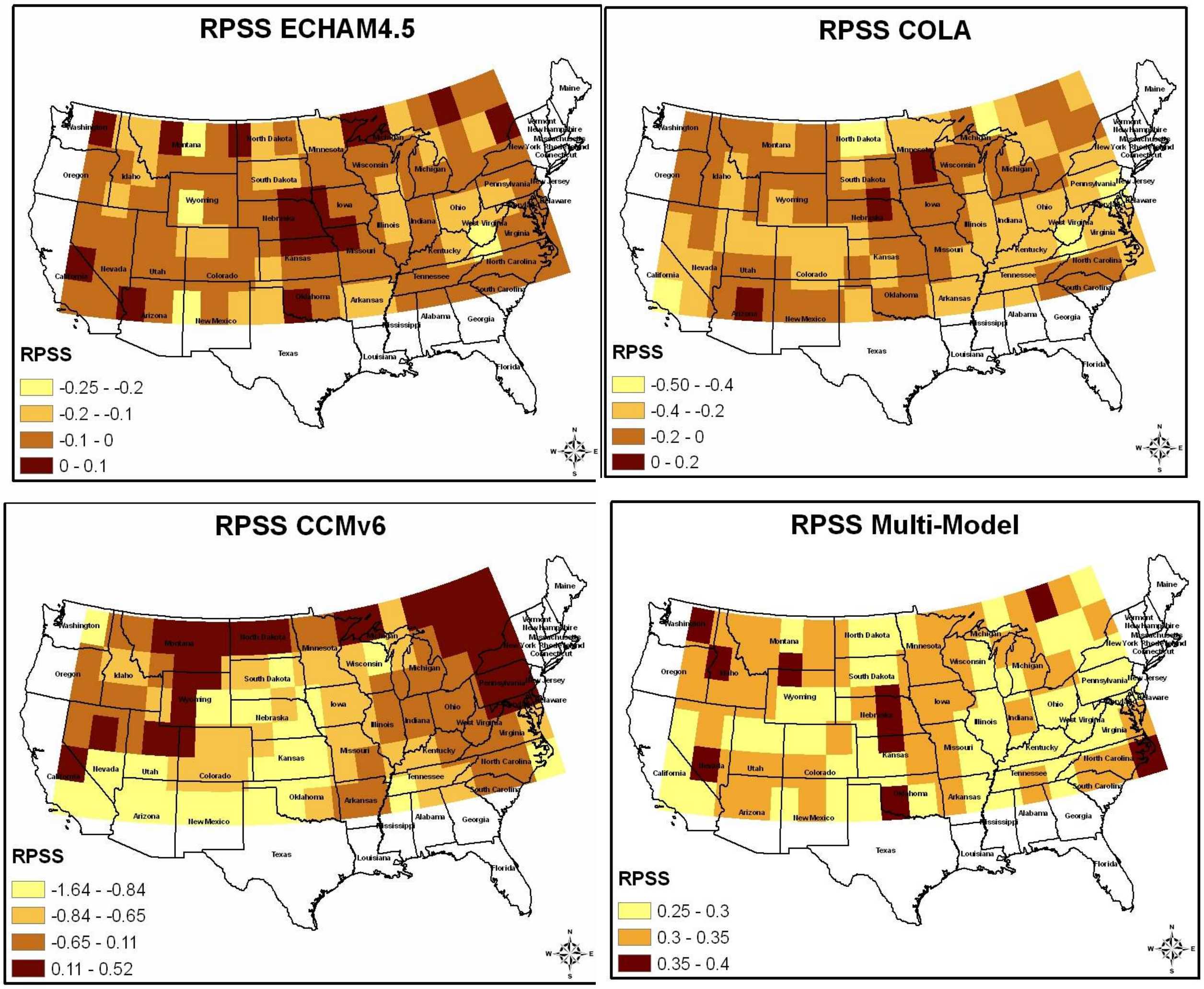


Figure 6.2 Skill of individual models and the combined multi-model

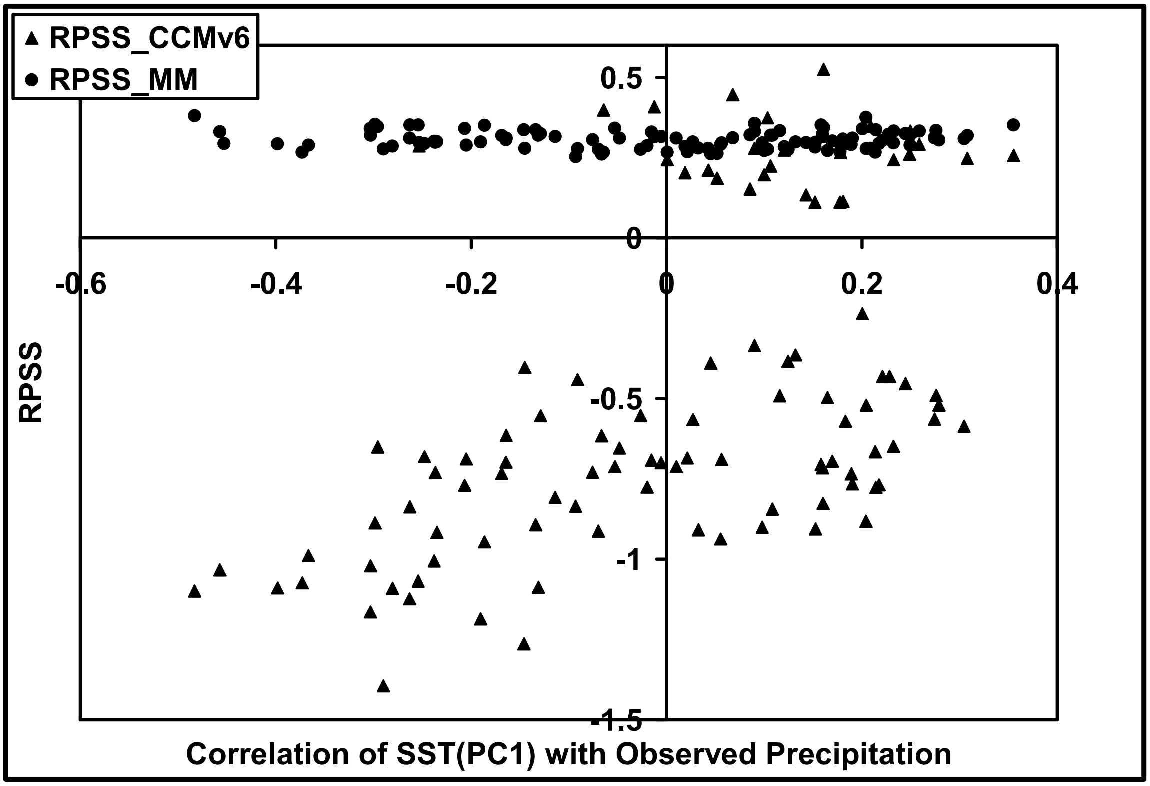


Figure 6.3 Performance of CCMv6 and multi-model forecasts under the causal relation of predictor SST with predictand observed precipitation.

CHAPTER 7

SUMMARY AND CONCLUSIONS

A new methodology for developing multi-model ensembles is presented and demonstrated that combines probabilistic streamflow forecasts from two low dimensional statistical seasonal streamflow forecasting models. The developed approach obtains multi-model ensembles by assessing the skill of the candidate forecasting models conditioned on the state of the predictor. To evaluate the model performance based on the state of the predictor, the multi-model ensembling algorithm employs Mahalonobis distance measure that computes the distance between the current state and the historical predictor vector by considering the covariance between the predictors. By choosing 'K' neighbors based on the distance metric, we assess the performance of the model by computing average RPS, in the neighborhood of the predictor state. The average RPS for each model are then converted into weights, using which appropriate number of ensembles are drawn from each candidate models to develop multi-model ensembles.

7.1 Streamflow forecasts for the Falls Lake

The proposed algorithm described in section 4.2 was employed to combine two low dimensional statistical models to develop multi-model ensembles of JAS streamflow forecasts for the Falls Lake of the Neuse river basin, NC. By comparing the performance of multi-model ensembles with individual model performance using various performance evaluation measures as well as using reliability diagrams, we show that the proposed multi-model ensembling approach develops probabilistic streamflow forecasts with much better predictability than what could be obtained from single model ensembles. We adopt a two step procedure by combining individual models first with climatology and then the resulting combinations are finally combined using the algorithm in section 4.2 to develop the final multimodel ensembles. This has been shown to improve the performance of multi-model ensembles as well as to ensure better stability of weights obtained for multi-model combination [Robertson et al., 2004]. Our approach also support these findings further by first eliminating the poorly performing model under a particular predictor conditions with climatological ensembles and then goes to the next step of combining the resulting forecasts into a final product of multimodel ensembles.

As shown in figure 5.1, if the predictability of all the models is poor under a particular condition, then our approach will eventually replace the multi-model ensembles with only climatological ensembles. This will help to reduce false alarms and missed targets in the issued forecasts and improves the reliability by ensuring better correspondence between the forecasted probability and its observed relative frequency. Further, the approach may use very small number of neighbors in assessing the predictability of model in the predictor state space, but the average skill of the model in those conditions are only used to arrive at the weights so that appropriate number of ensembles could be drawn from candidate model's streamflow forecasts which are in fact obtained based on the observed predictors and predictand employed for model fitting. By employing RPS to assess the skill of the probabilistic forecasts each year, the approach naturally considers predictability of entire conditional distribution of streamflows/precipitation.

7.2 Precipitation forecasts for the US

The methodology is also demonstrated for developing multimodel ensembles from three different precipitation forecasting GCMs. The study employs the proposed algorithm described in section 4.2 for combining multiple GCMs to develop multimodel climate forecasts for the US. As described above we employed a two step procedure by combining individual models first with climatology and then the resulting combinations are finally combined using the algorithm in section 4.2 to develop the final multimodel ensembles. The approach systematically eliminates the poorly performing models under a particular predictor conditions in the first step and then goes to the next step of combining the resulting forecasts into a final product of multimodel ensembles. To compare the skill of multi-model forecasts with individual model forecasts over the US we show the average RPSS maps for each model and multimodel forecasts. Preliminary results show that the skill of GCMs is poorer than climatology in many regions. By combining these poor models with climatology, we improved the resulting multimodel forecasts by analyzing the individual model's predictability conditioned on the predictor state.

7.3 Future work

The proposed multimodel ensembling scheme is general and applicable to most of the environmental and geosciences models. For example this work could be extended to combine multiple hydrologic models to improve streamflow prediction. Multiple hydrologic models can be combined based on their ability to predict the observed streamflow at various conditioning states of the two common predictors, precipitation and temperature. One could also extend this work with a Bayesian hierarchical model. Given that Bayesian hierarchical modeling facilitates multi-level modeling, we could extend the proposed multi-modeling scheme to take into account variability in forecasting skill that occur primarily due to variability in location, time and state of the predictor. By bringing the state of the art statistical methodologies on Bayesian model averaging for improving seasonal climate forecasts, we can improve seasonal to interannual climate forecasts, which is an important problem in geosciences community.

Our future work will focus on looking at the spatial and temporal organized modes exhibited by climate forecasts and to employ a Bayesian hierarchical framework to develop multi-model ensembles of climate forecasts.

7.4 Publications

Multi-model Ensembling of Probabilistic Streamflow Forecasts: Role of Predictor State Space in Skill Evaluation. Naresh Devineni, A.Sankarasubramanian, Sujit Gosh.2006 *Water Resources Research* (under 1st revision).

Multi-model Ensembling of Probabilistic Streamflow Forecasts: Role of Predictor State Space in Skill Evaluation. A.Sankarasubramanian, Naresh Devineni, Sujit Gosh.2006 Technical report, Institute of Statistics Mimeo Series 2595. North Carolina State University.

References

Anderson JL. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. Journal of Climate 1996;9(7):1518-1530.

Barnston AG, Mason SJ, Goddard L, DeWitt DG, Zebiak SE. Multimodel ensembling in seasonal climate forecasting at IRI. Bulletin of the American Meteorological Society 2003;84(12):1783-.

Boyle DP, Gupta HV, Sorooshian S. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. Water Resources Research 2000;36(12):3663-3674.

Brankovic C, Palmer TN. Seasonal skill and predictability of ECMWF PROVOST ensembles. Quarterly Journal of the Royal Meteorological Society 2000;126(567):2035-2067.

Candille G, Talagrand O. Evaluation of probabilistic prediction systems for a scalar variable. Quarterly Journal of the Royal Meteorological Society 2005;131(609):2131-2150.

Carpenter TM, Georgakakos KP. Assessment of Folsom lake response to historical and potential future climate scenarios: 1. Forecasting. Journal of Hydrology 2001;249(1-4):148-175.

Cayan DR, Redmond KT, Riddle LG. ENSO and hydrologic extremes in the western United States. Journal of Climate 1999;12(9):2881-2893.

Dettinger MD, Diaz HF. Global characteristics of stream flow seasonality and variability. Journal of Hydrometeorology 2000;1(4):289-310.

Dillon W and M. Goldstein. Multivariate Analysis: Methods and Applications: Wiley & Sons 1984.

Doblas-Reyes FJ, Deque M, Piedelievre JP. Multi-model spread and probabilistic seasonal forecasts in PROVOST. Quarterly Journal of the Royal Meteorological Society 2000;126(567):2069-2087.

Doblas-Reyes FJ, Hagedorn R, Palmer TN. The rationale behind the success of multi-model ensembles in seasonal forecasting - II. Calibration and combination. Tellus Series a-Dynamic Meteorology and Oceanography 2005;57(3):234-252.

Enfield DB. El-Nino, Past and Present. Reviews of Geophysics 1989;27(1):159-187.

Gangopadhyay S, Clark M, Rajagopalan B. Statistical downscaling using K-nearest neighbors. Water Resources Research 2005;41(2).

Georgakakos KP. Probabilistic climate-model diagnostics for hydrologic and water resources impact studies. Journal of Hydrometeorology 2003;4(1):92-105.

Giannini A, Saravanan R, Chang P. The preconditioning role of Tropical Atlantic Variability in the development of the ENSO teleconnection: implications for the prediction of Nordeste rainfall. Climate Dynamics 2004;22(8):839-855.

Goddard L, Barnston AG, Mason SJ. Evaluation of the IRI's "net assessment" seasonal climate forecasts 1997-2001. Bulletin of the American Meteorological Society 2003;84(12):1761-.

Guetter AK, Georgakakos KP. Are the El Nino and La Nina predictors of the Iowa River seasonal flow? Journal of Applied Meteorology 1996;35(5):690-705.

Hagedorn R, Doblas-Reyes FJ, Palmer TN. The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept. Tellus Series a-Dynamic Meteorology and Oceanography 2005;57(3):219-233.

Hamlet AF, Lettenmaier DP. Columbia River streamflow forecasting based on ENSO and PDO climate signals. Journal of Water Resources Planning and Management-Asce 1999;125(6):333-341.

Hansen JW, Hodges AW, Jones JW. ENSO influences on agriculture in the southeastern United States. Journal of Climate 1998;11(3):404-411.

Hayes MJ, Svoboda, M.D., Knutson, C.L., and Wilhite, D.A.,. Estimating the economic impacts of drought [abs.]. 2004(Proceedings of the 84th annual meeting of the American Meteorological Society,Seattle, Washington).

Hoeting JA, Madigan D, Raftery AE, Volinsky CT. Bayesian model averaging: A tutorial. Statistical Science 1999;14(4):382-401.

Jain S, Lall U. Floods in a changing climate: Does the past represent the future? Water Resources Research 2001;37(12):3193-3205.

Kiehl JT, Hack JJ, Bonan GB, Boville BA, Williamson DL, Rasch PJ. The National Center for Atmospheric Research Community Climate Model: CCM3. Journal of Climate 1998;11(6):1131-1149.

Krishnamurti TN, Kishtawal CM, LaRow TE, Bachiochi DR, Zhang Z, Williford CE, et al. Improved weather and seasonal climate forecasts from multimodel superensemble. Science 1999;285(5433):1548-1550.

Landman WA, Goddard L. Statistical recalibration of GCM forecasts over southern Africa using model output statistics. Journal of Climate 2002;15(15):2038-2055.

Lecce SA. Spatial variations in the timing of annual floods in the southeastern United States. Journal of Hydrology 2000;235(3-4):151-169.

Leung LR, Hamlet AF, Lettenmaier DP, Kumar A. Simulations of the ENSO hydroclimate signals in the Pacific Northwest Columbia River basin. Bulletin of the American Meteorological Society 1999;80(11):2313-2329.

Marshall L, Sharma A, Nott D. Modeling the catchment via mixtures: Issues of model specification and validation. Water Resources Research 2006;42(11).

Mason SJ, Mimmack GM. Comparison of some statistical methods of probabilistic forecasting of ENSO. Journal of Climate 2002;15(1):8-29.

Moura AD, Shukla J. On the Dynamics of Droughts in Northeast Brazil - Observations, Theory and Numerical Experiments with a General-Circulation Model. Journal of the Atmospheric Sciences 1981;38(12):2653-2675.

Murphy AH. Ranked Probability Score and Probability Score - a Comparison. Monthly Weather Review 1970;98(12):917-&.

Palmer TN, Brankovic C, Richardson DS. A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. Quarterly Journal of the Royal Meteorological Society 2000;126(567):2013-2033.

Piechota TC, Dracup JA. Drought and regional hydrologic variation in the United States: Associations with the El Nino Southern Oscillation. Water Resources Research 1996;32(5):1359-1373.

Pizaro G, and U.Lall. El Nino and Floods in the US West: What can we expect? Eos Trans 2002.

Quan X, Hoerling M, Whitaker J, Bates G, Xu T. Diagnosing sources of US seasonal forecast skill. Journal of Climate 2006;19(13):3279-3293.

Rajagopalan B, Lall U, Zebiak SE. Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. Monthly Weather Review 2002;130(7):1792-1811.

Rasmusson EM, Carpenter TH. Variations in Tropical Sea-Surface Temperature and Surface Wind Fields Associated with the Southern Oscillation El-Nino. Monthly Weather Review 1982;110(5):354-384.

Regonda SK, Rajagopalan B, Clark M, Zagona E. A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. Water Resources Research 2006;42(9).

Rhome JR, Niyogi DS, Raman S. Mesoclimatic analysis of severe weather and ENSO interactions in North Carolina. Geophysical Research Letters 2000;27(15):2269-2272.

Roads J, Chen S, Cocke S, Druyan L, Fulakeza M, LaRow T, et al. International Research Institute/Applied Research Centers (IRI/ARCs) regional model intercomparison over South America. Journal of Geophysical Research-Atmospheres 2003;108(D14).

Robertson AW, Kirshner S, Smyth P. Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. Journal of Climate 2004;17(22):4407-4424.

Robertson AW, Lall U, Zebiak SE, Goddard L. Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. Monthly Weather Review 2004;132(12):2732-2744.

Ropelewski CF, Halpert MS. Global and Regional Scale Precipitation Patterns Associated with the El-Nino Southern Oscillation. Monthly Weather Review 1987;115(8):1606-1626.

Roswintiarti O, Niyogi DS, Raman S. Teleconnections between tropical Pacific sea surface temperature anomalies and North Carolina precipitation anomalies during El Nino events. Geophysical Research Letters 1998;25(22):4201-4204.

Sankarasubramanian A, Lall U. Flood quantiles in a changing climate: Seasonal forecasts and causal relations. Water Resources Research 2003;39(5).

Schmidt N, Lipp EK, Rose JB, Luther ME. ENSO influences on seasonal rainfall and river discharge in Florida. Journal of Climate 2001;14(4):615-628.

Seo DJ, Koren V, Cajina N. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. Journal of Hydrometeorology 2003;4(3):627-641.

Sharma A. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 - A strategy for system predictor identification. Journal of Hydrology 2000;239(1-4):232-239.

Shukla J, Anderson J, Baumhefner D, Brankovic C, Chang Y, Kalnay E, et al. Dynamical seasonal prediction. Bulletin of the American Meteorological Society 2000;81(11):2593-2606.

Souza FA, Lall U. Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. Water Resources Research 2003;39(11).

Stephenson DB, Coelho CAS, Doblas-Reyes FJ, Balmaseda M. Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. Tellus Series a-Dynamic Meteorology and Oceanography 2005;57(3):253-264.

Trenberth KE, Guillemot CJ. Physical processes involved in the 1988 drought and 1993 floods in North America. Journal of Climate 1996;9(6):1288-1298.

Von Stroch H and Zwiers FW. Statistical Analysis in Climate Research: Cambridge
731 University Press; 2002.

Weaver CJ. The Drought of 1998–2002 in North Carolina—Precipitation and Hydrologic Conditions; 2005.

Wilks DS. Statistical Methods in the Atmospheric Sciences: Academic Press; 1995.

Wood AW, Kumar A, Lettenmaier DP. A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States. Journal of Geophysical Research-Atmospheres 2005;110(D4).

Wood AW, Maurer EP, Kumar A, Lettenmaier DP. Long-range experimental hydrologic forecasting for the eastern United States. Journal of Geophysical Research-Atmospheres 2002;107(D20).

Yonts W. Written Article: North Carolina Department of Environment and Natural Resources 2004.

Zorn MR, Waylen PR. Seasonal response of mean monthly streamflow to El Nino/Southern Oscillation in north central Florida. Professional Geographer 1997;49(1):51-62.

Appendix

Appendix A: Rank Probability Score and Rank Probability Skill Score

Given that seasonal forecasts are better represented probabilistically using ensembles, expressing the skill of the forecasts using correlation requires summarizing the forecasts using some measures of central tendency such as mean or median of the conditional distribution, which does not give any credit to the probabilistic information in the forecast. Rank Probabilistic Skill Score (RPSS) computes the cumulative squared error between the categorical forecast probabilities and the observed category in relevance to a reference forecast (Wilks, 1995). Here category represents dividing the climatological/observed streamflow, Q, into d=1, 2... D divisions and expressing the marginal probabilities as $P_d(Q)$. Typically, the divisions are made equal probabilistically with O=3 categories known as terciles with each category having 1/3 probability of occurrence. These three categories are known as below normal, normal and above-normal whose end points provide streamflow values corresponding to the particular category. Thus, for a total of D categories, the end points based on climatological observations for d^{th} category could be written as Q_d, Q_{d+1} (For d=1, Q_1= 0; d=D; $Q_{D+1} = Q_{max}$). Given streamflow forecasts at time't' from m^{th} model with i=1, 2... N ensembles, $Q^m_{i,t}$, then the forecast probabilities for the d^{th} category could be expressed as $FP^m_{d,t}(Q) = n^m_{d,t} / N$ by computing the number of ensembles between $Q_d \le Q^m_{i,t} \le Q_{d+1}$.To compute RPSS, the first step is to compute Rank Probability Score (RPS). Given D categories and $FP^m_{d,t}(Q)$ for a forecast, we can express the RPS for a particular year't' from m^{th} model as

$$RPS^m_t = \sum_{d=1}^{D} \left[CF^m_{d,t} - CO_d \right]^2 \qquad \text{... (A-1)}$$

where $CF_{d,t}^{m} = \sum_{q=1}^{d} FP_{d,t}^{m}$ is the cumulative probabilities of forecasts up to category d and CO_d is the cumulative probability of the observed event up to category d. Thus if Q_t, the observed streamflow falls in the d^{th} category, $CO_q = 0$ for $1 \le q \le d-1$ and $CO_q = 1$ for $d \le q \le D$. Given RPS, we can compute RPSS in relation to a reference forecast, which is usually climatological forecasts having equal probability of occurrence under each category $FP_{d,t}^{c\lim}(Q) = 1/D$.

$$RPSS_t^m = 1 - \frac{RPS_t^m}{RPS_t^{c\lim}} \qquad \text{...(A-2)}$$

Low RPS indicates high skill and vice versa. Similarly the range of RPSS varies from minus infinity to 1. RPSS of 0 indicates that there is no skill in the model when compared to the reference forecast. If RPSS is positive, then the forecast skill exceeds that of the climatological probabilities. RPSS of 1 indicates perfect forecast. RPSS could give an overly pessimistic view of the performance of the forecasts and it is a tough metric for evaluating probabilistic forecasts [Goddard et al., 2001]. In this study, we have computed RPS and RPSS for each year and both regression and resampling ensembles by assuming D=3. One can use RPSS to produce maps showing the special characteristics of the forecast skill [Goddard et al., 2001]. Using these maps of RPSS we can examine the spatial distribution of the skill of the forecast. One can also compare RPSS analogously to correlation. RPSS of 0.1 approximately corresponds to a correlation of 0.5 [Goddard et al., 2003; Barnston et al., 2003]. A detailed example on how to compute RPS and RPSS for given forecast, is given below [Goddard et al., 2003].

Illustration of RPS and RPSS for evaluation of probability forecasts

Let us consider a forecast precipitation for the upcoming season 't' have probabilities of 50%, 30% and 20% under below normal, normal and above normal categories respectively from the resampling model. For this forecast, we evaluate how RPS and RPSS will change if the observation falls in each of the categories. Probabilities of climatological ensembles naturally take 33%, 33% and 33%. From the given forecasts, cumulative forecasts, $CF_{d,t}^{m}$ under each category could be calculated as follows. $CF_{d,t}^{m} = \sum_{q=1}^{d} FP_{d,t}^{m}$

Thus, $CF_{1,t}^{1} = 0.5$, $CF_{2,t}^{1} = 0.8$, $CF_{3,t}^{1} = 1.0$ for the given model m = 1 representing resampling model. Similarly, we can also compute the cumulative probabilities under climatology with $CF_{1,t}^{1} = 0.33$ $CF_{2,t}^{1} = 0.66$ $CF_{3,t}^{1} = 1.0$.

Observed Category: Below normal

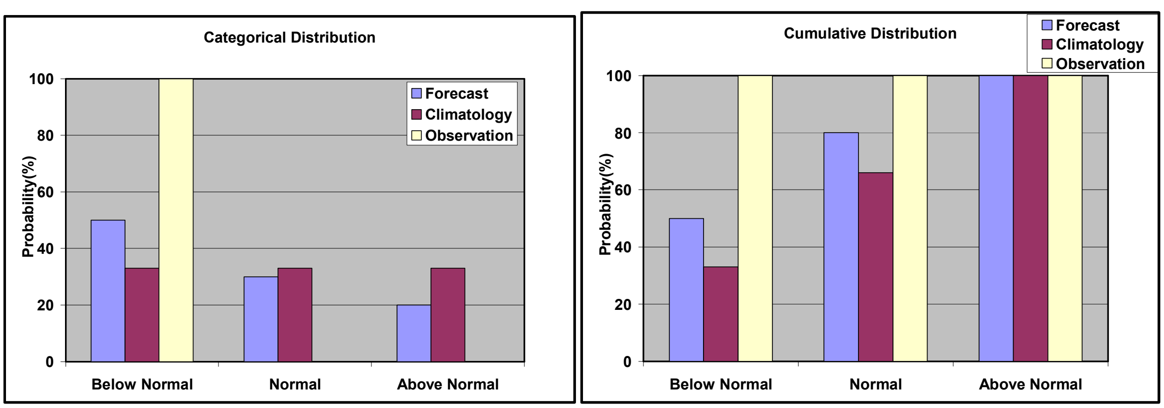


Figure A-1: Observed Category falling in Below Normal

Suppose if the observation falls under below-normal category as shown in figure A-1, then $CO_1 = 1$, $CO_2 = 1$ and $CO_3 = 1$ indicating the cumulative probabilities of observed event for each category'd'.

Hence $\quad RPS_{forecast} = (0.5-1)^2 + (0.8-1)^2 + (1-1)^2 = 0.25 + 0.04 + 0 = 0.29$

Similarly $\quad RPS_{climatology} = (0.33-1)^2 + (0.67-1)^2 + (1-1)^2 = 0.4489 + 0.1089 + 0 = 0.5578$

$$RPSS_{forecast} = 1- RPS_{forecast}/RPS_{climatology}$$

$$1-(0.29/0.5578) = 0.48$$

Thus RPS of the forecast is smaller than the RPS of climatology with smaller error in probabilities of forecasts. This leads to a positive RPSS which compares the performance of candidate forecasts with climatology.

Observed Category: Normal

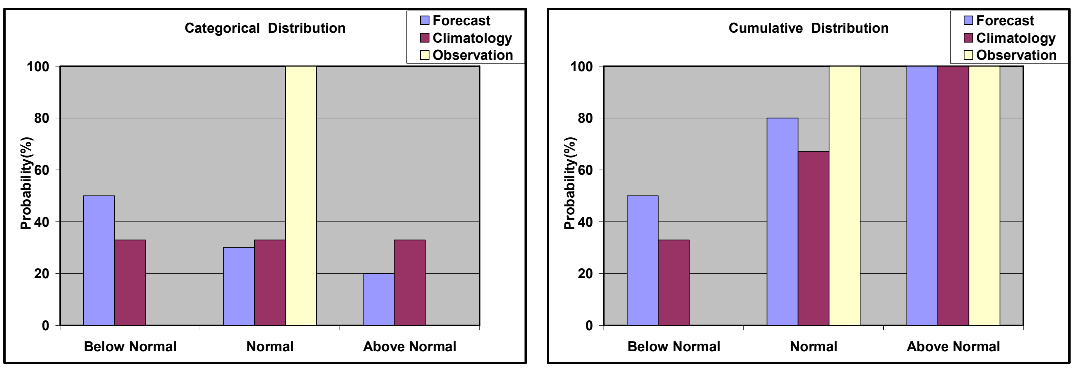


Figure A-2: Observed Category falling in Normal

Now, we consider the observation to be falling under normal category. This changes the cumulative probabilities of observed event, $CO_1 = 0$, $CO_2 = 1$ and $CO_3 = 1$ under the three categories.

Computing

$$RPS_{forecast} = (0.5-0)^2 + (0.8-1)^2 + (1-1)^2 = 0.25 + 0.04 + 0 = 0.29$$

Similarly $\quad RPS_{climatology} = (0.33-0)^2 + (0.67-1)^2 + (1-1)^2 = 0.1089 + 0.1089 + 0 = 0.22$

Hence $\quad RPSS_{forecast} = 1- RPS_{forecast}/RPS_{climatology}$

$$1-(0.29/0.22) = -0.32$$

This shows clearly that if the observation falls in a category which is different from the category in which forecast has higher probabilities, then RPS of the forecast increases leading to reduced RPSS.

Observed Category: Above normal

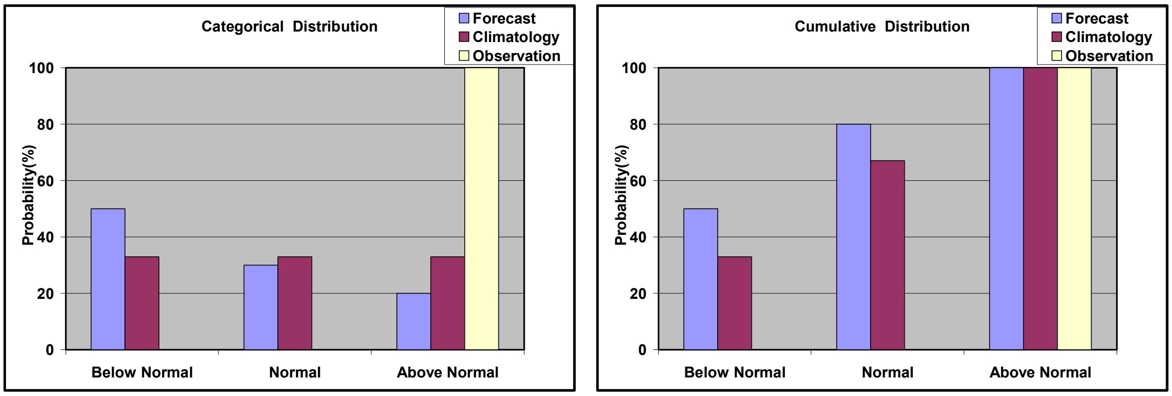


Figure A-3: Observed Category falling in the Above Normal

Now, we consider the observation to be falling under above normal category. The cumulative probabilities of observed events are, $CO_1 = 0$, $CO_2 = 0$ and $CO_3 = 1$ under the three categories.

Hence $RPS_{forecast} = (0.5-0)^2 + (0.8-0)^2 + (1-1)^2 = 0.25 + 0.64 + 0 = 0.89$

$RPS_{climatology} = (0.33-0)^2 + (0.67-0)^2 + (1-1)^2 = 0.1089 + 0.4489 + 0 = 0.56$

$RPSS_{forecast} = 1 - RPS_{forecast}/RPS_{climatology}$

$1-(0.89/0.56) = -0.59$

Thus in this case, the forecast is completely wrong with the prediction exactly opposite of the forecasts. This leads to RPS of the forecast being higher than that of RPS of climatology. Thus RPS is nothing but denoting the error in cumulative probabilities. If both observations falls under a category in which forecast has higher density, then RPS is less. Hence, if one can predict when such situation can occur, it may be advisable to use climatology than using forecasts.