ABSTRACT

DEVINENI, NARESH. Seasonal Hydroclimatology of the Continental United States: Forecasting and its Relevance to Water Management. (Under the direction of Sankar Arumugam).

Recent research in seasonal climate prediction has focused on combining multiple atmospheric General Circulation Models (GCMs) to develop multimodel ensembles. A new approach to combine multiple GCMs is proposed by analyzing the skill of candidate models contingent on the relevant predictor(s) state. To demonstrate this approach, we combine historical simulations of winter (December-February, DJF) precipitation and temperature from seven GCMs by evaluating their skill – represented by Mean Square Error (MSE) – over similar predictor (DJF Nino3.4) conditions. The MSE estimates are converted into weights for each GCM for developing multimodel tercile probabilities. A total of six multimodel schemes are considered that includes combinations based on pooling of ensembles as well as based on the long-term skill of the models. To ensure the improved skill exhibited by the multimodel scheme is statistically significant, we perform rigorous hypothesis tests comparing the skill of multimodels with individual models' skill. The multimodel combination contingent on Nino3.4 show improved skill particularly for regions whose winter precipitation and temperature exhibit significant correlation with Nino3.4.

Analyses of weights also show that the proposed multimodel combination methodology assigns higher weights for GCMs and lesser weights for climatology during El Nino and La Nina conditions. On the other hand, due to the limited skill of GCMs during neutral conditions over the tropical Pacific, the methodology assigns higher weights for climatology resulting in improved skill from the multimodel combinations. The proposed methodology is also evaluated within a forecasting context by combining real-time

precipitation forecasts from five different coupled GCMs contingent on the forecasted Nino3.4. Thus, analyzing GCMs' skill contingent on the relevant predictor state provide an alternate approach for multimodel combination such that years with limited skill could be replaced with climatology.

The utility of the proposed multimodel combination methodology in the context of short-term (monthly to seasonal) water management is investigated by utilizing 3-month ahead probabilistic multimodel streamflow forecasts developed using climate information – sea surface temperature conditions in the tropical Pacific, tropical Atlantic, and over the North Carolina coast – to invoke restrictions for Falls Lake Reservoir in the Neuse River Basin, NC. Multimodel streamflow forecasts developed from two single models, a parametric regression approach and semiparametric resampling approach, are forced with a reservoir management model that takes ensembles to estimate the reliability of meeting the water quality and supply releases and the end of the season target storage. The study suggests that, by constraining the end of the season target storage conditions being met with high probability, the climate information based streamflow forecasts could be utilized for invoking restrictions during below-normal inflow years. Further, multimodel forecasts perform better in detecting the below-normal inflow conditions in comparison to single model forecasts by reducing false alarms and missed targets which could improve public confidence in utilizing climate forecasts for developing proactive water management strategies. This research also presents a systematic analysis for understanding the seasonal hydroclimatology of the continental United States. The relationship of seasonality in precipitation and temperature to mean monthly runoff are analyzed for 1373 watersheds across the U.S. using a physical model with no calibration.

Seasonal Hydroclimatology of the Continental United States: Forecasting and its Relevance to Water Management

by
Naresh Devineni

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Civil Engineering

Raleigh, North Carolina

May, 2010

APPROVED BY:

_____	_____
Dr. Sankar Arumugam Committee Chair	Dr. S. Ranji Ranjithan
_____	_____
Dr. E. Downey Brill	Dr. Fred H. M. Semazzi

DEDICATION

This work is dedicated to my parents, Murali Krishna and Jayasree Ramu Devineni.

BIOGRAPHY

Naresh Devineni is born on June 23rd 1983 in Guntur, Andhra Pradesh, India. He is the younger of the two sons to Mr. Murali Krishna Devineni and Mrs. Jayasree Ramu Devineni. He got his bachelor's degree in Civil Engineering from Osmania University, Hyderabad, India in 2005. Soon after that, he moved to the United States to pursue higher education. He obtained his Master of Science in Civil Engineering with a specialization in Water Resources Engineering in 2007 from North Carolina State University, Raleigh, North Carolina. He is currently pursuing his PhD in Water Resources Engineering from North Carolina State University and expects to graduate in May 2010. As a graduate student, he worked as a Research and Teaching Assistant for the Department of Civil, Construction and Environmental Engineering at North Carolina State University. He is one of the recipients of the outstanding student paper award at the American Geophysical Union Fall Meeting, 2007, San Francisco, CA, for his research. He also worked as a Consultant for the World Bank, Washington DC and carried out reviews focusing on improving the Water Resources Management in developing countries of Africa. He is interested in pursuing a research and academic career in the field of Climate, Hydrology and Water Resources.

ACKNOWLEDGMENTS

I would like to express my profound sense of gratitude to Dr. Sankar Arumugam for being an excellent mentor and for providing me with necessary guidance and support at every stage of my graduate education. I am deeply indebted to him for transforming me into responsible research personnel. I am grateful to Dr. Ranji Ranjithan who has given feedback from time to time on my performance. I would also like to thank Dr. Downey Brill and Dr. Fred Semazzi for their encouraging words on my performance. I feel honored to work under the guidance of these people. Their constant encouragement was a great support for me.

I would like to acknowledge the Water Resources Research Institute (WRRI) of North Carolina for funding the project. I also thank Dr. Lisa Goddard of IRI for her useful discussions and feedback on the research. 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), International Research Institute for Climate and Society(IRI), European Center for Medium Range Weather Forecasts (ECMWF) for providing us with the data. I also thank Dr. Satish Regonda of NOAA for his valuable feedback and discussions on my work.

I am always grateful to my parents and grandparents, with whose constant support, encouragement and blessings I was able to come to the United States to pursue my higher education. I also thank my brother Ravi Chandra Devineni for his constant encouragement at every stage of my education. I would like to thank all my friends for their help and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Motivation and Problem Context

Seasonal to interannual climatic variations resulting from changing Sea Surface Temperature (SST) and atmospheric conditions influence the regional to continental scale hydroclimatology. For instance, it is well known that quasi-oscillatory modes such as the El Nino Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), and the Pacific Decadal Oscillation (PDO) influence the seasonal to interannual precipitation variability of the continental United States (U.S). These ocean – atmosphere interactions modulate the moisture delivery pathway and have significant influence on the rainfall [Trenberth and Guillemot 1996; Cayan et al. 1999] and streamflow patterns at regional and global scales [Dettinger et al. 2000]. In addition, initial land surface moisture conditions such as snowpack and soil moisture storage also play an important role in determining the seasonal streamflow potential of a basin. Hence, identification of these dominant climatic modes and estimation of land surface states provide vital information for predicting hydrologic fluxes well in advance which could be beneficial in developing improved water management strategies. Both national and international agencies monitor these climatic variations to issue seasonal to interannual climate forecasts. Efforts in the development and application of seasonal to long-lead climate forecasts have grown tremendously in recent times.

Seasonal climate forecasts, typically represented as ensembles of precipitation and temperature, are obtained by forcing SST forecasts with atmospheric General Circulation Models (GCMs). Seasonal climate forecasts could also be obtained from coupled GCMs, which obtains SSTs (i.e. boundary conditions) using ocean-atmosphere-land surface coupling. Errors resulting from climate forecasts can be classified into two types. The first

type of error is the error due to uncertainty in initial and boundary conditions, and the second type of error is the model error [Hagedorn et al. 2005]. While the first source of error can be resolved by representing the uncertainties in initial and boundary 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 forcing). A common approach to reduce model uncertainty is through refinement of parameterizations and process representations in the considered GCM. 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]. Thus, combining climate forecasts from multiple models seems to be a good alternative in improving the overall predictability of seasonal forecasts and reducing the overall error in prediction. Application of streamflow forecasts developed from multiple models has been shown to be beneficial in invoking restrictions in meeting the summer demand and in improving the end of season storage [Golembesky et al. 2009].

One of the main objectives of this research is to improve the predictability of winter precipitation and temperature over the continental U.S. by optimally combining multiple atmospheric GCMs. Since the skill of the GCMs over the U.S. is primarily dependent on the state of SST conditions in the tropical Pacific [Quan et al. 2006; Shukla et al. 2000; Brankovic and Palmer 2000], we combine multiple GCMs using the algorithm developed by Devineni et al. [2008]. The proposed methodology assigns weights for each GCM by evaluating their skill over similar predictor conditions. By combining the GCMs with

climatological ensembles, we develop different multimodel combination schemes that give higher weights for GCMs that performs better during ENSO states.

Seasonal climate/streamflow forecasts could be utilized effectively for managing the water and energy supply systems and to provide useful information to planners and operational agencies towards developing contingency measures and improved management strategies during extreme hydroclimatic conditions (e.g. floods and droughts). This research explores the utility of the above mentioned multimodel combination methodology in the context of short-term water management by investigating various drought management strategies for Falls Lake Reservoir in Neuse River basin, NC. Under this study, we apply climate information based multimodel streamflow forecasts (developed based on the above methodology) for invoking restrictions on water supply releases from the Falls Lake Reservoir.

Finally, this dissertation also focuses on understanding the seasonal hydroclimatology of the continental U.S. An investigation on the 1373 watersheds across the U.S. showed that the observed monthly climatology of streamflows over the eastern U.S. showed significant differences from the month of precipitation climatology. Though the distribution of monthly precipitation is uniform throughout the year over the eastern U.S, streamflow exhibit pronounced seasonality with peak flow seasons occurring during the winter over the Southeast and during the spring over the Mid-Atlantic and Northeast regions. To physically explain this discordance in streamflow seasonality, we systematically analyze the basin response by forcing the water balance models with synthetically generated inputs – precipitation and temperature – and the observed time series.

1.2 Outline of the Dissertation

The dissertation is organized as follows: Chapter 2 presents the detailed analysis of the multimodel combinations using precipitation and temperature simulations (i.e. forced with observed SSTs) from seven different atmospheric GCMs. Since the simulated variables from GCMs overestimate the potential forecasting skill, we investigated the utility of the methodology in combining the real-time precipitation forecasts (i.e. forced with forecasted SSTs) from five different coupled GCMs. This analysis is presented in Chapter 3. Chapter 4 focuses on exploring the utility of the multimodel combination methodology for improving the drought management of Falls Lake in the Neuse River basin, NC. In this analysis, three-month ahead multimodel streamflow forecasts are used with a reservoir simulation model for setting up restrictions on water supply releases from the Falls Lake Reservoir. Chapter 5 presents the analysis on the role of seasonality of precipitation and temperature in explaining the climatology of monthly runoff. Finally, in Chapter 6, we present the summary and overview of the entire research work.

1.3 References

Barnston, A.G., S.J.Mason, L.Goddard, D.G.DeWitt and S.E.Zebiak (2003), Multimodel ensembling in seasonal climate forecasting at IRI, *Bulletin of the American Meteorological Society*; 84(12):1783-+.

Branković Č., and T. N. Palmer, 2000: Seasonal skill and predictability of ECMWF PROVOST ensembles. *Quart. J. Roy. Meteor. Soc.*, **126**, 2035–2067.

Cayan, D. R., K. T. Redmond, and L. G. Riddle (1999), ENSO and Hydrologic Extremes in the Western United States, *Journal of Climate*, 12, 2881-2893.

Dettinger, M.D. and H.F. Diaz (2000), Global characteristics of stream flow seasonality and variability, *Journal of Hydrometeorology*, 1 (4): 289-310.

Devineni, N., A. Sankarasubramanian, and S. Ghosh (2008), Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations, *Water Resour. Res.*, 44, W09404, doi:10.1029/2006WR005855.

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

Golembesky, K, A.Sankarasubramanian, and N.Devineni (2009), Improved Drought Management of Falls Lake Reservoir: Role of Multimodel Streamflow Forecasts in Setting up Restrictions, *Journal of Water Resources Planning and Management,* 135, 188, DOI:10.1061/(ASCE)0733-9496(2009)135:3(188).

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

Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran (1999), Improved weather and seasonal climate forecasts from multimodel superensemble, *Science*, 285, 1548-1550.

Quan X., M. Hoerling, J. Whitaker, G. Bates, and T. Xu, 2006: Diagnosing sources of US seasonal forecast skill, *J. Climate*, **19,** 3279-3293.

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

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

Shukla J., J. Anderson, D. Baumhefner, C. Brankovic, Y. Chang, E. Kalnay, L. Marx, T. Palmer, D. Paolino, J. Ploshay, S. Schubert, D. Straus, M. Suarez, and J. Tribbia, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593-2606.

Trenberth, K. E., and C. J. Guillemot (1996), Physical Processes involved in the 1988 Drought and 1993 Floods in North America, *Journal of Climate*, 9, 6, 1288–1298.

CHAPTER 2

Improving the Prediction of Winter Precipitation and Temperature over the continental United States: Role of ENSO State in Developing Multimodel Combinations

2.1 Introduction

Planning and management of water and energy systems are usually carried out based upon the seasonal climate (precipitation and temperature) forecasts over a particular region. Several national and international agencies routinely issue climate forecasts using coupled General Circulation Models (GCMs) [e.g., Saha et al. 2006] as well as using atmospheric GCMs (AGCMs) [e.g., Goddard et al. 2003]. Forecasts from AGCMs are typically developed in a two-tiered process with sea surface temperature (SST) being predicted first from an ocean-atmosphere model and then the predicted SSTs are forced as boundary conditions into the AGCMs. This two-tiered approach primarily emphasizes that much of the predictability at seasonal time scales primarily stems from the oceanic conditions with the ensembles representing the atmospheric internal variability. However, the skill of the climate forecasts could vary substantially depending on the location, time and the GCMs itself [Doblas-Reyes et al. 2000; Robertson et al. 2004].

Reducing model uncertainties through the conventional approach of refinement of parameterizations and improved process representation is time consuming which led recent efforts to focus on the combination of AGCMs for improving seasonal climatic prediction [Krishnamurthi et al. 1999; Doblas-Reyes et al. 2000; Rajagopalan et al. 2002]. Research studies from PROVOST (PRediction Of climate Variations On Seasonal to interannual Time-scales) and from International Research Institute for Climate and Society (IRI) show that multimodel combination of AGCMs provide better calibration (i.e., improved reliability and resolution) than individual model predictions [Doblas-Reyes et al. 2000; Barnston et al. 2003].

Studies from DEMETER (Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction) experiments show that multimodel combination of CGCMs also improve the reliability and skill in predicting summer precipitation in the tropics and winter precipitation in the northern extratropics [Palmer et al. 2004]. Hagedorn et al. [2005] demonstrate that the superiority of multimodels primarily arise from error cancellation, not due to increased ensemble members (i.e., pooling of ensemble from single models), resulting in improved reliability and consistency. Reviewing the skill of various forecast products – medium-range, monthly and seasonal – over the Europe, Rodwell and Dobas-Reyes [2006] show that multimodel ensembles of CGCMs exhibit higher skill in predicting precipitation and temperature during both winter and summer seasons. Recently, multimodel ensembles developed using 46-year hindcasts from five CGCMs run from the European Union's ENSEMBLES project show better skill in predicting tropical SSTs than the multimodel ensembles developed using re-forecasts from the DEMETER project [Weisheimer et al. 2009].

The simplest approach to develop multimodel combination is to pool the ensembles from all the models by giving equal weights for all the models [Palmer et al. 2000]. A different approach to develop multimodel ensembles is to optimally combine multiple GCMs so that the resulting multimodel forecast has better skill than individual model forecasts [Rajagopalan et al. 2002; Robertson et al. 2004; DelSole, 2007]. Under optimal combination approach, weights are obtained for each GCM as a fraction such that the chosen skill metric of the multimodel ensembles is maximized [Rajagopalan et al. 2002; Robertson et al. 2004]. Doblas-Reyes et al. [2005] compare the performance of two multi-model combination

techniques – equal weighting of all models and optimal combination using multiple linear regression – and show that, except in tropical regions, it is difficult to improve the performance of optimal combination due to small sample size. Studies have also employed simple statistical techniques such as linear regression [Krishnamurthi et al. 1999] to advanced statistical techniques such as canonical variate method [Mason and Mimmack, 2002] and Bayesian techniques [Hoeting et al. 1999; Stephenson et al. 2005; Luo et al. 2007] for developing multimodel combinations. DelSole [2007] proposed a Bayesian multimodel regression framework that incorporates prior beliefs about the model weights for estimating regression parameters.

It is well known that anomalous conditions in the tropical Pacific influence the skill of GCMs in predicting precipitation and temperature over North America [Shukla et al. 2000; Quan et al. 2006]. Upon diagnosing the sources of seasonal prediction skill over the United States, Quan et al. [2006] show that the entire skill of the AGCMs could be explained by El-Nino Southern Oscillation (ENSO) alone. For this analysis, Quan et al. [2006] considered multimodel mean obtained by averaging all the ensembles (total of 48 simulations) from four AGCMs. Recently, Devineni et al. [2008] proposed a new approach to develop multimodel ensembles of streamflow that combines forecasts from individual models by evaluating their skill contingent on the predictor state.

The main intent of this study is to explore strategies for improving the skill in predicting winter precipitation and temperature over the continental United States by optimally combining multiple GCMs. Given that predicting winter precipitation and temperature over the United States primarily depends on SST conditions over the tropical

Pacific [Quan et al. 2006], we combine multiple GCMs by evaluating the skill of seven AGCMs conditioned on the ENSO state based on the algorithm outlined in Devineni et al. [2008]. The skill of the GCMs contingent on the ENSO state is assessed by averaging the Mean Square Error (MSE) in predictions under similar tropical SST conditions. For this purpose, we consider simulated precipitation and temperature (i.e., e forced with observed SSTs) from seven different AGCMs for developing multimodel combinations. The performance of the developed multimodel tercile probabilities of winter precipitation and temperature are compared with the performance of individual models' as well as with two of the commonly employed techniques for multimodel combination.

For better readability from here on in this dissertation, we often refer to both individual and multimodel simulations of precipitation and temperature as "predictions/forecasts" with an understanding that simulated GCM variables overestimate the potential forecasting skill. Section 2.2 describes the data and the GCMs used for the study along with the description of the multimodel combination methodology. Section 2.3 presents the results and analysis by comparing the skill of individual GCMs and multimodels in predicting the observed winter precipitation and temperature. Finally, in Section 2.4, we summarize the findings and conclusions from the study.

2.2 Multimodel Combination Contingent on Predictor State: Basis and Methodology

2.2.1 Data

Seven AGCMs that are commonly employed by various research institutes and agencies are considered for developing multimodel winter (December – January – February,

(DJF)) precipitation and temperature forecasts over the continental United States. Table 2.1 gives the details on each model along with the number of ensembles available in predicting precipitation and temperature. Historical monthly simulations of winter precipitation and temperature which are developed by forcing the AGCMs with observed SSTs are obtained from [http://iridl.ldeo.columbia.edu/SOURCES/.IRI/.FD/] IRI data library. Figure 2.1a shows the grid points (a total of 192) that are considered for developing multimodel predictions. Observed monthly precipitation and temperature at 0.5°X0.5° available from University of East Anglia (UEA), Climate Research Unit (CRU) [New et al. 2000], is used to assess the skill of each model. Monthly climate anomalies, relative to the 1961–90 mean [New et al. 1999], were interpolated from the station data to develop monthly terrestrial surface climate grids for the period 1901-1996. Recent studies on multimodel combination have used the observed precipitation and temperature for UEA database for showing the improvements resulting from multimodel combination [Rajagopalan et al. 2002; Robertson et al. 2004; DelSole, 2007]. Grid points (0.5°X0.5°) of monthly precipitation and temperature from UEA were spatially averaged to map the grid points of the GCMs.

We consider Nino3.4, the index commonly used to denote the ENSO state, as the primary predictor influencing the winter precipitation and temperature over the United States. Nino3.4 denotes the anomalous SST conditions in the tropical Pacific which are obtained by averaging the SSTs over 5S-5N and 170W-120W. Average DJF Nino3.4, which is computed using Kaplan's SST database [Kaplan et al. 1998], is obtained from IRI data library for the forty six years (DJF 1951 – 1996) [http://iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.Indices/.NINO34/] considered for

verification. El Nino (Nino3.4 > 0.5), La Nina (Nino3.4 < - 0.5) and neutral conditions ((|Nino 3.4| ≤ 0.5) are identified resulting in a total of 14 years of El Nino, 12 years of La Nina and 20 years of neutral conditions from the DJF Nino3.4 time series.

2.2.2 Basis behind Combining Individual GCMs Contingent on the Predictor State

The multimodel combination methodology proposed in this study is motivated based on the premise that model uncertainties could be better reduced by combining the GCMs based on their ability to predict under a given predictor state. Recent studies on seasonal to interannual climate prediction over North America clearly show that the skill of GCMs is enhanced during ENSO years [Brankovic and Palmer, 2000; Shukla et al. 2000; Quan et al. 2006]. To understand this further, Figure 2.1a shows the correlation between the observed precipitation and ensemble mean of the GCM predicted precipitation unconditional of the ENSO state (i.e., over the entire period of record), whereas Figures 2.1b, 2.1c and 2.1d show the skill (correlation) of two GCMs, ECHAM4.5 and ECPC, in simulating winter precipitation under El Nino, La Nina and neutral conditions respectively. The correlations ($1.96/\sqrt{n_s-3}$ where 'n_s' denotes the number of samples under each category) that are statistically significant at 95% confidence interval under El Nino (n_s = 14), La Nina (n_s = 12), neutral conditions (n_s = 20) and over the entire record (n_s = 46) are 0.59, 0.65, 0.48 and 0.30 respectively.

Though Figure 2.1a shows significant correlation at many grid points (> 0.30) for both models, the performance of the models under those grid points are not consistent under three different ENSO conditions. Further, the skill of both GCMs is not-significant/negative

(< 0.50) for most grid points under neutral conditions with the skill being mostly significant only under El Nino (> 0.59) and La-Nina conditions (> 0.65). We can also infer from Figure 2.1 that the significant skill exhibited by both GCMs also varies spatially. Thus, anyone combining the GCMs purely based on the overall skill would end up giving higher weights for the best performing GCM at that grid point, which would naturally result in poor prediction during neutral conditions.

We also compare whether the difference in positive correlations exhibited by these two models are statistically significant using the Hotelling-Williams test [Bobko, 1995]. Details of the Hotelling-Williams test are given in Appendix A. Hotelling-Williams test statistic, $(r_{12} - r_{13})\sqrt{\frac{(N-1)(1+r_{23})}{2(N-1)/(N-3)|R| + \bar{r}^2(1-r_{23})^2}}$, follows a *'t'* distribution with *(N-3)* degrees of freedom where r_{12} *(*r_{13}*)* denoting the correlation between the observed precipitation and ensemble mean from ECHAM4.5 (ECPC), r_{23} denoting the correlation between the ensemble means of ECHAM4.5 and ECPC and '*N*' denoting the total number of years of observation with $\bar{r} = (r_{12} + r_{23})/2$ and $R = (1 - r_{12}^2 - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23})$. A plus (triangle) sign on the grid points in Figure 2.1b indicates that the difference between r_{12} *(*r_{13}*)* and r_{13} *(*r_{12}*)* is greater than zero indicating the better performance of ECHAM4.5 (ECPC). We can see from Figure 2.1 that under El Nino conditions, there are 6 (10) grid points with r_{12} *(*r_{13}*)* being significant over r_{13} *(*r_{12}*).* This primarily shows that the skills exhibited by these models under different ENSO states are statistically significant and also could be completely different from the overall skill of the model. Hence, we propose a methodology that evaluates the performance of GCMs contingent on the dominant predictor state(s) and assigns higher

weights for the best performing model under those predictor conditions. We also consider climatology as one of the candidate models for developing multimodel ensembles. By including climatology, we believe that if the skill of all models is poor under neutral ENSO conditions, the algorithm could give higher weights for climatology in developing multimodel combinations. In the next section, we describe the modified version of multimodel combination algorithm presented in Devineni et al. [2008].

2.2.3 Description of the Multimodel Combination Algorithm

Figure 2.2 provides the flow chart of the multimodel combination algorithm that combines tercile predictions/forecasts from multiple GCMs. Historical simulations of winter precipitation and temperature available for each GCM (1951 - 1996) are converted into tercile categories, $Q_{i,t}^{m}$, where $m=1,2...,M$ (M=8) denotes the model index including climatology, with $i=1, 2..., N$ (N=3) representing the categories in year 't' which specifies the time index with $t=1,2\ldots, T$ (T= 46 years) (for additional details, see Figure 2.2). Tercile categories, $Q_{i,t}^{m}$, are computed from the ensembles of GCMs after removal of systematic bias (i.e., each ensemble is represented as anomaly from the model's seasonal climatology). The squared error (SE_t^m) in predicting the observed precipitation/temperature is computed from the ensemble mean of the simulated precipitation/temperature for each year at 192 grid points over the United States. Based on Quan et al. [2006], we consider ENSO state indicated by Nino3.4 as the dominant predictor in influencing the winter precipitation and temperature over the United States.

The objective is to combine individual model simulations by evaluating their skill – represented by mean squared error (MSE) ($l_{t,K}^{m}$ in Figure 2.2*)* – over '*K*' neighboring predictor conditions. Devineni et al. [2008] considered various skill metrics for evaluating the performance of candidate models over similar predictor conditions and find that mean squared error and average rank probability score perform well in improving the skill of multimodel combination. Further MSE, which is obtained based on the average error in the conditional mean of the forecast over similar predictor conditions, is also a proper skill score [Brocker and Smith, 2007]. A skill score is proper if it maximizes the expected score for an observation drawn from a particular distribution only if the issued probabilistic forecast is of the same distribution [Brocker and Smith, 2007]. Given that we have candidate GCMs with different ensemble size, we did not consider other strictly proper scores such as RPS since their sampling variability heavily depend on the number of ensembles [Weigel et al. 2007]. We identify '*K*' similar ENSO conditions by calculating the euclidean distance between DJF Nino3.4 in the conditioning year'*t*' and the rest of DJF Nino3.4 observed during 1951-1996. *It is important to note that in computing the MSE from 'K' similar climatic conditions, we leave out the skill of the model (SE_t^m) in that conditioning year.*

Based on the MSE computed over '*K*' neighbors, weights ($w_{t,K}^{m}$ in Figure 2.2) for each model are computed for each year using which the multimodel tercile categories are computed. Basically, the weighting scheme should be inversely proportional to a chosen increasing function (e.g., linearly or logarithmic) of the prediction error metric. The idea behind the proposed approach in Figure 2.2 is that the weights for each model vary according

to ENSO conditions. Thus, if a particular GCM performs well during El Nino conditions in a given grid point, then higher weights ($w_{t,K}^{m}$ in Figure 2.2) will be given to the tercile probabilities from that GCM in developing multimodel combinations. Using the algorithm in Figure 2.2, we develop six multimodel combinations of winter precipitation (shown in Table 2.2) and temperature for 192 grid points over the continental United States. We discuss the issue of the selection of optimal number of neighbors '*K*' and various multimodel combination schemes (in Table 2.2) in the next Section.

2.2.4 Multimodel Schemes and Performance Analysis

Table 2.2 provides brief description on the different multimodel schemes considered in the study. Four multimodel (MM) schemes, MM-1, MM-2, MM-3, and MM-4, are developed using the algorithm in Figure 2.2. MM-1 and MM-2 employ fixed neighbors '*K*' to obtain the MSE. Under MM-1 and MM-2, if the Nino3.4 in the conditioning year is under El Nino, La Nina and neutral states, then we estimate the MSE ($l_{t,K}^{m}$) of the GCM during the observed El Nino (K =13 years – leaving the conditioning year out), La Nina (K =11 years) and neutral (K =19 years) years respectively. Thus, for MM-1 and MM-2, we evaluate the skill of the model only under similar ENSO conditions.

With multimodel schemes, MM-3 and MM-4, we obtain the neighbors, K_t, by performing two-deep cross-validation [Stone, 1974]. The two-deep cross validation is a rigorous model validation technique which is generally employed to choose optimum model parameters as well as to reduce over-fitting that typically results in multimodel combination [Delsole, 2007]. The two-deep cross validation technique obtains model predictions

recursively in two stages. In the outer loop, we leave out the predictor (Nino3.4) and the predictand (DJF precipitation/temperature) in year'*t*' and use the remaining '*T-1*' years (*T*=46) to estimate the optimum 'K_t'. For the samples (which constitute *T-1* years of GCM predictions and Nino3.4) in the inner loop, we obtain K_t that minimizes the MSE of multimodel predictions over '*T-1*' years through a leave-one-out cross validation (i.e., model fitting is done with '*T-2*' years and validated for the left out year from the '*T-1*' sample). We employ this 'K_t' from the inner loop to develop the multimodel predictions for year'*t*' in the outer loop. This procedure is repeated for all possible left out samples in the outer loop to obtain multimodel predictions for each year. Thus, under MM-3 and MM-4, the number of neighbors 'K_t' varies from year to year.

We also employ different strategies in combining individual model simulations with climatological ensembles. MM-1 and MM-3 utilize seven different GCMs (Table 2.1) along with climatological ensembles to develop multimodel ensembles. MM-2 and MM-4 combine each model with climatology in the first step and then combine the resulting seven models in the second step. Recent studies have shown that a two step procedure of combining each individual model forecasts separately with climatology and then combining the resulting '*M*' combinations at the second step improves the skill of multimodel ensembles [Robertson et al. 2004]. For climatology, we simply consider the 45 years (leaving the conditioning year out) of observed precipitation and temperature at each grid point from UEA. To compute the squared error for each year at the second step of combination for MM-2 and MM-4, we assume the conditional distribution obtained from first step as normal. MM-P is the multimodel combination scheme that is obtained by pooling all the ensembles from seven

individual models and climatology. The reason we consider climatology under MM-P is to be consistent for comparison with other multimodel schemes (MM-1 to MM-4). Hence, in MM-P scheme, we have an increased number of ensembles (203) since we are now pooling ensembles from all the models.

MM-OS combines individual models based on their overall skill (unconditional of the ENSO state), which is specified based on the MSE for the period 1951-1996 in predicting winter precipitation/temperature at a given grid point. Thus, under MM-OS, the weight $\left(\frac{\left(\overline{MSE}^{m}\right)^{-1}}{\sum_{m=1}^{M}\left(\overline{MSE}^{m}\right)^{-1}} \right)$ for a given model 'm' is obtained based on the inverse of the MSE of model 'm' to the sum of the inverse of MSE of all the models. MM-P and MM-OS provide the baseline comparison with some of the commonly employed techniques in developing multimodel combinations [Palmer et al. 2000; Rajagopalan et al. 2002; Robertson et al. 2004]. The performance of multimodel predictions are compared with individual models' skill using standard verification measures such as average Rank Probability Score ($\overline{RPS}$), average Rank Probability Skill Score ($\overline{RPSS}$), reliability diagrams and average Brier scores. Expressions related to these metrics could be found in Wilks [1995]. Details on estimating the $\overline{RPSS}$ for a given forecast is also given in Appendix B. The next section discusses the performance of multimodels in predicting winter precipitation and temperature over the continental United States.

2.3 Results and Analysis

Six multimodel predictions (in Table 2.2) of winter precipitation and temperature are developed by combing seven AGCMs and climatology based on the algorithm shown in Figure 2.2. The developed multimodel predictions are represented as tercile probabilities in 192 grid points over the continental United States for the period 1951 – 1996.

2.3.1 Baseline Comparison between Multimodels and Individual Models

Figure 2.3 shows the box plot of $\overline{RPSS}$ for the seven individual models and for six multimodels over the entire United States. $\overline{RPSS}$ computes the cumulative squared error between the categorical forecast probabilities and the observed category in relevance to a reference forecast [Wilks, 1995]. The reference forecast is usually climatological ensembles that have equal probability of occurrence under each category. A positive score of $\overline{RPSS}$ indicates that the forecast skill exceeds that of the climatological probabilities. Alternately if the $\overline{RPSS}$ is negative, it indicates that the forecast skill is less than that of climatology. $\overline{RPSS}$ is a rigorous metric for evaluating categorical forecasts, since it evaluates the performance of entire conditional distribution. Using the multimodels' and individual models' tercile probabilities, we compute the $\overline{RPSS}$ for the period 1951-1996.

Figures 2.3a and 2.3b show the box plots of $\overline{RPSS}$ in predicting winter precipitation and temperature respectively over the U.S for the period 1951-1996. Figure 2.3 also show the number of grid points that have $\overline{RPSS}$ greater than zero. Similarly, Figures 2.4a and 2.4b also show the box plots of mean squared error based skill score (MSSS) in predicting winter

precipitation and temperature respectively. For computing MSSS, we assume the conditional distribution resulting from the multimodel combination as normal. From both Figures 2.3 and 2.4, we can infer that the individual models' $\overline{RPSS}$ and MSSS is lesser than zero in most of the grid points which implies that the skill of the AGCMs is poorer than climatology. Among the individual models, ECHAM4.5 and CCM3v6 perform better than other GCMs in predicting winter precipitation and temperature. Further, we can also see that all six multimodels (in Table 2.2) perform better than the GCMs with more grid points having positive $\overline{RPSS}$ and MSSS in predicting winter precipitation and temperature.

Comparing the performance of multimodels in predicting precipitation (Figures 2.3a and 2.4a) and temperature (Figures 2.3b and 2.4b), we infer that the two-step multimodel combination schemes, MM-2 and MM-4, perform better than the currently employed techniques, MM-P and MM-OS, with more number of grid points having positive $\overline{RPSS}$ and MSSS. MM-2, which uses fixed neighbors contingent on ENSO conditions in evaluating the skill of the models, also perform better than MM-P and MM-OS in predicting winter precipitation and temperature. Further, we can notice from both Figures 2.3 and 2.4 that both individual models' and multimodels' have better skill in predicting winter temperature in comparison to the skill in predicting winter precipitation. Among the individual models, we see that ECHAM4.5 and CCM3v6 are the best individual models in predicting winter precipitation and temperature. So, all further analyses in quantifying the improvements resulting from multimodels will focus only on comparing with the performance of ECHAM4.5 and CCM3v6.

2.3.2 Statistical Significance of Multimodel Predictions – Hypothesis Testing

To ensure that the improved $\overline{RPSS}$ exhibited by the multimodel schemes, MM-1-MM-4, is statistically significant over the skill of ECHAM4.5, we perform detailed nonparametric hypothesis tests [Hamill, 1999] by testing the null hypothesis that $\overline{RPS}$ of a multimodel scheme is equal to $\overline{RPS}$ of ECHAM4.5 in predicting precipitation/ temperature at each grid point. With Model A denoting ECHAM4.5 and Model B denoting one of multimodel schemes (in Table 2.2), the null hypothesis for testing $\overline{RPS}$ could be written as:

$$H_O : \overline{RPS}^A - \overline{RPS}^B = 0 \quad \text{... (2.1)}$$

$$H_A : \overline{RPS}^A - \overline{RPS}^B \neq 0 \quad \text{... (2.2)}$$

The distribution of null hypothesis, $\overline{RPS}^{1,*} - \overline{RPS}^{2,*}$, is constructed by resampling equally likely each year from the RPS_t^A of Model A (i.e., ECHAM4.5) and RPS_t^B of Model B (MM-1 to MM-4, MM-P and MM-OS). In other words, $\overline{RPS}^{1,*}$ and $\overline{RPS}^{2,*}$, the average rank probability score ($\overline{RPS}$) estimated from 46 years to construct the null distribution, incorporate RPS_t^A and RPS_t^B equally likely. A total of 10000 estimates of $\overline{RPS}^{1,*} - \overline{RPS}^{2,*}$ are obtained to develop the null distribution for each grid point over the United States. The percentiles at which the observed test statistic at each grid point, $\overline{RPS}^A - \overline{RPS}^B$, has fallen in the constructed null distribution is computed. Results from the hypothesis tests, the percentiles of the observed test statistic $\overline{RPS}^A - \overline{RPS}^B$ on the constructed null distribution, are plotted on the United States map to identify grid points showing significant improvement

from multimodel combination (Figures 2.5 and 2.6). For a significance level of 10%, if the percentile of the observed test statistic is between 0.9-1 (0-0.1) at a given grid point, then Model B (Model A)'s $\overline{RPS}$ is statistically lesser than Model A (Model B)'s $\overline{RPS}$. For additional details on the performed nonparametric hypothesis test, see Hamill [1999].

Tables 2.3 (precipitation) and 2.4 (temperature) summarize the results from hypothesis tests across the six multimodels. Entries in the upper triangle in Tables 2.3 and 2.4 provide the number of grid points having the percentiles of observed test statistic between 0.9-1 on the constructed null distribution which implies that $\overline{RPS}$ of Model B – represented as column – is statistically significant than the $\overline{RPS}$ of Model A, which is represented as row entry. For instance, from the upper triangle in Table 2.3 from the hypothesis tests between MM-1 (Model A) and MM-P (Model B), we find that MM-P's $\overline{RPS}$ is statistically smaller than the $\overline{RPS}$ of MM-1 in 24 grid points (with the percentiles of the observed test statistic between 0.9-1 in the null distribution). Similarly, results from the same hypothesis tests are also summarized in the lower triangle between the two models indicating the number of grid points over which the percentiles of the observed test statistic fell between 0-0.1 on the constructed null distribution, which implies the MM-1's (Model A) $\overline{RPS}$ is statistically smaller than the $\overline{RPS}$ of MM-P (Model B) in 20 grid points. For both Tables 2.3 and 2.4, the best-performing model in terms of increased number of significant grid points is underlined by its column entry. Thus, between MM-1 and MM-P, we infer that MM-P (underlined by the column) performs better in more grid points in comparison to MM-1 in predicting precipitation.

Figure 2.5 shows the relative performance of six multimodel combination schemes over the best individual model, ECHAM4.5, in predicting winter precipitation over the entire United States. From Figure 2.5a and Table 2.3, 39 (5) grid points have the percentiles of the test statistic falling between 0.9-1(0- 0.1) which indicate that MM-1(ECHAM4.5) performs better than ECHAM4.5 (MM-1) in those grid points by rejecting the null hypothesis that the difference in $\overline{RPS}$ between the ECHAM4.5 and MM-1 is zero for a significance level of 10%. We can also see that many grid points fell between 0.1 – 0.9 indicating the difference in skill in not statistically significant at 10%. However, a plus symbol is shown in Figure 2.5 to indicate $\overline{RPS}$ of the corresponding multimodel at that grid is lesser than $\overline{RPS}$ of ECHAM4.5. Even though the difference in $\overline{RPS}$ is statistically not significant at 10%, we observed that the percentage reduction in using multimodel combinations is around 5 – 15% for grid points with the percentiles of the observed test statistic, $\overline{RPS}^{A} - \overline{RPS}^{B}$,from 0.5 – 0.9.

Among the multimodels, MM-2 and MM-4 perform better than the rest of the multimodels, which is indicated by more number of grid points (Figure 2.3 and Table 2.3) having statistically significant $\overline{RPS}$ than the $\overline{RPS}$ of the rest of the multimodels and ECHAM4.5. From Table 2.3, we clearly understand that multimodel scheme proposed in this study, MM-2 and MM-4, perform better than the existing techniques on multimodel combinations (MM-P and MM-OS). It is important to note that both MM-2 and MM-4 employ two-step combination to develop multimodel predictions. Comparing between MM-2 and MM-4, we infer that in 52(25) grid points MM-2's (MM-4) $\overline{RPS}$ is statistically

significant than the $\overline{RPS}$ of MM-4 (MM-2) with the observed test statistic between the two models falling between 0.9-1 (0-0.1) on the constructed null distribution. This indicates that two-step combination seems to be more effective in reducing the $\overline{RPS}$ of multimodels in predicting precipitation. Recently, Chowdhury and Sharma [2009] show that combining multimodels that have least covariance at the first step seem to be more effective in developing better multimodel predictions. Given this, it seems obvious that climatology will have the smallest covariance with individual model predictions, thereby two-step combination being very effective in reducing the $\overline{RPS}$ of multimodels.

Figure 2.6 and Table 2.4, which are similar to Figure 2.5 and Table 2.3, summarize the multimodel combination results for temperature. From Figure 2.6, it is very clear that all the multimodels perform better than ECHAM4.5 in predicting temperature. Among the multimodels, MM-1 proposed in the study, perform better than the rest of the multimodels. From Table 2.4, we also infer that the performance of MM-OS is also equally good in predicting winter temperature. Comparing the performance of MM-1 and MM-2, we infer that in 48 (24) grid points MM-OS (MM-1)'s $\overline{RPS}$ is statistically significant than the $\overline{RPS}$ of MM-1(MM-OS) indicating that combining models purely based on their long-term skill seem to be a good strategy in multimodel combination. However, among these grid points, if we drop grid points with $\overline{RPSS}$ of both models being negative, then we result with 27(19) grid points that show the $\overline{RPS}$ of MM-OS (MM-1) being statistically significant than the $\overline{RPS}$ of MM-1 (MM-OS). This indicates MM-1's better performance is more in grid points exhibiting positive $\overline{RPSS}$. From Tables 2.3 and 2.4, we also understand that the

improvements in predicting winter temperature from multimodel combination is more in comparison to the improvements in predicting winter precipitation. In Section 2.3.5, we discuss in detail improvements resulting from multimodel combination from a regional perspective over the continental United States particularly for grid points that exhibit positive $\overline{RPSS}$.

It is important to note that Figures 2.5 and 2.6 show spatial correlation in the reported percentiles of the test statistic. This is because we resample RPS_t from Model A and B available at each grid point to construct the null distribution. Performing hypothesis tests with spatially correlated forecast error metric would reduce the effective number of independent samples [Wilks, 1997; Hamill, 1999]. One way to overcome the spatially correlated prediction error metric is to spatially average the verification measure over a region and perform the hypothesis tests over the spatially averaged verification measure. However, we felt that such an approach would require first identification of homogenous regions for spatial averaging of the error metric, so it is not pursued here.

2.3.3 Comparison of Forecast Reliability between Multimodels and Individual Models

Rank probability score just quantifies the squared error in forecasted cumulative probabilities for categorical forecasts. However, it does not provide information on how the forecasted probabilities for a particular category correspond to their observed frequencies. For this purpose, this section compares the reliability, resolution of multimodel predictions with the reliability and resolution of individual model predictions. Reliability diagrams provide information on the correspondence between the forecasted probabilities for a

particular category (e.g., above-normal, normal and below-normal categories) and how frequently that category is observed under the issued forecasted probability [Wilks, 1995]. For instance, if we forecast the probability of below-normal category as 0.9 over '*n1*' years, then we expect the actual outcome to fall under below-normal category for *0.9*n1* times over the entire forecast verification period.

Figures 2.7a and 2.7b compare the reliabilities of three multimodels (MM-2, MM-4, MM-OS) with the reliabilities of ECHAM4.5 and CCM3v6 in predicting precipitation for below-normal and above-normal categories respectively. Similarly, Figures 2.8a and 2.8b compare the reliabilities of MM-1, MM-3 and MM-OS with the reliabilities of ECHAM4.5 and CCM3v6 in predicting temperature under below normal and above normal categories respectively. We did not consider MM-P since it did not reduce the $\overline{RPS}$ over many grid points in comparison to the rest of the multimodels in predicting precipitation and temperature (Tables 2.3 and 2.4).

For developing reliability plots, the tercile probabilities for 46 years under each category are grouped at an interval of 0.1 over all grid points (46* 192 = 8832 forecasts for a tercile category for each model). The observed category is also recorded using which the observed relative frequency under each forecasted probability is calculated for each tercile category. Inset in each reliability plots show the attribute diagram indicating the logarithm of the number of forecasts that fell under each forecast probability bin for a given model. Figures 2.7 and 2.8 also show the perfect (diagonal) reliability line with one to one correspondence between forecasted probability and its observed relative frequency.

From Figures 2.7 and 2.8, we observe that the selected multimodels improve the reliability of forecasts showing better correspondence between forecasted probabilities and their observed relative frequencies. The basic reason multimodel predictions ensure better reliability is by reducing the overconfidence of individual model predictions. This could be understood from the attribute diagram which clearly shows reduction in the number of predictions with high forecast probabilities (0.8-1) under individual models (ECHAM4.5 and CCM3v6). These findings are in line with earlier studies [Weigel et al. 2008]. On the other hand, multimodels show increase in the number of predictions under moderate forecast probabilities (0.4-0.7), thereby resulting in the reduction of false alarms. Similarly, under low forecast probabilities, individual models seem to be less reliable indicating higher frequency of occurrence, whereas multimodels have better reliability resulting in a reduction of missed targets. To better quantify the information in Figures 2.7 and 2.8, we summarize the ability of a model to predict a particular tercile category using average Brier Score ($\overline{BS}$) [Wilks, 1995].

Brier Score, summarizing the squared error in categorical forecast probabilities, can be decomposed into reliability, resolution and uncertainty [Wilks, 1995]. For $\overline{BS}$ to be close to zero, it is important that the reliability term should be close to zero and resolution term should be large. Figures 2.9a (2.9c) and 2.9b (2.9d) provide the reliability, resolution and $\overline{BS}$ for ECHAM4.5, CCM3v6 and all the six multimodels in predicting below normal and above normal categories of precipitation (temperature). From Figures 2.9a and 2.9b, we infer that all multimodels have smaller reliability score in comparison to the reliability scores of

individual models under both tercile categories, thereby contributing to the reduction in $\overline{BS}$. Among the multimodels, MM-2 has the smallest reliability score than the rest of the five multimodels in predicting precipitation. In terms of resolution, ECHAM4.5 has larger resolution score than the resolution scores of CCM3v6 and other multimodels in predicting precipitation. Among the multimodels, we clearly see that, MM-2 has the largest resolution score which leads to MM-2 and MM-4 having the lowest $\overline{BS}$ in predicting precipitation. Similarly, from Figures 2.9c and 2.9d, we infer that MM-1 and MM-OS have the smallest $\overline{BS}$ which results primarily from smaller reliability score and larger resolution score in predicting temperature under below-normal and above-normal categories.

To summarize the multimodel schemes, MM2 and MM4 (MM1 and MM-OS), perform better than individual models as well as over the rest of the multimodels in predicting winter precipitation (temperature) over the continental United States (Figures 2.9a - 2.9d). The proposed multimodel combination schemes in this study, MM-2 and MM-4, have the lowest $\overline{BS}$ among all the models in predicting precipitation, whereas MM-1 also perform equally well (in comparison to the best multimodel MM-OS) in predicting winter temperature. To understand why the multimodel combination schemes proposed in Figure 2.2 result in improved predictions, we plot weights (w_k^m) obtained for each multimodel schemes and analyze how they vary conditioned on ENSO state in the next section.

2.3.4 Analysis of Weights

Figure 2.10 shows box plots of the ratio of weights $(w_{t,k}^{m} \mid MM\text{-}1)$ for each model under MM-1 scheme to the weights $(w^{m} \mid MM\text{-}OS)$ obtained for each model based on MM-OS scheme in predicting temperature. The weight ratios plotted in Figure 2.10 are grouped under two categories namely grid points exhibiting significant skill under El Nino years (Figure 2.10a) and La Nina years (Figure 2.10b) by ECHAM4.5. This resulted in a total of 20 and 43 grid points exhibiting significant skill (based on the correlation between the observed temperature and the ensemble mean) under El Nino and La Nina years respectively. The weights $(w_{t,k}^{m} \mid MM-1, w^{m} \mid MM-OS)$ are pooled over 46 years and the computed weight ratios for these grid points are shown as separate box plots (as columns) conditioned on the ENSO state (El Nino, La Nina and Neutral). Weight ratio above 1 indicates that MM-1 weights for a given model are higher than the weights assigned by the MM-OS scheme.

From Figure 2.10a, which shows the weight ratios for grid points showing significant skill during El Nino years, we can clearly see that weight ratios are greater than 1 for ECHAM4.5 around 25% of the time and lesser than 1 for climatology around 85% of the time during El Nino conditions (first column in Figure 2.10a). However, the weight ratios for GFDL model is higher than 1 (around 75% of the time) indicating GFDL's better performance during El Nino years for grid points considered in Figure 2.10a. This implies that if a particular GCM performs well during El Nino years, then higher weights are assigned for that GCM during those conditions in comparison to the weights based on long-term skill of the model (MM-OS). Further, the weights assigned for climatology under MM-1

scheme is lesser since all GCMs have good skill in predicting temperature during El Nino conditions.

On the other hand, during neutral conditions (last column in Figure 2.10a), the weight ratios are substantially lesser than 1 for both ECHAM4.5 and GFDL, whereas the weight ratios are greater than 1 for climatology (around 90% of the time). Under La Nina conditions for grid points exhibiting significant skill during El Nino years (Figure 2.10a, middle column), we can clearly infer that the weights for ECHAM4.5 from MM-1 schemes are higher in comparison to the weights for ECHAM4.5 received from MM-OS scheme during La Nina years. This analysis again confirms our argument that if GCM's skill is poor during certain predictor conditions, then it is better to consider climatology as the best information available. Our multimodel combination algorithm shown in Figure 2.2 basically implements this by evaluating the models' skill contingent on the dominant predictor state and assigns higher weight for the model/climatology that exhibits better skill during those predictor conditions.

Figure 2.10b shows similar results for grid points exhibiting significant skill in predicting DJF temperature by ECHAM4.5 during La Nina years. From Figure 2.10b, we can clearly see that the weight ratios for ECHAM4.5 are mostly lesser than 1 with ECPC's ratio being higher than 1 during El Nino conditions (Figure 2.10b, leftmost column). On the other hand, during La Nina conditions, ECHAM4.5 weight ratios are greater than 1 (around 85% of the time) which forces weight ratios for climatology being substantially lesser than 1 (around 60% of the time) (Figure 2.10b, middle column). Under neutral conditions (Figure 2.10b, rightmost column) with none of the models exhibiting significant skill, the weights

assigned by MM-1 scheme for climatology are higher than the weights assigned by MM-OS scheme (around 90% of neutral conditions in those grid points). Similar analysis on weights under MM-2 in predicting precipitation showed the same pattern (figure not shown). Thus, our study clearly shows that combining multiple climate models by evaluating them contingent on the predictor state is a potential strategy in improving the skill of multimodel climate forecasts.

2.3.5 Discussion

The main advantage of the proposed multimodel combination technique is in its ability to evaluate GCMs' skill contingent on the predictor state and assign higher weights for GCMs that perform better during those (predictor) conditions. Similarly, the methodology could also assign higher weights to climatology if all the GCMs have limited skill under that conditioning state. On the other hand, pursuing multimodel combination purely based on the overall skill (MM-OS) could result in higher weights for GCMs under conditions during which the model might exhibit poor/ limited skill. Further, the proposed approach combines both models' skill (as quantified by MSE) and optimization (choosing the number of neighbors, K_t, in MM-3 and MM-4 under two-deep cross-validation) to estimate weights as opposed to obtaining model weights purely based on optimization [e.g., Rajagopalan et al.2002], which could end up choosing one or two models alone in multimodel combination. To overcome these difficulties, Robertson et al. [2004] proposed two-step combination. Analysis of weights (Figure 2.10) shows clearly that model weights are linked to their skill

with GCMs weights being higher during ENSO conditions and climatology receiving higher weights during neutral conditions.

Figure 2.11 shows the skill, expressed as $\overline{RPSS}$, in predicting DJF precipitation (2.11a) and temperature (2.11b) with each grid point's $\overline{RPSS}$ being indicated by the best performing individual model or the multimodel. Table 2.5 shows the number of grid points with each individual model and multimodel having the highest $\overline{RPSS}$ over 192 grid points shown in Figure 2.11. Figure 2.11 and Table 2.5 summarize the performance of models (individual model and multimodels) only if the $\overline{RPSS}$ of at least one model is greater than zero at a given grid point. Thus, if $\overline{RPSS}$ of all individual models and multimodels are lesser than zero, then climatology provides the best information for those grid points. From Figure 2.11 and Table 2.5, we can clearly see that multimodels proposed in the study (MM-1, MM-2, MM-3 and MM-4) perform better than the individual models and over the existing multimodel combination techniques (MM-P and MM-OS). Among the multimodels, MM-4 seems to be best performing multimodel in predicting precipitation and temperature, whereas COLA and ECHAM4.5 seem to be the best performing individual models in predicting precipitation and temperature respectively. Comparing Figures 2.11a and 2.11b, we infer that prediction of temperature seems to benefit more from multimodel combination in comparison to the improvements resulting in precipitation.

From Figure 2.11a, we understand that the improvements resulting from multimodel combination in predicting DJF precipitation predominantly lies over the Southern United States as well as over certain grid points in Midwest and Northeast. In the case of

temperature (Figure 2.11b), with the exception of Midwest, we infer that $\overline{RPSS}$ is greater than zero for most of the regions indicating better skill (in comparison to climatology) demonstrated by both individual models and multimodels. From Figure 2.11, we can understand that there is a significant improvement in $\overline{RPSS}$ of multimodel proposed in the study (shown as open circles) over Southeast, Southwest regions of the United States and over Northwest Mexico. Figures 2.5 and 2.6 also show similar spatial structure with the $\overline{RPS}$ of multimodels being statistically significant than the $\overline{RPS}$ of ECHAM4.5. The reason for this improved performance over these regions is primarily due to strong correlation between ensemble mean of the individual models with the observed precipitation/temperature under ENSO conditions.

It is important to note that this study has employed historical simulations of precipitation and temperature from AGCMs to demonstrate the utility of multimodel combination algorithm presented in Figure 2.2. Historical simulations from AGCMs which employ observed SSTs as boundary conditions typically overestimate the real predictive skill [Goddard and Mason, 2002; Sankarasubramanian et al., 2008]. Further, to apply the proposed methodology in a forecasting context, one may have to use the forecasted Nino3.4 from multiple CGCMs as the conditioning variable [Tippet and Barnston, 2008]. Given the peak ENSO activity typically coincides during DJF season, 3-month ahead multimodel forecasts of ENSO indices issued in December exhibit very high skill with correlations ranging above 0.8 and root mean square error around 0.2-0.4° C [Jin et al. 2008; Weisheimer et al. 2009]. Due to this, the identified similar DJF ENSO conditions using forecasted multimodel mean

of DJF Nino3.4 could slightly differ from the identified conditions using observed DJF Nino3.4. More importantly, employing retrospective forecasts from AGCMs forced with forecasted SSTs could result with reduced skill from the proposed multimodel scheme if the skills of retrospective forecasts from AGCMs are better than that of climatology under the conditioned state. But, if the skills of retrospective forecasts from AGCMs are poorer than that of climatology (which is highly likely based on Figures 2.3 and 2.4), then we expect the proposed multimodel scheme is likely to be more beneficial in replacing AGCMs forecasts with climatology. The study in chapter 3 evaluates the utility of the proposed methodology in combining real-time precipitation forecasts from CGCMs contingent on the forecasted DJF Nino3.4 state.

2.4 Summary and Conclusions

A methodology to combine multiple GCMs is proposed and evaluated for predicting winter precipitation and temperature over the United States. The methodology assigns weights for each GCM by evaluating their skill, quantified by Mean Square Error, over similar predictor conditions. Considering Nino3.4 as the primary predictor influencing the winter precipitation and temperature [Quan et al. 2006], the study combines seven GCMs with climatological ensembles to develop multimodel predictions over the continental United States. Totally, six different multimodel schemes are considered with their performance being compared with individual models based on various verification measures such as rank probability skill score, reliability and resolution scores and brier score. The improvements

resulting (from reduction in $\overline{RPS}$) from multimodel combination over individual model is also tested through a rigorous nonparametric hypothesis testing based on resampling.

The study clearly shows that the proposed multimodel combination algorithm perform better, in terms of improving the $\overline{RPSS}$, than individual models and over multimodel combinations based on pooling and long-term skill. Further, the proposed multimodel combination methodology also improves the reliability and resolution of tercile probabilities resulting with reduced Brier scores. The improved reliability of multimodel predictions primarily arises from reducing the overconfidence of individual model predictions, which in turn results with reduced number of false alarms and missed targets in categorical forecasts. Analysis of weights also show that the proposed methodology assigns higher (lower) weights for GCMs and lesser (higher) weights for climatology during anomalous (neutral) ENSO conditions in grid points. These analyses show that the combining multimodels contingent on the dominant predictor state is an attractive strategy in improving the skill of multimodel forecasts.

2.5 References

Anderson J. L., and Coauthors, 2005: The new GFDL global atmospehre and land model AM2-LM2: Evaluation with prescribed SST simulations. *J. Climate*, **17**, 4641-4673.

Bacmeister J., P. J. Pegion, S. D. Schubert, and M. J. Suarez, 2000: Atlas of Seasonal Means Simulated by the NSIPP 1 Atmospheric GCM. *NASA/TM-2000-* 104505, **17**, 194 pp.

Barnston A. G., S. J. Mason, L. Goddard, D. G. Dewitt, and S. E. Zebiak, 2003: Multimodel ensembling in seasonal climate forecasting at IRI. *Bull. Amer. Meteor. Soc.,* **84**, 1783–1796.

Bobko P, 1995: Correlation and Regression: Principles and Applications for Industrial/ Organizational Psychology and Management. New York: McGraw Hill, Inc.

Branković Č., and T. N. Palmer, 2000: Seasonal skill and predictability of ECMWF PROVOST ensembles. *Quart. J. Roy. Meteor. Soc.*, **126**, 2035–2067.

Bröcker, J., and L.A. Smith: Scoring Probabilistic Forecasts; On the Importance of Being Proper. Weather and Forecasting, 22, (2), 2007.

Chowdhury S., and A. Sharma, 2009: Long-Range Niño-3.4 Predictions Using Pairwise Dynamic Combinations of Multiple Models. *J. Climate*, **22**, 793–805.

DelSole T., 2007: A Bayesian framework for multimodel regression. *J. Climate,* **20**, 2810–2826.

Devineni N., A. Sankarasubramanian, and S. Ghosh, 2008: Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations, *Water Resour. Res.*, **44**, W09404, doi:10.1029/2006WR005855.

Doblas-Reyes F.J., M.Deque, and J.P.Piedelievre, 2000: Multimodel spread and probabilistic seasonal forecasts in PROVOST. *Quart. J. Roy. Meteor. Soc.*, **126(567)**:2069-2087.

Doblas-Reyes F. J., R. Hagedorn, and R. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination. *Tellus*, **57A**, 234–252.

Goddard L., A. G. Barnston, and S. J. Mason, 2003: Evaluation of the IRI's "net assessment" seasonal climate forecasts: 1997– 2001. *Bull. Amer. Meteor. Soc.*, **84**, 1761–1781.

Hagedorn R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus A*, **57**, 219–233, doi:10.1111/j.1600–0870.2005.00103.x.

Hamill T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167.

Hoeting J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky, 1999: Bayesian model averaging: A tutorial, *Stat. Science*, **14**, 382-401.

Jin E. K., J. L. Kinter, B. Wang, C. K. Park, I. S. Kang, B. P. Kirtman, J. S. Kug, A. Kumar, J. J. Luo, J. Schemm, J. Shukla, and T. Yamagata, 2008: Current status of ENSO prediction skill in coupled ocean-atmosphere models, *Climate Dyn*, ***31***, 647-664.

Kanamitsu M., and K. C. Mo, 2003: Dynamical effect of land surface processes on summer precipitation over the southwestern United States. *J. Climate*, **16**, 496–509.

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991, *Journal of Geophysical Research*, **103**, 18,567-18,589.

Kiehl J.T., J.J.Hack, G.B.Bonan, B.A.Boville, D.L.Williamson, and P.J.Rasch, 1998: The National Center for Atmospheric Research Community Climate Model: CCM3. *J. Climate*, **11(6)**:1131-1149.

Krishnamurti T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from a multi-model superensemble. *Science*, **286**, 1548–1550.

Luo L., E. F. Wood, and M. Pan, 2007: Bayesian Merging of Multiple Climate Model Forecast for Seasonal Hydrological Predictions. *J. Geophys. Res.*, **112**, D10102, doi: 10.1029/2006JD007655.

Mason S.J., and G.M.Mimmack, 2002: Comparison of some statistical methods of probabilistic forecasting of ENSO. *J. Climate*, **15**:8-29.

New M., M. Hulme, and P. D. Jones, 1999: Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate*, **12**, 829–856.

New M., M. Hulme, and P. D. Jones, 2000: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238.

Palmer T.N., C.Brankovic, D.S.Richardson, 2000: A probability and decision-model analysis of PROVOST seasonal multimodel ensemble integrations. *Quart. J. Roy. Meteor. Soc.*, **126(567),**2013-2033.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Delecluse, M. Deque, E. Diez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J. F. Gueremy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J. M. Terres, and M. C. Thomson, 2004: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER), *Bulletin of the American Meteorological Society,* **85,** 853-872.

Quan X., M. Hoerling, J. Whitaker, G. Bates, and T. Xu, 2006: Diagnosing sources of US seasonal forecast skill, *J. Climate*, **19,** 3279-3293.

Rajagopalan B., U. Lall, and S. E. Zebiak, 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.*, **130,** 1792–1811.

Robertson A. W., U. Lall, S. E. Zebiak, and L. Goddard, 2004: Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Wea. Rev.*, **132**, 2732–2744.

Rodwell, M. J., and F. J. Doblas-Reyes (2006), Medium-range, monthly, and seasonal prediction for Europe and the use of forecast information, *Journal of Climate*, **19**, 6025-6046.

Roeckner E., and Coauthors, 1996: The atmospheric general circulation model ECHAM4: Model description and simulation of present-day climate. *Max-Planck-Institut fur Meteorologie Rep*. **218**, Hamburg, Germany, 90 pp.

Saha S., S. Nadiga, C. Thiaw, and J. Wang, 2006: The NCEP Climate Forecast System. *J. Climate,* **19**, 3483-3517.

Schneider E. K., 2002: Understanding differences between the equatorial Pacific as simulated by two coupled GCMs. *J. Climate*, **15**, 449-469.

Sankarasubramanian, A., U.Lall, and S.Espuneva, 2008: Role of Retrospective Forecasts of GCM Forced with Persisted SST anomalies in Operational Streamflow Forecasts Development. *Journal of Hydrometeorology*, **9(2)**, 212-227.

Shukla J., J. Anderson, D. Baumhefner, C. Brankovic, Y. Chang, E. Kalnay, L. Marx, T. Palmer, D. Paolino, J. Ploshay, S. Schubert, D. Straus, M. Suarez, and J. Tribbia, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593-2606.

Stephenson D. B., C. A. S. Coelho, F. J. Doblas-Reyes, and M. Malmaseda, 2005: Forecast assimilation: A unified framework for the combination of multi-model weather and climate predictions. *Tellus A*, **57**, 253–264.

Stone M., 1974: Cross-validatory choice and assessment of statistical predictions (with discussion). *J. Roy. Stat. Soc.*, **36**, 111–147.

Tippett M. K., and A. G. Barnston , 2008: Skill of Multimodel ENSO Probability Forecasts, *Mon. Wea. Rev.*, **136**, 3933-3946.

Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores, *Monthly Weather Review*, **135**, 118-124.

Weigel, A. P., M. A. Liniger, and C. Appenzeller , 2008: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts?, *Quarterly Journal of the Royal Meteorological Society*, **134**, 241-260.

Weisheimer, A., F. J. Doblas-Reyes, T. N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel , 2009: ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett*, 36, L21711, doi:10.1029/2009GL040896.

Wilks D. S., 1995: *Statistical Methods in the Atmospheric Science*. Academic Press, 467 pp.

Wilks D.S., 1997: Resamping hypothesis tests for auto-correlated fields, *J. Climate*, **10**, 66-82.

Table 2.1: Details of atmospheric GCMs considered for the study. All models span from longitude (123.75W to 66.25W) and latitude (25N to 45N) resulting in a total of 192 grid points. Historical simulations of winter (DJF) precipitation and temperature from the seven GCMs are considered for multimodel combination.

Historical Simulations			
Model	**Source**	**Ensemble Size**	**Reference**
ECHAM4.5	Max-Plank Institute	85	Roeckner et al. (1996)
CCM3v6	National Center for Atmospheric Research, NCAR	24	Kiehl et al. (1998)
COLA	Center for Ocean-Land-Atmosphere Studies	10	Schneider (2002)
GFDL, AM2p12b	Princeton University	10	Anderson et al. (2005)
ECPC	Scripps Institution of Oceanography, UC, San Diego	10	Kanamitsu et al. (2003)
NCEP	NOAA	10	Saha et al. (2006)
NSIPP-1	NASA GSFC	9	Bacmeister et al. (2000)

Table 2.2: List of multimodel combinations considered for the study.

Multimodel Indices/Schemes	**Brief Description**
MM-1	Individual Models + Climatology in one step *with fixed number of neighbors contingent on ENSO state*
MM-2	Individual Model + Climatology combination in the first step and the resulting model outputs combined at the second step *with fixed number of neighbors contingent on ENSO state*
MM-3	Individual Models + Climatology in one step *using optimized neighbors obtained by two-deep cross-validation*
MM-4	Individual Model + Climatology combination in the first step and the resulting model outputs combined at the second step *using optimized neighbors obtained by two-deep cross-validation*
MM-P	Multimodel combination using pooled ensembles
MM-OS	Multimodel combination using weights based on the overall skill in the calibration period.

Table 2.3: Number of grid points showing significant difference in $\overline{RPS}$ in predicting precipitation based on the hypothesis testing between ECHAM4.5 and various multimodel schemes given in Table 2.2. Entries in the upper (lower) triangle of the table summarize the number of grid points having the percentile value of the test statistic $\overline{RPS}^{A} - \overline{RPS}^{B}$ between 0.9-1 (0-0.1) in the resampled null distribution for hypothesis testing between Model A and Model B. For values in the upper (lower) triangle, Model A (Model B) is indicated by its row entry and Model B (Model A) is indicated by its column entry. The best-performing model in terms of increased number of significant grid points is underlined by its column entry.

Models	**ECHAM4.5**	**MM-1**	**MM-2**	**MM-3**	**MM-4**	**MM-P**	**MM-OS**
ECHAM4.5		39	42	36	41	53	40
MM-1	5		50	22	51	24	59
MM-2	5	6		7	25	11	9
MM-3	10	65	57		64	43	81
MM-4	5	16	52	6		11	16
MM-P	6	20	33	19	37		20
MM-OS	8	17	35	16	43	17	

Table 2.4: Number of grid points showing significant difference in $\overline{RPS}$ in predicting temperature based on the hypothesis testing between ECHAM4.5 and various multimodel schemes given in Table 2.2. Entries in the upper (lower) triangle of the table summarize the number of grid points having the percentile value of the test statistic $\overline{RPS}^{A} - \overline{RPS}^{B}$ between 0.9-1 (0-0.1) in the resampled null distribution for hypothesis testing between Model A and Model B. For values in the upper (lower) triangle, Model A (Model B) is indicated by its row entry and Model B (Model A) is indicated by its column entry. The best-performing model in terms of increased number of significant grid points is underlined by its column entry.

Models	**ECHAM4.5**	**MM-1**	**MM-2**	**MM-3**	**MM-4**	**MM-P**	**MM-OS**
ECHAM4.5		<u>70</u>	<u>48</u>	<u>65</u>	<u>59</u>	<u>81</u>	<u>77</u>
MM-1	5		6	27	17	3	<u>48</u>
MM-2	6	<u>39</u>		<u>28</u>	33	<u>19</u>	<u>41</u>
MM-3	7	<u>51</u>	9		23	8	<u>58</u>
MM-4	9	<u>44</u>	<u>49</u>	<u>36</u>		24	<u>46</u>
MM-P	5	<u>42</u>	18	<u>34</u>	<u>31</u>		<u>52</u>
MM-OS	5	24	4	17	9	3	

Table 2.5: Number of grid points with each individual GCM and multimodels having the highest $\overline{RPSS}$ in predicting winter precipitation and temperature.

Models	**Precipitation**	**Temperature**
ECHAM4.5	8	32
CCM3v6	9	6
COLA	11	8
ECPC	0	2
GFDL	1	3
NCEP	3	3
NSIPP-1	6	0
MM-1	4	8
MM-2	12	20
MM-3	9	18
MM-4	27	31
MM-P	3	10
MM-OS	4	15

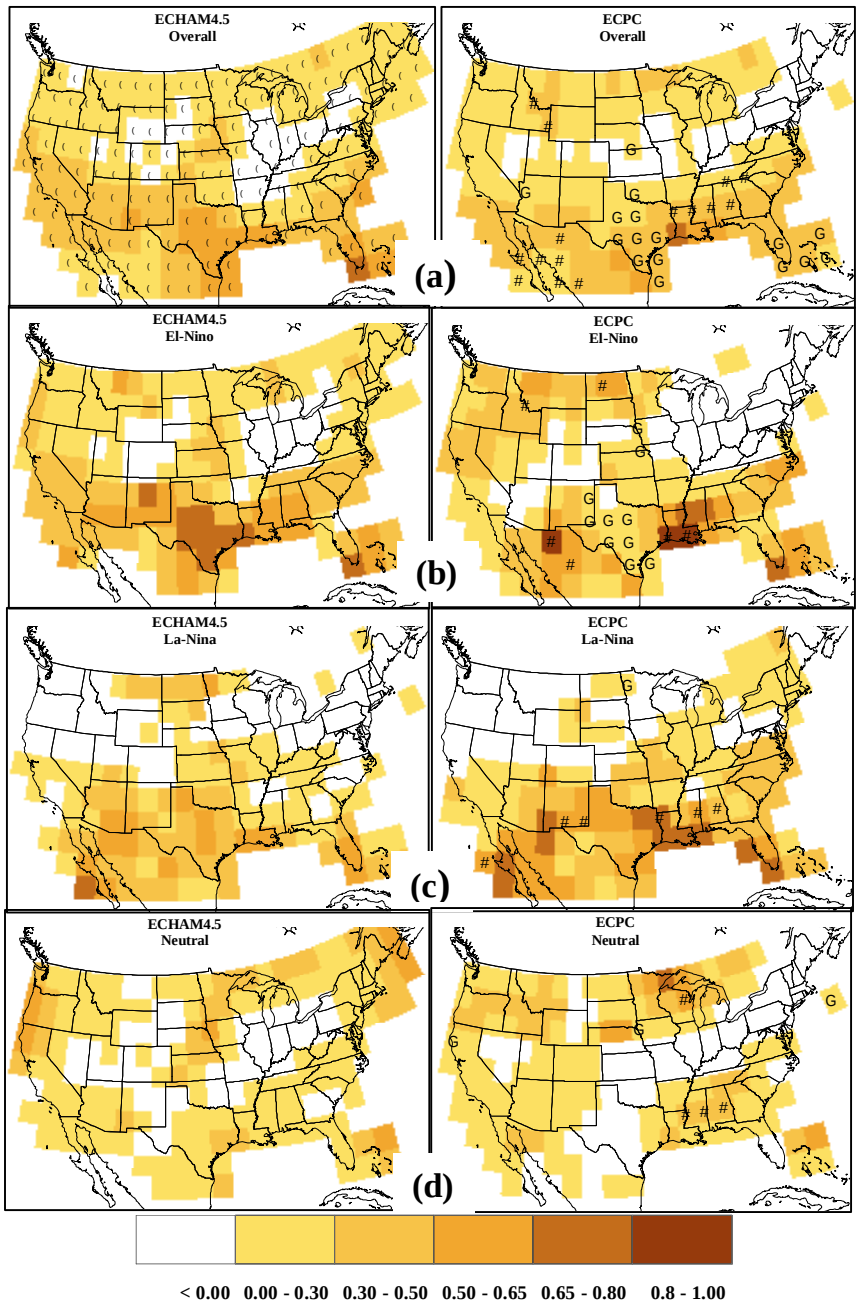

Figure 2.1: Skill, expressed as correlation between ensemble mean of the GCM and observed precipitation, in simulating the DJF winter precipitation by two GCMs, ECHAM 4.5 (left) and ECPC (right), over the entire period of record (Figure 2.1a), under El-Nino (Figure 2.1b), La Nina (Figure 2.1c) and Neutral conditions (Figure 2.1d). Figure on the right under each category shows plus (triangle) sign which indicates the correlation between DJF precipitation and the ensemble mean of ECHAM4.5 (ECPC) is statistically higher than the correlation between the DJF precipitation and the ensemble mean of ECPC (ECHAM4.5) for that category.

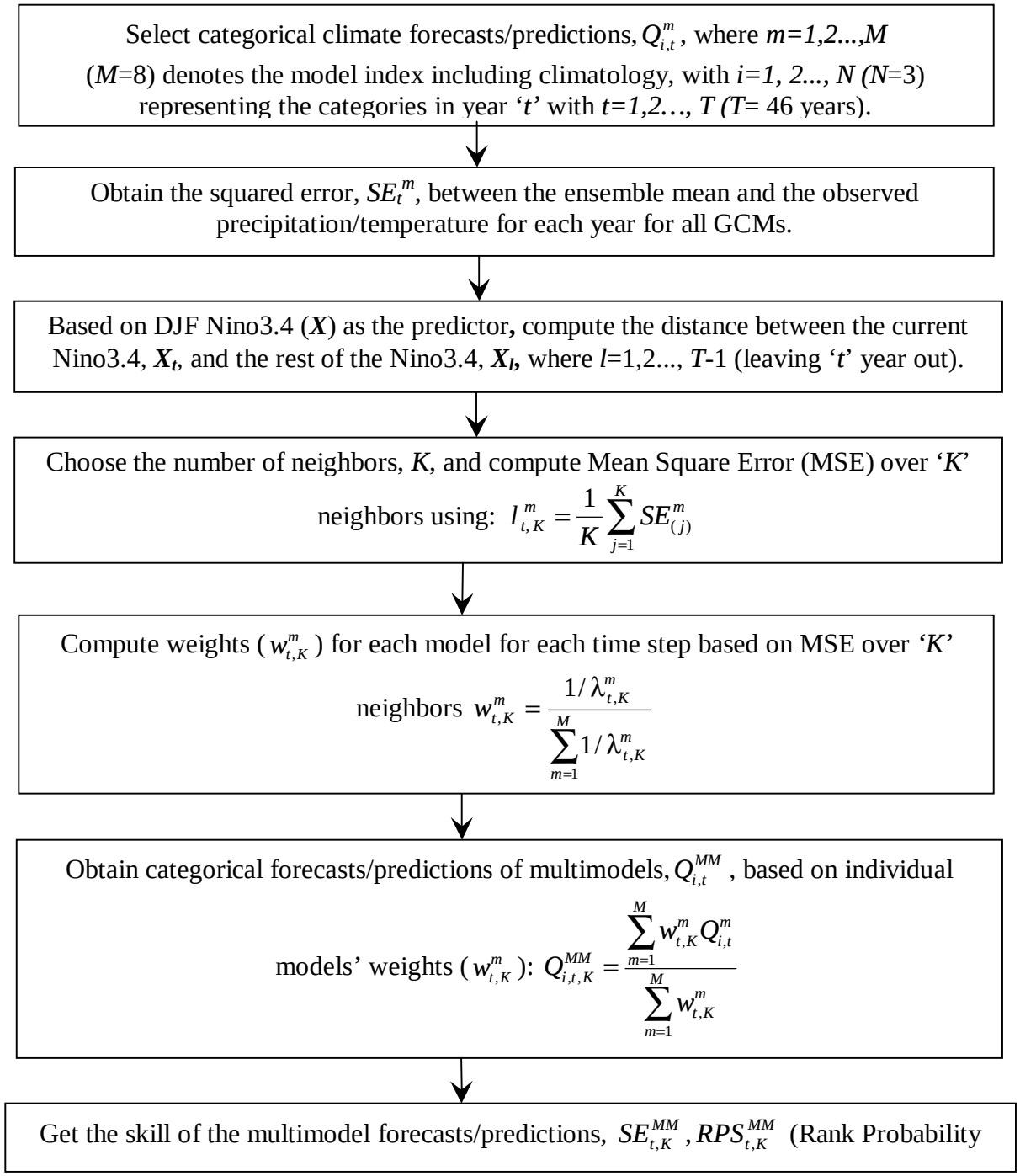

Figure 2.2: Flowchart of the multimodel combination algorithm employed in the study (modified from Devineni et al. 2008).

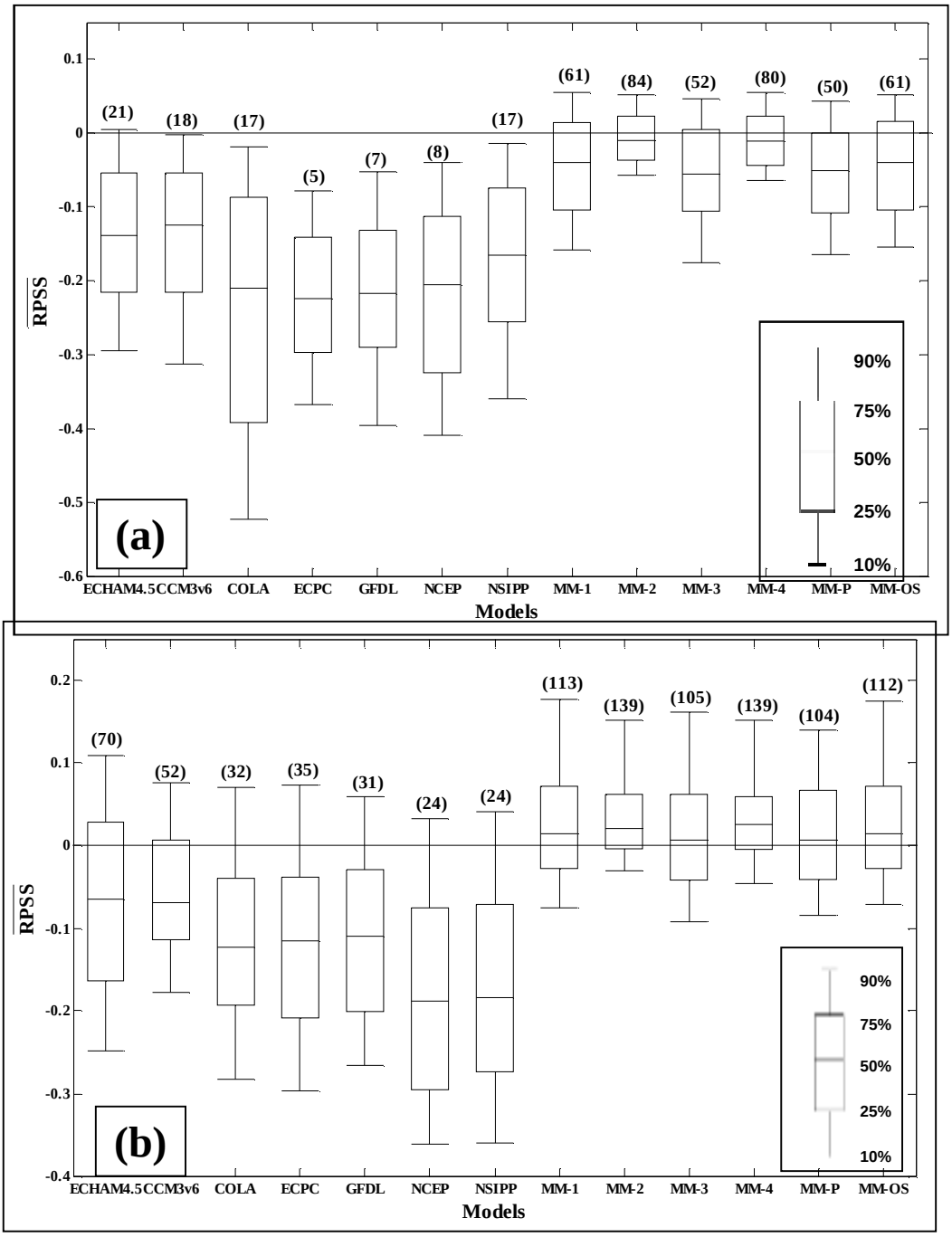

Figure 2.3: Box plots of $\overline{RPSS}$ in predicting winter precipitation (2.3a) and temperature (2.3b) for individual GCMs and various multimodel schemes given in Table 2.2. Numbers in parenthesis above each box plot indicate the number of grid points having $\overline{RPSS}$ greater than zero.

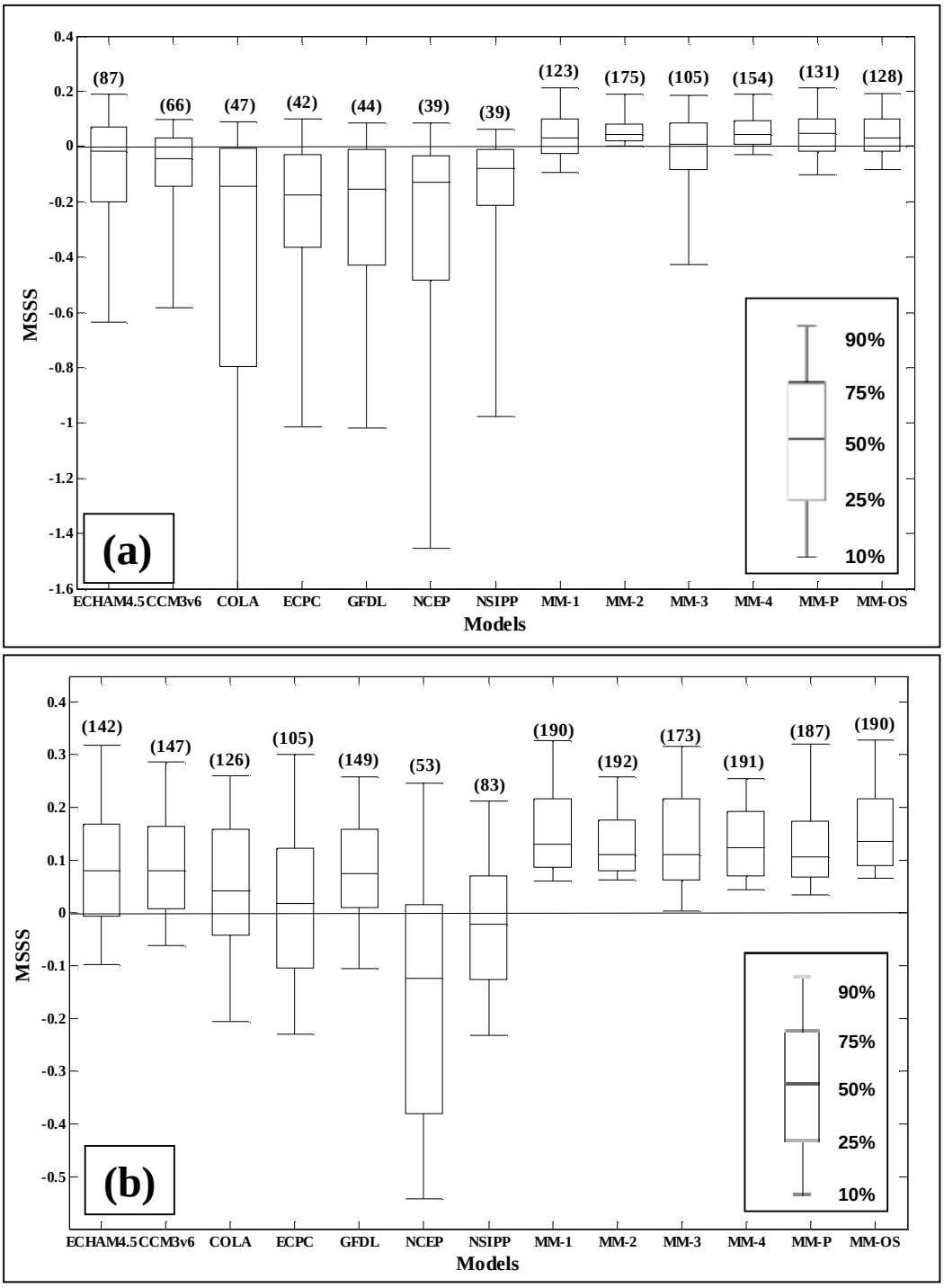

Figure 2.4: Box plots of MSSS in predicting winter precipitation (2.4a) and temperature (2.4b) for individual GCMs and various multimodel schemes given in Table 2.2. Numbers in parenthesis above each box plot indicate the number of grid points having MSSS greater than zero.

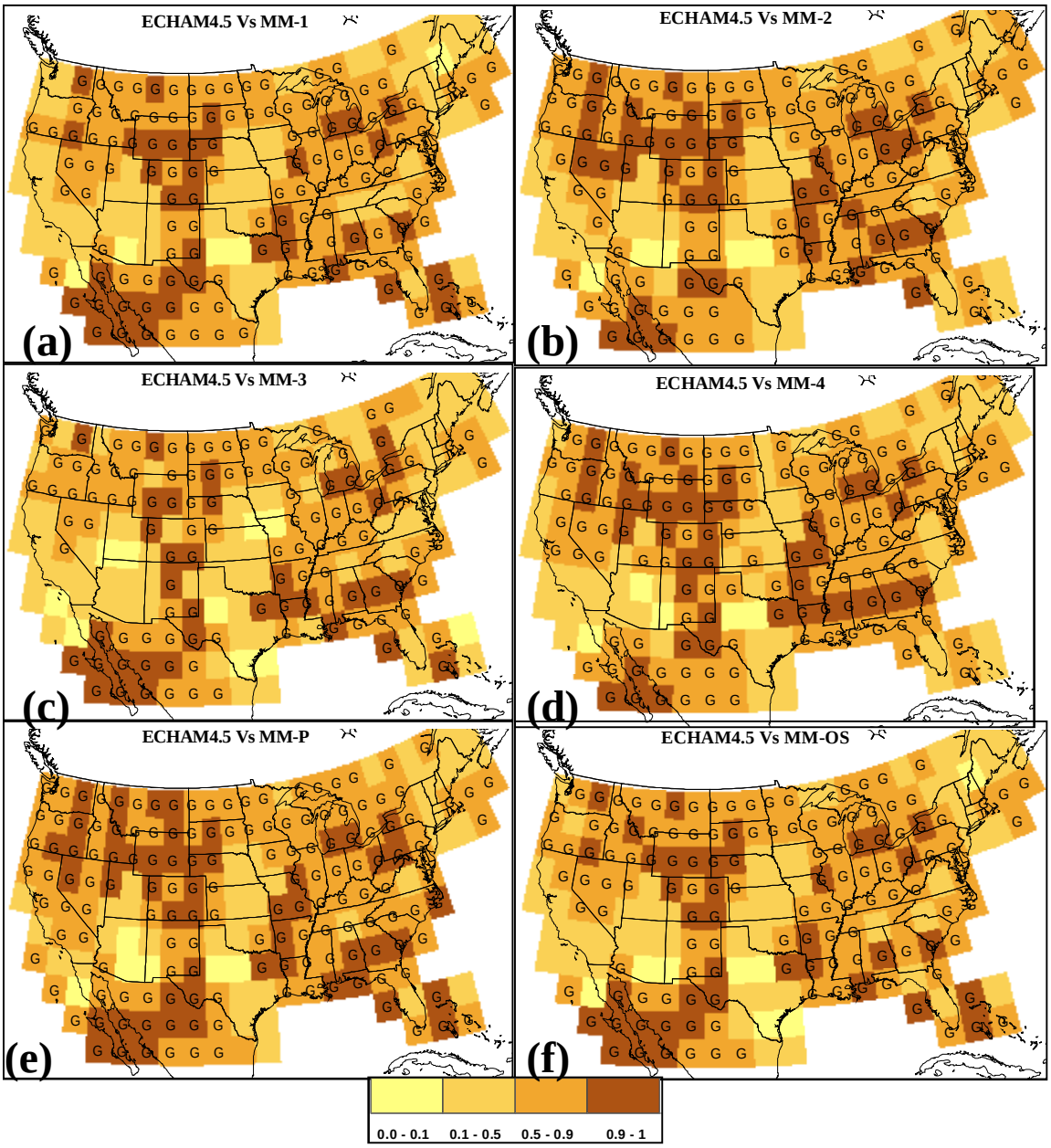

Figure 2.5: Performance comparison of multimodels with the best individual model, ECHAM4.5, in predicting U.S winter precipitation. The background color indicates the percentile of the test statistic ($\overline{RPS}^{ECHAM\,4.5} - \overline{RPS}^{MM}$) obtained from the resampled null distribution that represents $\overline{RPS}^{ECHAM\,4.5} = \overline{RPS}^{MM}$. A lower (higher) value of the percentiles from the test statistic indicates ECHAM4.5 (multimodel) performs better than multimodel (ECHAM4.5). A plus (blank) sign indicates that $\overline{RPS}$ from multimodel (ECHAM4.5) is lesser than the $\overline{RPS}$ of ECHAM4.5 (multimodel).

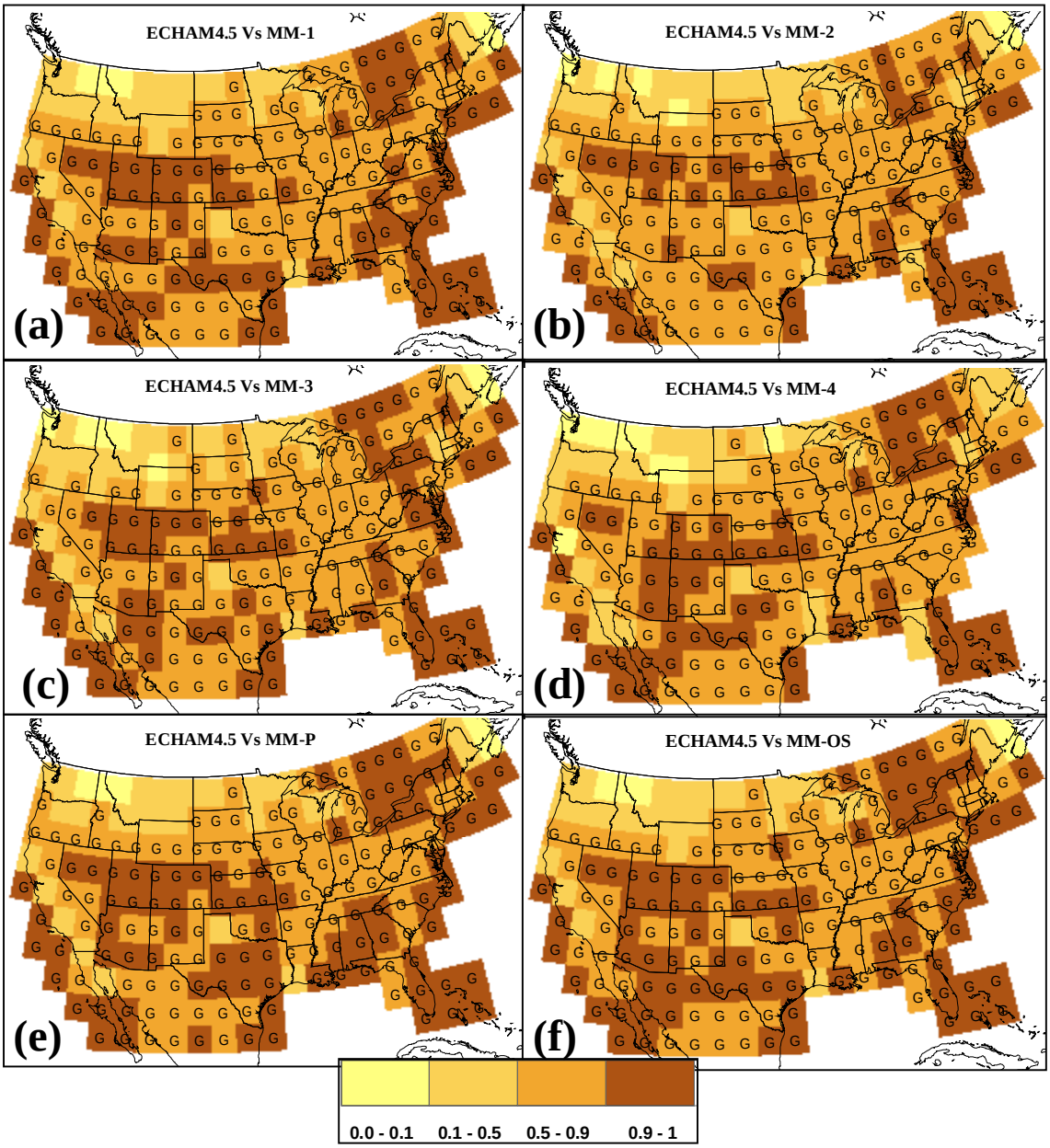

Figure 2.6: Performance comparison of Multimodels with the best individual model, ECHAM4.5, in predicting U.S winter temperature. The background color indicates the percentile of the test statistic ($\overline{RPS}^{ECHAM4.5} - \overline{RPS}^{MM}$) obtained from the resampled null distribution that represents $\overline{RPS}^{ECHAM4.5} = \overline{RPS}^{MM}$. A lower (higher) value of the percentiles from the test statistic indicates ECHAM4.5 (multimodel) performs better than multimodel (ECHAM4.5). A plus (blank) sign indicates that $\overline{RPS}$ from multimodel (ECHAM4.5) is lesser than the $\overline{RPS}$ of ECHAM4.5 (multimodel).

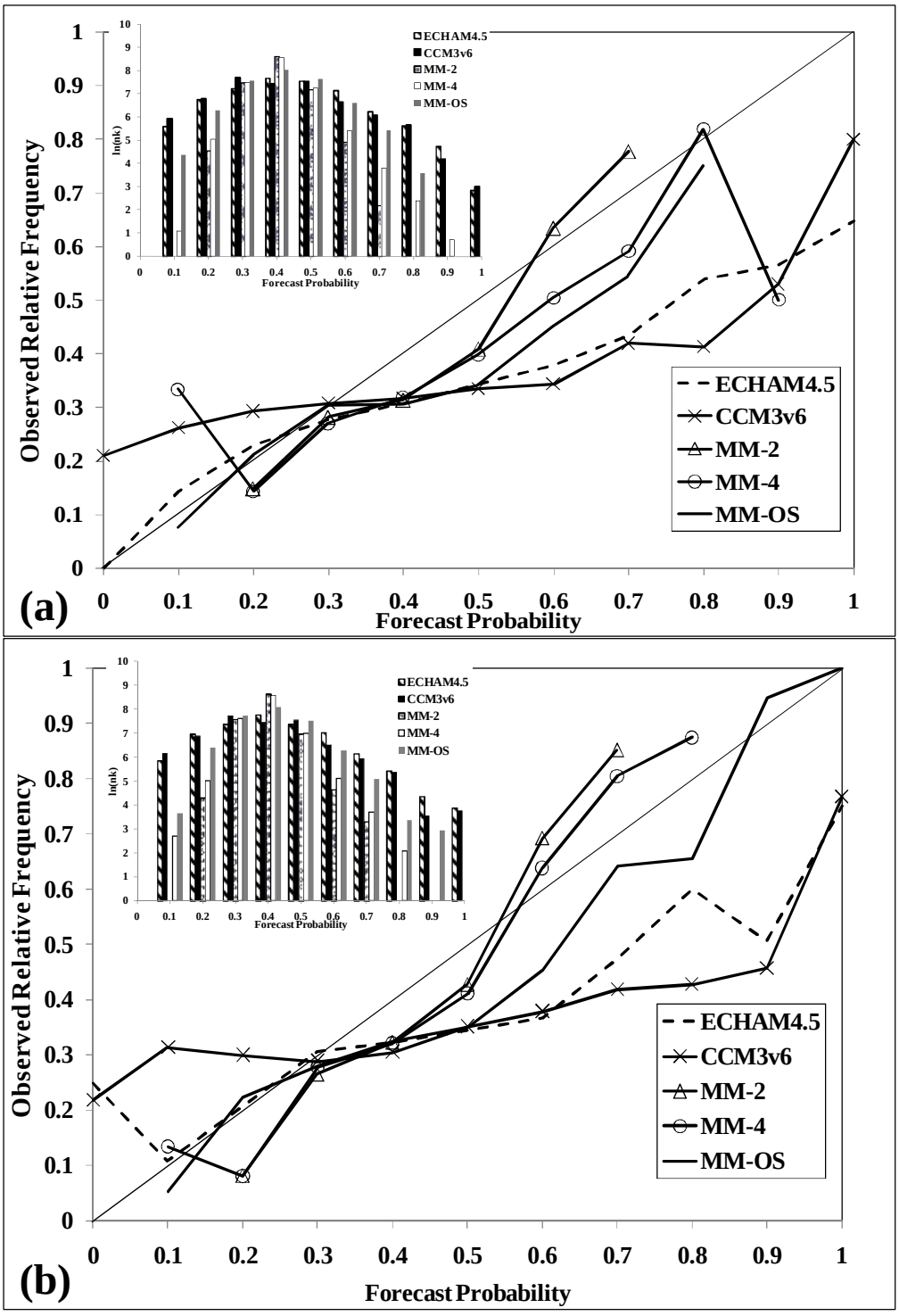

Figure 2.7: Reliability Diagram for individual models, ECHAM4.5 and CCM3v6, and for various multimodel combination schemes in predicting below-normal (Figure 2.7a) and above-normal (Figure 2.7b) categories of precipitation.

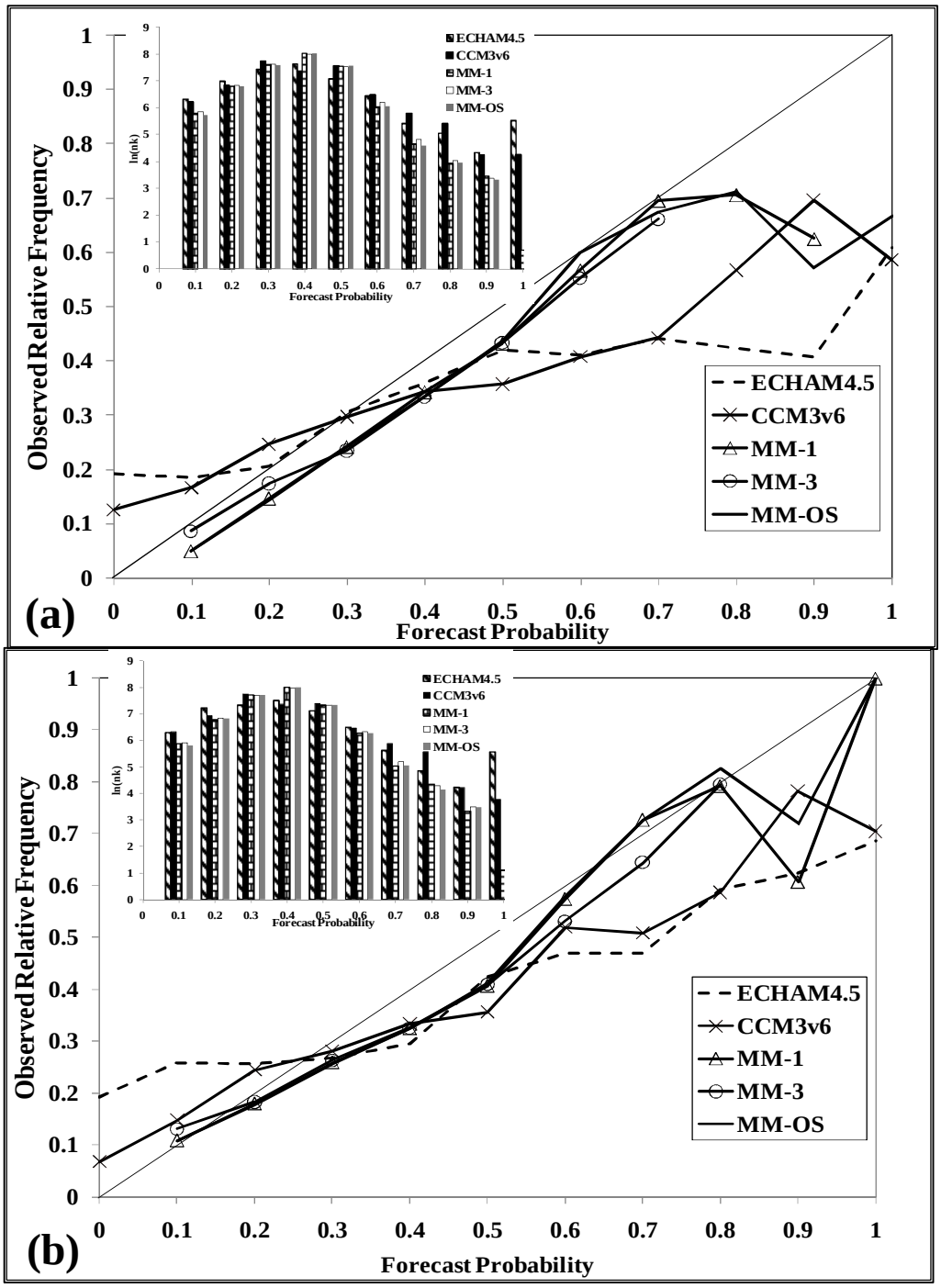

Figure 2.8: Reliability Diagram for individual models, ECHAM4.5 and CCM3v6, and for various multimodel combination schemes in predicting below-normal (Figure 2.8a) and above-normal (Figure 2.8b) categories of temperature.

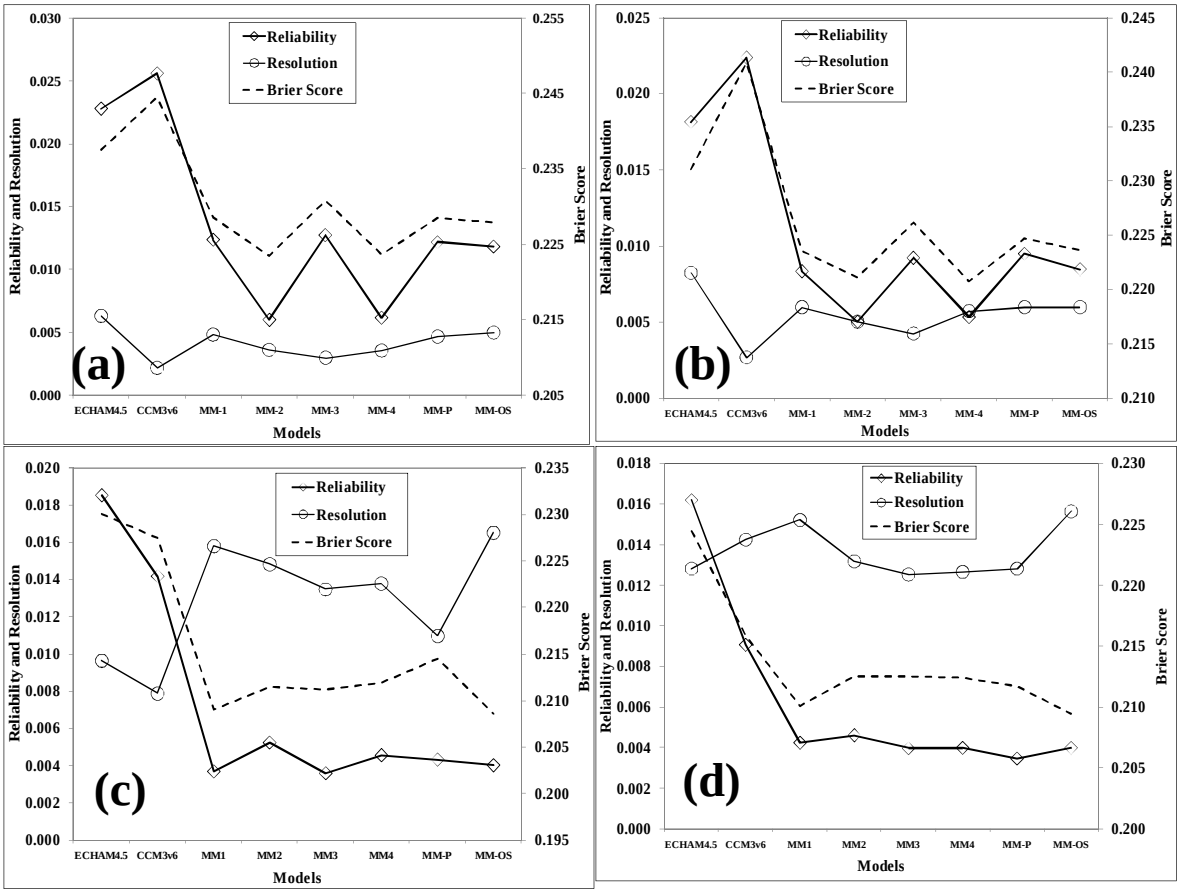

Figure 2.9: Performance comparison of individual models, ECHAM4.5 and CCM3v6, with various multimodels based on Brier Score and its components – reliability and resolution – in predicting below normal (Figure 2.9a – precipitation, Figure 2.9c – temperature) and above normal (Figure 2.9b – precipitation, Figure 2.9d – temperature) events.

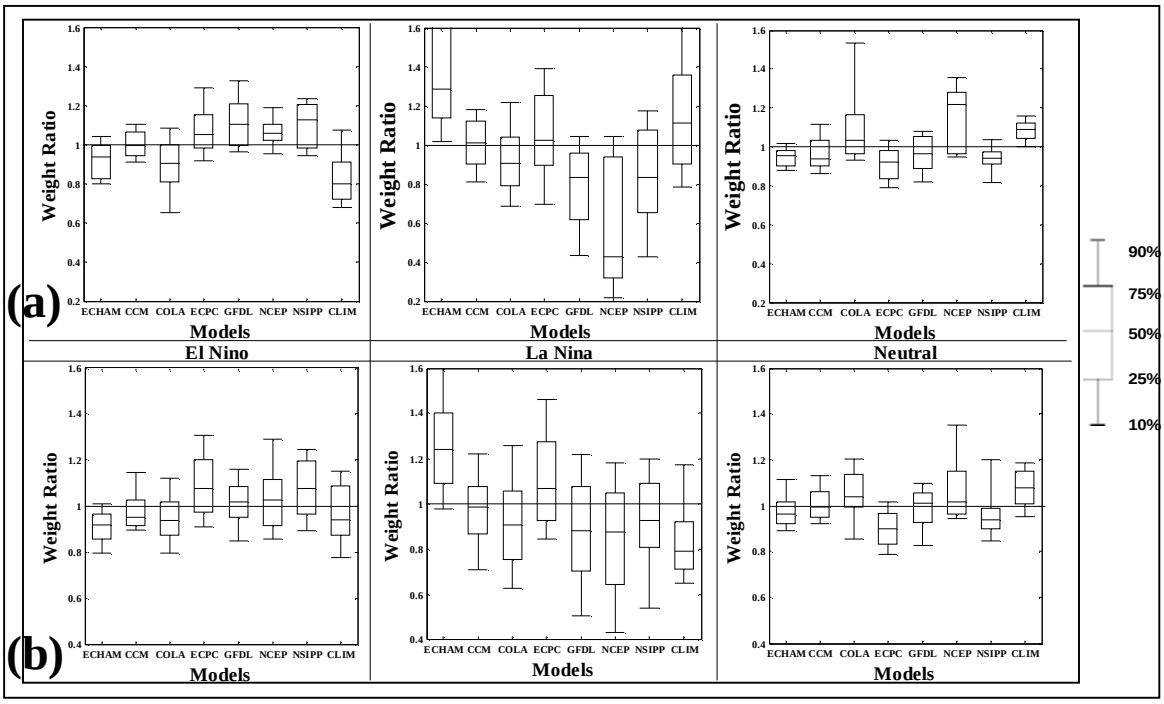

Figure 2.10: Box-plots of the ratio of weights ($w_{t,k}^{m}$ |*MM-1)* for each model under MM-1 scheme to the weights for each model (w^{m} |*MM-OS)* under MM-OS scheme in predicting temperature. Figure 2.10a (Figure 2.10b) shows the weights ratio, pooled – as columns – under three ENSO conditions, for grid points having significant correlation with observed precipitation and the ensemble mean from ECHAM4.5 during El Nino (La Nina) years. Weight ratio above 1, for a given GCM scheme, indicates that the weight under MM-1 is higher than the weight assigned based on MM-OS scheme.

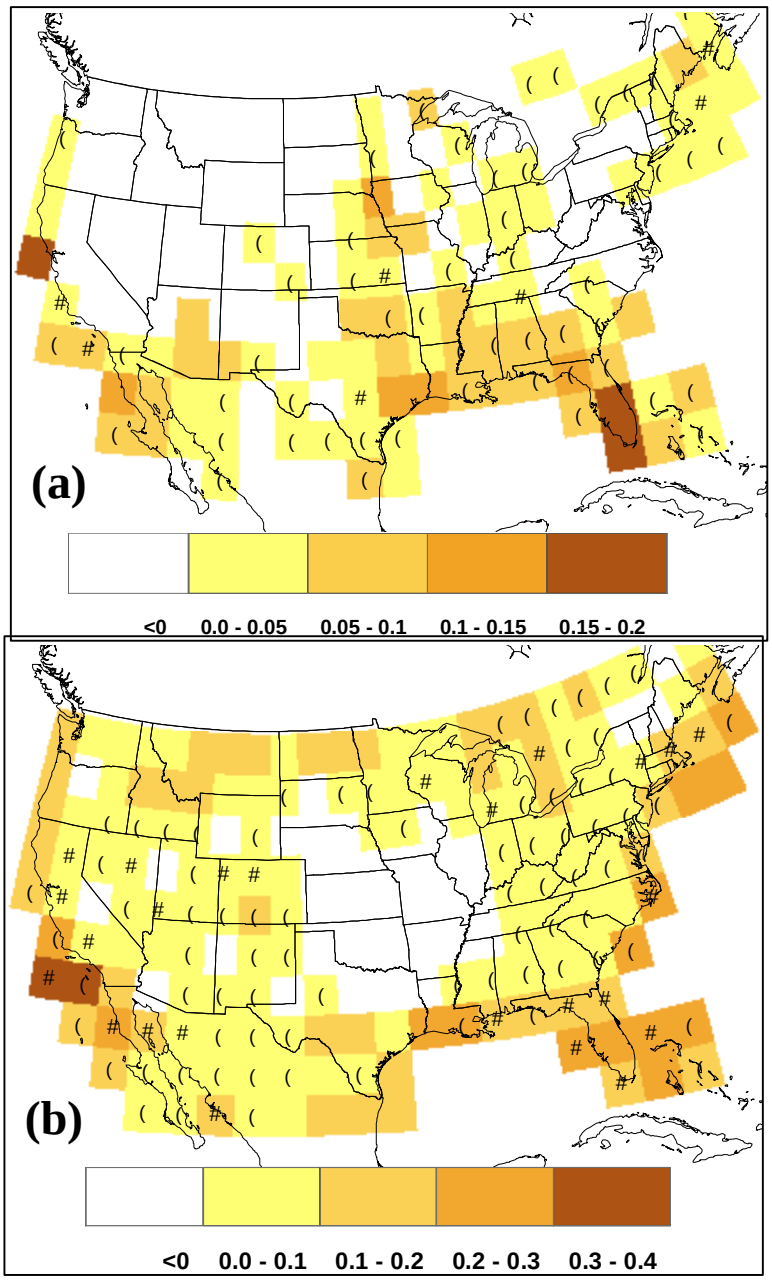

Figure 2.11: Performance of multimodels and individual models, expressed as $\overline{RPSS}$, in predicting DJF winter precipitation (Figure 2.11a) and temperature (Figure 2.11b). Grid points with no markers, open circle and triangle indicate the best performing model (having the highest $\overline{RPSS}$) at that grid point being individual GCMs (all the seven models in Table 2.1), multimodels proposed in this study (MM-1, MM-2, MM-3 and MM4) and existing multimodel techniques (MM-P and MM-OS) respectively.

CHAPTER 3

Multimodel Combinations of Retrospective Precipitation Forecasts from coupled Ocean-Atmosphere Models: Skill Evaluation over the continental United States

3.1 Introduction

Seasonal to interannual climate forecasts up to several months ahead are issued regularly by various national and international agencies based on atmospheric general circulation models (AGCMs) [Goddard et al. 2003], or comprehensive coupled ocean-atmosphere general circulation models (CGCMs) [Saha et al. 2006; Kanamitsu et al. 2002]. Recent efforts have focused on combining multiple GCMs to issue real-time multimodel climate forecasts. Studies have demonstrated that multimodel forecasts developed using approaches such as pooling of the ensembles, optimal weights, or using various statistical techniques have improved predictability and reduced overall error compared to individual model forecasts [Rajagopalan et al., 2002; Robertson et al., 2004; Barnston et al., 2003; Doblas-Reyes et al., 2000; Krishnamurthi et al., 1999]. In Chapter 2, a methodology to combine multiple AGCMs is proposed and evaluated for predicting winter precipitation and temperature over the United States. The methodology assigned weights for each GCM by evaluating their skill over similar predictor conditions using historical simulations of winter precipitation and temperature. The proposed multimodel combination algorithm performed better than individual models and over multimodel combinations based on pooling and long-term skill.

The main goal of this study is to evaluate the utility of the proposed multimodel combination methodology (Chapter 2) in combining real-time precipitation forecasts from CGCMs contingent on the forecasted Nino3.4 state. Using the retrospective precipitation forecasts (forced with forecasted SSTs) from five CGCMs, we demonstrate the multimodel combination methodology based on an adaptive forecasting scheme. The performance of the

developed multimodel winter precipitation forecasts are compared with the performance of individual models' as well as with two of the commonly employed techniques for multimodel combination. Section 3.2 describes the data and the CGCMs used for the study. Section 3.3 presents the different adaptive multimodel combination schemes developed in this study. In Section 3.4, we present the results and analysis by comparing the skill of individual CGCMs and multimodels in forecasting the observed winter precipitation. Section 3.5, we summarize the findings and conclusions from the study.

3.2 Data Description

Retrospective forecasts from the European Union's ENSEMBLES project are used to develop the multimodel precipitation forecasts for winter season over the continental U.S. We considered the same grid points over the U.S. (from Chapter 2) to develop the forecasts. The ENSEMBLES project consists of global coupled ocean atmosphere model forecasts from the UK Met Office (UKMO), Meteo France (MF), European Centre for Medium-Range Weather Forecasts (ECMWF), the Leibniz Institute of Marine Sciences at Keiel University (IFM-GEOMAR), and the European Mediterranean Centre for Climate Change (CMCC – INGV). Forecasts from each model are issued on 1st February, 1st May, 1st August and 1st November and extend up to seven months lead from their respective start times. All the models are represented by nine ensemble members and available for the period of 46 years, 1960 – 2005. In this study, we considered precipitation forecasts issued on 1st November with a lead time of four months. Hence the multimodel precipitation forecasts are developed for the winter months of November (N), December (D), January (J) and February (F) using the

forecasts issued on 1st November. Table 3.1 has the details on each model. Observed monthly precipitation at 0.5°X0.5° available from University of East Anglia (UEA), Climate Research Unit (CRU) [New et al. 2000], is used to assess the skill of each model. Grid points (0.5°X0.5°) of monthly precipitation from UEA were spatially averaged to map the grid points of the CGCMs. In this study, we used the forecasted Nino3.4 (obtained from the ocean model of the CGCM) for the four winter months (N, D, J and F) as the conditioning variable to assess the skill of the respective models. All the forecasts are obtained from the ECMWF Meteorological Archival and Retrieval System.

3.3 Multimodel Combination of Retrospective Precipitation Forecasts: Methodology

3.3.1 Forecasted Nino3.4 as conditioning variable

We used the retrospective climate forecasts from five different CGCMs to demonstrate the utility of the multimodel combination algorithm in a forecasting context. These models obtain the forecasted boundary SST conditions in a coupled mode and have different biases in forecasting the SST states since they differ in their parameterization and process representation. In this study, we used the forecasted Nino3.4 obtained from the ocean models of multiple CGCMs as the conditioning variable [Tippet and Barnston, 2008].

Figure 3.1 shows the skill of the five models in terms of mean squared error in forecasting Nino3.4 for different lead times. We can see that the 4-month ahead forecasts of the ENSO conditions (Nino3.4index) issued in November exhibit high skill since the peak ENSO activity typically coincides during the DJF season. Figure 3.1 also shows the skill of multimodel Nno3.4 forecast obtained by pooling ensembles from all the models. It is evident

that the skill of the multimodel forecast is better than all the models. Enhanced skill of multimodel is observed mainly due to the cancellation of differing biases [Hagedorn et al. 2005]. Jin et al. [2008] and Weisheimer et al. [2009] also showed that the 3-month ahead multimodel forecasts of ENSO conditions issued in November exhibit very high skill with correlations ranging above 0.8. In this study, we employed the multimodel combination algorithm on retrospective precipitation forecast using forecasted Nino3.4 from individual models and multimodel (obtained from pooling) as the conditioning variable. However, the identified similar ENSO conditions using forecasted multimodel mean of Nino3.4 could slightly differ from the identified conditions using observed Nino3.4.

3.3.2 Adaptive Forecasting Multimodel Schemes

The main objective of this study is to demonstrate the utility of the multimodel combination methodology in a forecasting context. Using forecasted Nino3.4 as the conditioning variable, we develop multimodel precipitation forecasts for Nr, D, J and F separately for the period of 1997 – 2005, by training the individual models based on the data available from 1960 - 1996. Table 3.2 provides brief description on the different multimodel schemes considered in the study. Four multimodel (M) schemes, M-1, M-2, M-3, and M-4, are developed by employing fixed neighbors *'K'* to obtain the MSE. If the forecasted Nino3.4 in the conditioning year is under El Nino, La Nina and neutral states, then we estimate the MSE ($l^{m}_{t,K}$) of the CGCM during the forecasted El Nino (*K* = El Nino years in the calibration period, 1960 – 1997), La Nina (*K* = La Nina years) and neutral (*K* = neutral years) years respectively. Thus, we evaluate the skill of the models under similar ENSO conditions in the

calibration/training period. For example, we developed the multimodel precipitation forecasts for the month of November from 1997 – 2005 by evaluating the skill of the individual models in November during the 36 years training period. The conditioning variable is the Nino3.4 forecast for November, and similar neighbors are identified from the 36 years of forecasts for that month. It is important to note that the number of neighbors will be different for different months of N, D, J, and F. The skill of the multimodel forecast in the winter season, NDJF is verified using all the forecasts from N, D, J and F. Hence we have a total of 40 forecasts (10 forecasts from each month).

Multimodels M-1 and M-2 are developed for each month separately based on the skill under ENSO years in the calibration period. However, ENSO conditions are identified based on the multimodel Nino3.4 forecasts for M-1 and individual model Nino3.4 forecasts for M-2. For example, under M-2 scheme, the skill of the precipitation forecasts from UKMO in the calibration period is obtained by identifying similar ENSO years from the forecasted Nino3.4 of the corresponding ocean model of UKMO. This scheme is chosen based on the premise that the errors in forecasting SSTs are one source of error in the CGCM forecast. Hence in a coupled mode, the skill of a particular CGCM is heavily dependent skill of the forecasted SST from its corresponding ocean model. For multimodels M-3 and M-4, the weights are developed for the NDJF seasonal average precipitation forecasts and disaggregated for all the four months. Hence, under this scheme, the weights for N, D, J and F for each year are equal. M-3 uses the multimodel forecasted Nino3.4 for NDJF as the conditioning variable and M-4 uses the individual forecasted Nino3.4 for NDJF from corresponding ocean models to identify similar ENSO conditions. All the multimodel schemes (M-1 – M-4) combine

individual model forecasts along with climatological ensembles. For climatology, we simply consider the 36 years (training period) of observed precipitation at each grid point from UEA. The multimodel schemes developed in study (M-1 – M-4) are different from the multimodel schemes (MM-1 – MM-4) shown in the previous chapter 2.

We developed MM-P and MM-OS, similar to chapter 2 in order to have a baseline comparison with some of the commonly employed techniques in developing multimodel combinations [Palmer et al. 2000; Rajagopalan et al. 2002; Robertson et al. 2004]. MM-P is the multimodel combination scheme that is obtained by pooling all the ensembles from five individual models and climatology. Hence, in MM-P scheme, we have an increased number of ensembles (81) since we are now pooling ensembles from all the models. MM-OS combines individual models based on their overall skill (unconditional of the ENSO state), which is specified based on the MSE for the period 1960-1996 in forecasting N, D, J and F precipitation at a given grid point. The performance of multimodel forecasts in the validation period for the NDJF aggregate season (total of 40 forecasts) are compared with individual models' skill using standard verification measures such as average Rank Probability Score ($\overline{RPS}$), average Rank Probability Skill Score ($\overline{RPSS}$),and average Brier scores. The next section discusses the performance of multimodels in forecasting winter precipitation over the continental U.S.

3.4 Results and Analysis

Six multimodel forecasts (in Table 3.2) of winter precipitation are developed by combing five CGCMs and climatology based on the methodology described in chapter 2. The

developed multimodel forecasts are represented as tercile probabilities in 187 grid points over the continental United States for the four months (N, D, J and F) during the period 1997 – 2005.

3.4.1 Baseline Comparison between Multimodels and Individual Models

Figure 3.2 shows the box plot of $\overline{RPSS}$ for the five individual models and for six multimodels over the entire U.S. Using the multimodels' and individual models' tercile probabilities, we compute the $\overline{RPSS}$ for the NDJF season (using a total of 40 forecasts) based on the forecast developed for each month for the period 1997-2005. Figures 3.2a show the box plot of $\overline{RPSS}$ over the U.S. and figure 3.2b show the box plot of mean squared error based skill score (MSSS) in forecasting winter precipitation respectively. We assumed the conditional distribution resulting from the multimodel combination as normal for computing MSSS. From figure 3.2, we can infer that the individual models' $\overline{RPSS}$ and MSSS is lesser than zero in most of the grid points which implies that the skill of the CGCMs is poorer than climatology. Among the individual models, ECMWF perform better than other CGCMs in forecasting winter precipitation. We can also see that all six multimodels (in Table 3.2) perform better than the CGCMs with more grid points having positive $\overline{RPSS}$ and MSSS in forecasting winter precipitation. Among the six multimodels, we infer that all the multimodels schemes perform better than MM-P. Since ECMWF is the model that is performing better than other CGCMs, further analyses in quantifying the improvements resulting from multimodels will focus only on comparing with the performance of ECMWF.

This is similar to the analysis in Chapter 2 where we compared the performance of ECHAM4.5 and CCM3v6 with all the multimodels.

3.4.2 Statistical Significance of Multimodel forecasts – Hypothesis Testing

To ensure that the improved $\overline{RPSS}$ exhibited by the multimodel schemes, M-1-M-4, is statistically significant over the skill of ECMWF, we perform detailed nonparametric hypothesis tests [Hamill, 1999] by testing the null hypothesis that $\overline{RPS}$ of a multimodel scheme is equal to $\overline{RPS}$ of ECMWF in forecasting precipitation at each grid point. Results from the hypothesis tests, the percentiles of the observed test statistic $\overline{RPS}^{A} - \overline{RPS}^{B}$ on the constructed null distribution, are plotted on the United States map as shown in Figure 3.3 to identify grid points showing significant improvement from multimodel combination. Further, Tables 3.3 summarize the results from hypothesis tests across the six multimodels. Entries in the upper triangle in Tables 3.3 provide the number of grid points having the percentiles of observed test statistic between 0.9-1 on the constructed null distribution. Similarly, results from the same hypothesis tests are also summarized in the lower triangle between the two models indicating the number of grid points over which the percentiles of the observed test statistic fell between 0-0.1 on the constructed null distribution. The best-performing model in terms of increased number of significant grid points is underlined by its column entry. In addition to this, Table 3.4 summarizes the results for the same hypothesis test, after dropping grid points with $\overline{RPSS}$ of both models being negative.

Figure 3.3 shows the relative performance of six multimodel combination schemes over the best individual model, ECMWF, in forecasting winter precipitation over the entire United States. From Figure 3.3a and Table 3.3, 71 (2) grid points have the percentiles of the test statistic falling between 0.9-1(0- 0.1) which indicate that M-1(ECMWF) performs better than ECMWF (M-1) in those grid points by rejecting the null hypothesis that the difference in $\overline{RPS}$ between the ECMWF and M-1 is zero for a significance level of 10%. A triangle symbol is shown in Figure 3.3 to indicate the grid points where the multimodel is performing better than climatology. These grid points are obtained after dropping the grid points with $\overline{RPSS}$ of both models being negative. For example, there are 26 grid points where the multimodel M-1 has $\overline{RPSS}$ that is greater than ECMWF and is positive, indicating that the M-1 is performing better than climatology in those grid points. For the remaining 45 gird points though the percentile of the test statistic is between 0.9 – 1, the multimodel is still poorer than climatology.

Among the multimodels, M-3 and M-4 perform better than the rest of the multimodels. M-3 and M-4 use same weights (developed using average NDJF precipitation) for all the four months. From Table 3.3, we also infer that the performance of MM-OS is equally good in predicting winter temperature. Comparing the performance of M-3 and MM-OS, we infer that in 20 (13) grid points MM-OS (M-3)'s $\overline{RPS}$ is statistically significant than the $\overline{RPS}$ of M-3(MM-OS) indicating that combining models purely based on their long-term skill seem to be a good strategy in multimodel combination. Similar observation is made from the results in Chapter 2 for predicting winter temperature. However, among these

grid points, after dropping grid points with $\overline{RPSS}$ of both models being negative, then we result with 9(6) grid points that show the $\overline{RPS}$ of MM-OS (M-3) being statistically significant than the $\overline{RPS}$ of M-3 (MM-OS). This indicates M-3's better performance is more in grid points exhibiting positive $\overline{RPSS}$. The comparison of all multimodel schemes (M-1 – M-4) with MM-OS for grid points where at least one model's $\overline{RPSS}$ is greater than zero is shown in Table 3.4. It is evident that the multimodel schemes proposed in this study have better performance in grid points exhibiting positive $\overline{RPSS}$.

3.4.3 Comparison of Brier Score between Multimodels and Individual Models

In this section we compare the average Brier Score of multimodel forecasts with the average Brier Score of individual model forecasts. Brier Score summarizes the squared error in categorical forecast probabilities. It can be decomposed into three components, reliability, resolution and uncertainty [Wilks, 1995]. For $\overline{BS}$ to be close to zero, it is important that the reliability term should be close to zero and resolution term should be large. Figures 3.4a and 3.4b provide the reliability, resolution and $\overline{BS}$ for ECMWF and all the six multimodels in forecasting below normal and above normal categories of precipitation. From Figures 3.4, we infer that all multimodels have smaller reliability score in comparison to the reliability scores of individual model under both tercile categories, thereby contributing to the reduction in $\overline{BS}$. In terms of resolution, ECMWF has smaller resolution score than the resolution scores of other multimodels in forecasting precipitation. Similarly, we infer that M-1 and MM-OS have the smallest $\overline{BS}$ which results primarily from smaller reliability score and larger

resolution score under below-normal and above-normal categories. M-1 and MM-OS perform better than individual model as well as over the rest of the multimodels in forecasting winter precipitation over the continental U.S. In the next section, we discuss in detail improvements resulting from multimodel combination from a regional perspective over the continental U.S. particularly for grid points that exhibit positive $\overline{RPSS}$.

3.4.4 Improvements in multimodel precipitation forecasts: Regional Analysis

Figure 3.5 shows the skill, expressed as $\overline{RPSS}$, in forecasting NDJF precipitation with each grid point's $\overline{RPSS}$ being indicated by the best performing individual model or the multimodel. Table 3.5 shows the number of grid points with each individual model and multimodel having the highest $\overline{RPSS}$ over 187 grid points shown in Figure 3.5. Figure 3.5 and Table 3.5 summarize the performance of models only if the $\overline{RPSS}$ of at least one model is greater than zero at a given grid point. From Figure 3.5 and Table 3.5, we can clearly see that multimodels proposed in the study (M-1, M-2, M-3 and M-4) perform better than the individual models and over the existing multimodel combination techniques (MM-P and MM-OS). Among the multimodels, M-2 seems to be best performing multimodel in forecasting precipitation. Under M-2 scheme, the skill of the precipitation forecasts from each model in the calibration period is obtained by identifying similar ENSO years from the forecasted Nino3.4 of the corresponding ocean model.

From Figure 3.5, we understand that the improvements resulting from multimodel combination in forecasting NDJF precipitation predominantly lies over the South and Eastern

United States and some regions over the Western U.S, particularly over California and Arizona. From Figure 3.5, we can understand that there is a significant improvement in $\overline{RPSS}$ of multimodel proposed in the study (shown as open circles) over Eastern, Southwest regions of the United States and over Northwest Mexico. The reason for this improved performance over these regions is primarily due to strong correlation between ensemble mean of the individual models with the observed precipitation under ENSO conditions.

3.5 Summary and Findings

The methodology to combine multiple GCMs that is proposed in Chapter 2 is evaluated for retrospective winter precipitation forecasts over the United States. Considering forecasted Nino3.4 as the primary predictor influencing the winter precipitation, the study combines five CGCMs with climatological ensembles to develop multimodel forecasts over the continental United States. Six different multimodel schemes are developed based on an adaptive forecasting scheme for the period of 1997 – 2005 by training the retrospective forecasts from 1960 – 1996. The skill of the forecast are evaluated by using the forecasts from all the winter months (N, D, J and F) and their performance is compared with individual models based on various verification measures such as rank probability skill score, reliability and resolution scores and brier score. The improvements resulting (from reduction in $\overline{RPS}$) from multimodel combination over individual model is also tested through a rigorous nonparametric hypothesis testing based on resampling. In this study we developed the forecast for each month by identifying similar neighbors from the calibration period of 36 years. In addition to this, the developed methodologies use fixed number of neighbors to

evaluate the skill of the model. The study clearly shows that the proposed multimodel combination algorithm perform better than individual models and over multimodel combinations based on pooling and long-term skill in the forecasting context.

The analyses presented in Chapter 2 and Chapter 3 show that the combining multimodels contingent on the dominant predictor state is an attractive strategy in improving the skill of multimodel forecasts. The improved reliability in multimodel forecasts results in reduced number of false alarms and missed targets in the categorical forecasts. Thus, applying multimodel forecasts would reduce uncertainty from individual models which could lead to better decisions and could also improve public confidence in utilizing seasonal forecasts for water management application. Chapter 4 presents the applications of the multimodel combination methodology in improving the water management of Falls Lake Reservoir in the Neuse River Basin, NC.

3.6 References

A. Alessandri et al. The INGV-CMCC Seasonal Prediction System: Improved ocean initial conditions, submitted to Monthly Weather Review, 2009

Balmaseda, M. A., A. Vidard, and D. L. T. Anderson (2008), The ECMWF ORA-S3 ocean analysis system, Mon. Weather Rev., 136, 3018–3034, doi:10.1175/2008MWR2433.1.

Barnston A. G., S. J. Mason, L. Goddard, D. G. Dewitt, and S. E. Zebiak, 2003: Multimodel ensembling in seasonal climate forecasting at IRI. *Bull. Amer. Meteor. Soc.*, **84**, 1783–1796.

Collins, W. J., et al. (2008), Evaluation of the HadGEM2 model, Tech. Note HCTN 74, Met Off. Hadley Cent., Exeter, U. K.. (Available at http://www.metoffice.gov.uk/publications/HCTN/index.html)

Daget, N., A. T. Weaver, and M. A. Balmaseda (2009), Ensemble estimation of background-error variances in a three-dimensional variational data assimilation system for the global ocean, Q. J. R. Meteorol. Soc., 135, 1071–1094, doi:10.1002/qj.412.

Doblas-Reyes F.J., M.Deque, and J.P.Piedelievre, 2000: Multimodel spread and probabilistic seasonal forecasts in PROVOST. *Quart. J. Roy. Meteor. Soc.*, **126(567)**:2069-2087.

Goddard L., A. G. Barnston, and S. J. Mason, 2003: Evaluation of the IRI's "net assessment" seasonal climate forecasts: 1997– 2001. *Bull. Amer. Meteor. Soc.*, **84**, 1761–1781.

Hagedorn R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus A*, **57**, 219–233, doi:10.1111/j.1600–0870.2005.00103.x.

Hamill T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167.

Jin E. K., J. L. Kinter, B. Wang, C. K. Park, I. S. Kang, B. P. Kirtman, J. S. Kug, A. Kumar, J. J. Luo, J. Schemm, J. Shukla, and T. Yamagata, 2008: Current status of ENSO prediction skill in coupled ocean-atmosphere models, *Climate Dyn*, ***31***, 647-664.

Kanamitsu M., and K. C. Mo, 2003: Dynamical effect of land surface processes on summer precipitation over the southwestern United States. *J. Climate*, **16**, 496–509.

Keenlyside, N. S., M. Latif, M. Botzet, J. Jungclaus, and U. Schulzweida (2005), A coupled method for initializing El Niño Southern Oscillation forecasts using sea surface temperature, Tellus, Ser. A, 57, 340–356.

Krishnamurti T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved weather and seasonal climate forecasts from a multi-model superensemble. *Science*, **286**, 1548–1550.

New M., M. Hulme, and P. D. Jones, 2000: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238.

Palmer T.N., C.Brankovic, D.S.Richardson, 2000: A probability and decision-model analysis of PROVOST seasonal multimodel ensemble integrations. *Quart. J. Roy. Meteor. Soc.*, **126(567),**2013-2033.

Rajagopalan B., U. Lall, and S. E. Zebiak, 2002: Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.*, **130,** 1792–1811.

Robertson A. W., U. Lall, S. E. Zebiak, and L. Goddard, 2004: Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. *Mon. Wea. Rev.*, **132**, 2732–2744.

Saha S., S. Nadiga, C. Thiaw, and J. Wang, 2006: The NCEP Climate Forecast System. *J. Climate,* **19**, 3483-3517.

Schneider E. K., 2002: Understanding differences between the equatorial Pacific as simulated by two coupled GCMs. *J. Climate*, **15**, 449-469.

Tippett M. K., and A. G. Barnston , 2008: Skill of Multimodel ENSO Probability Forecasts, *Mon. Wea. Rev.*, **136**, 3933-3946.

Weisheimer, A., F. J. Doblas-Reyes, T. N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel , 2009: ENSEMBLES: A new multi-model ensemble for seasonal-to-annual predictions—Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys. Res. Lett*, 36, L21711, doi:10.1029/2009GL040896.

Wilks D. S., 1995: *Statistical Methods in the Atmospheric Science*. Academic Press, 467 pp.

Table 3.1: Details of coupled GCMs considered in the study. All models span from longitude (123.75W to 66.25W) and latitude (25N to 45N) resulting in a total of 187 grid points. Retrospective forecasts of precipitation and Nino3.4 from five CGCMs are used for multimodel combination.

Retrospective Forecasts			
Ocean Model	**Atmospheric Model**	**Source**	**Reference**
HOPE	IFS CY31R1	ECMWF	Balmaseda et al. [2008]
HadGEM2-O	HadGEM2-A	UKMO	Collins et al. [2008]
OPA8.2	ARPEGE4.6	MF	Daget et al. [2009]
MPI-OMI	ECHAM5	IFM-GEOMAR	Keenlyside et al. [2005]
OPA8.2	ECHAM5	CMCC-INGV	Alessandri et al. [2009]

Table 3.2: List of multimodel schemes considered for the study. Multimodel schemes shown here are different from the multimodel schemes developed in Chapter 2.

Multimodel Schemes	**Brief Description**
M-1	Multimodel combination using weights based on the skill under ENSO years in the calibration period for each month (N, D, J, and F) separately. ENSO conditions identified based on the multimodel Nino3.4 forecasts.
M-2	Multimodel combination using weights based on the skill under ENSO years in the calibration period for each month (N, D, J, and F) separately. ENSO conditions identified based on the individual model Nino3.4 forecasts.
M-3	Multimodel combination using weights based on the skill under ENSO years in the calibration period for the winter season (NDJF). ENSO conditions identified based on the multimodel Nino3.4 forecasts.
M-4	Multimodel combination using weights based on the skill under ENSO years in the calibration period for the winter season (NDJF). ENSO conditions identified based on the individual model Nino3.4 forecasts.
MM-P	Multimodel combination using pooled ensembles.
MM-OS	Multimodel combination using weights based on the overall skill in the calibration period.

Table 3.3: Number of grid points showing significant difference in $\overline{RPS}$ in forecasting precipitation based on the hypothesis testing between ECMWF and various multimodel schemes given in Table 3.2. Entries in the upper (lower) triangle of the table summarize the number of grid points having the percentile value of the test statistic $\overline{RPS}^{A} - \overline{RPS}^{B}$ between 0.9-1 (0-0.1) in the resampled null distribution for hypothesis testing between Model A and Model B. For values in the upper (lower) triangle, Model A (Model B) is indicated by its row entry and Model B (Model A) is indicated by its column entry. The best-performing model in terms of increased number of significant grid points is underlined by its column entry.

Models	**ECMWF**	**M-1**	**M-2**	**M-3**	**M-4**	**MM-P**	**MM-OS**
ECMWF		71	73	74	75	38	78
M-1	2		17	22	19	0	26
M-2	3	21		24	22	2	24
M-3	1	23	21		16	3	20
M-4	2	29	27	29		5	32
MM-P	8	97	90	100	95		112
MM-OS	2	17	15	13	18	2	

Table 3.4: Number of grid points showing significant difference in $\overline{RPS}$ in forecasting precipitation based on the hypothesis testing between ECMWF and various multimodel schemes given in Table 3.2. Values presented here are based on grid points where the $\overline{RPSS}$ of at least one model is greater than zero.

Models	**ECMWF**	**M-1**	**M-2**	**M-3**	**M-4**	**MM-P**	**MM-OS**
ECMWF		26	25	26	24	5	25
M-1	1		8	10	7	0	12
M-2	1	3		8	6	0	7
M-3	1	9	10		3	0	9
M-4	1	12	13	14		1	15
MM-P	5	59	56	58	54		72
MM-OS	1	5	7	6	7	0	

Table 3.5: Number of grid points with each individual CGCM and multimodels having the highest $\overline{RPSS}$ in forecasting winter precipitation.

Models	**Precipitation**
UKMO	4
ECMWF	5
FRANCE	10
GEOMAR	6
CMCC	0
M-1	8
M-2	18
M-3	8
M-4	10
MM-P	6
MM-OS	11

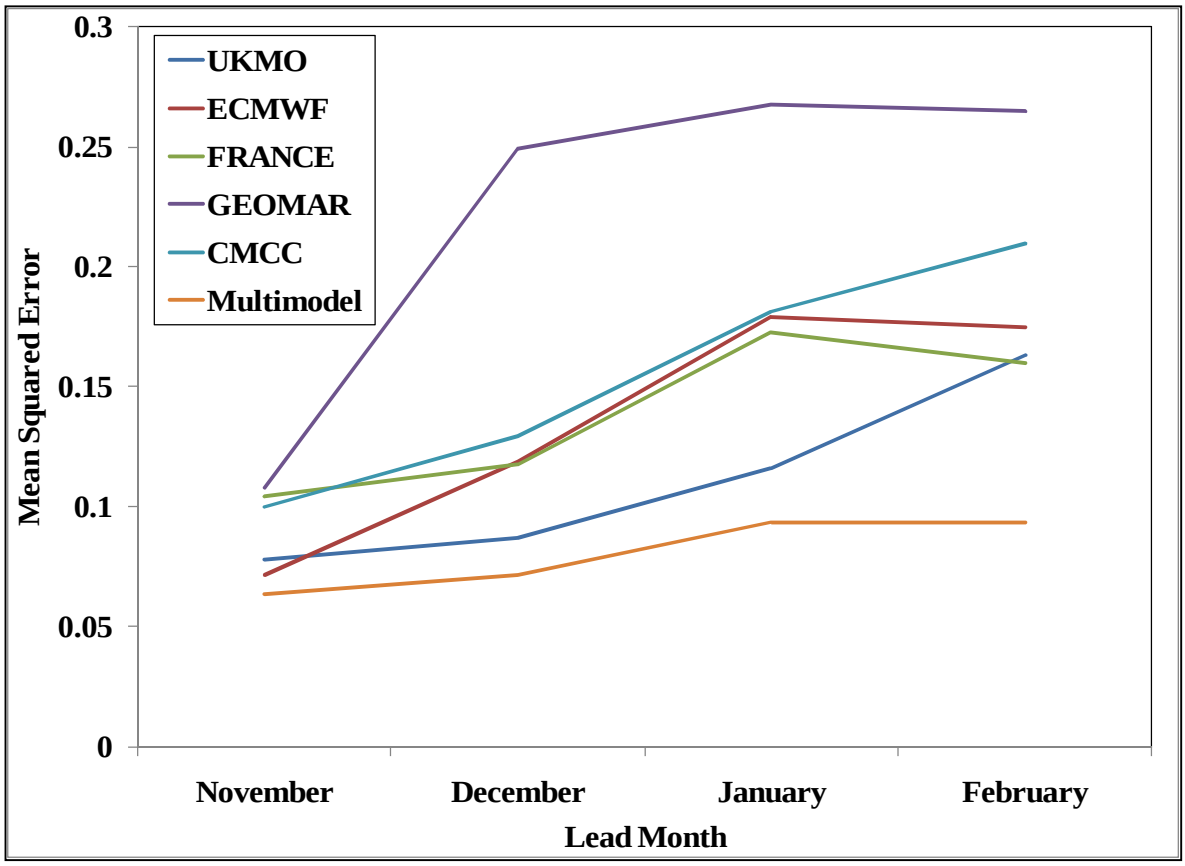

Figure 3.1: Skill expressed as mean squared error between the forecasted Nino3.4 from various ocean models and observed Nino3.4 for the lead time of four months. Retrospective forecasts of Nino3.4 start on 1st November of each year.

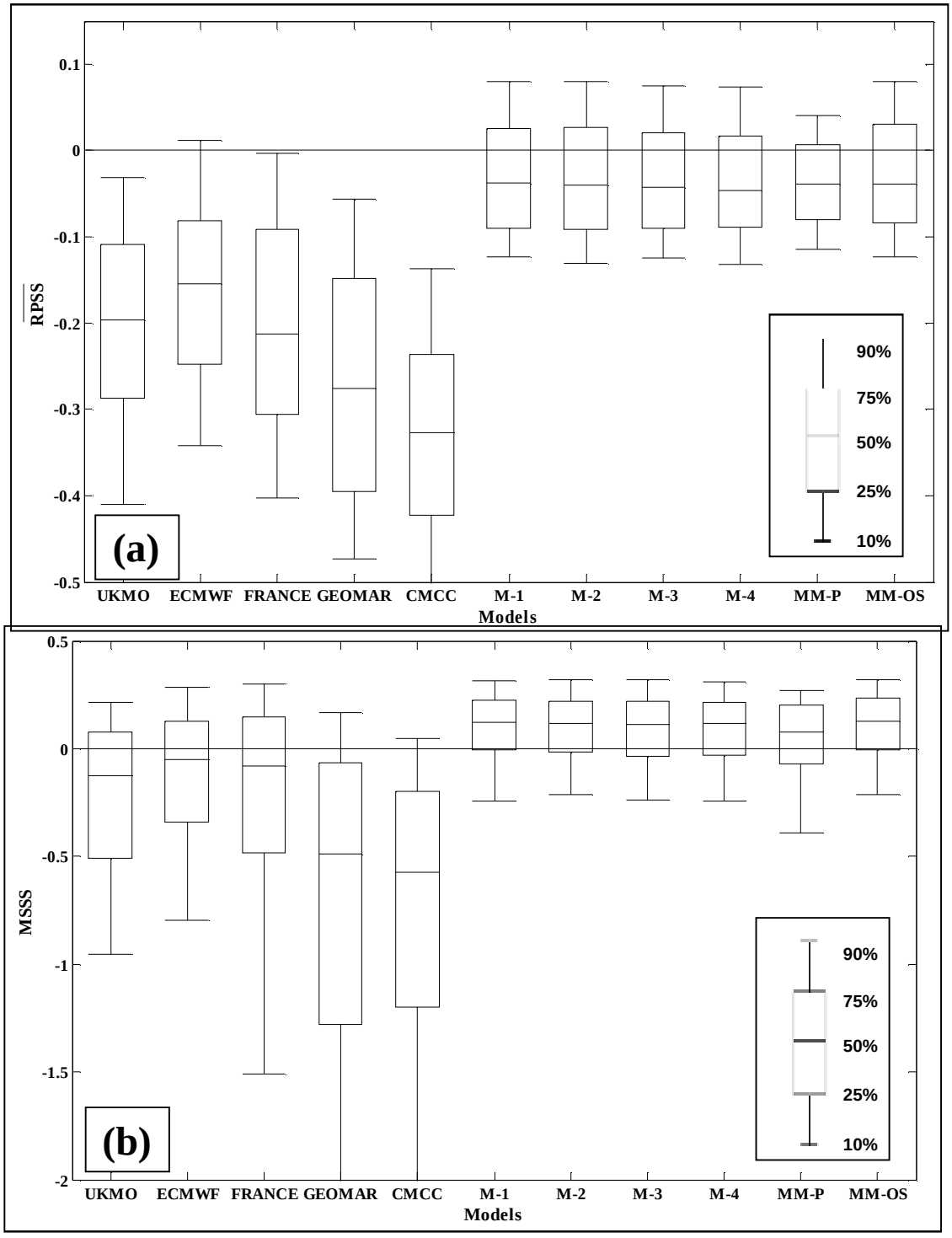

Figure 3.2: Box plots of $\overline{RPSS}$ (3.2a) and MSSS (3.2b) in forecasting winter precipitation for the CGCMs and various multimodel schemes given in Table 3.2.

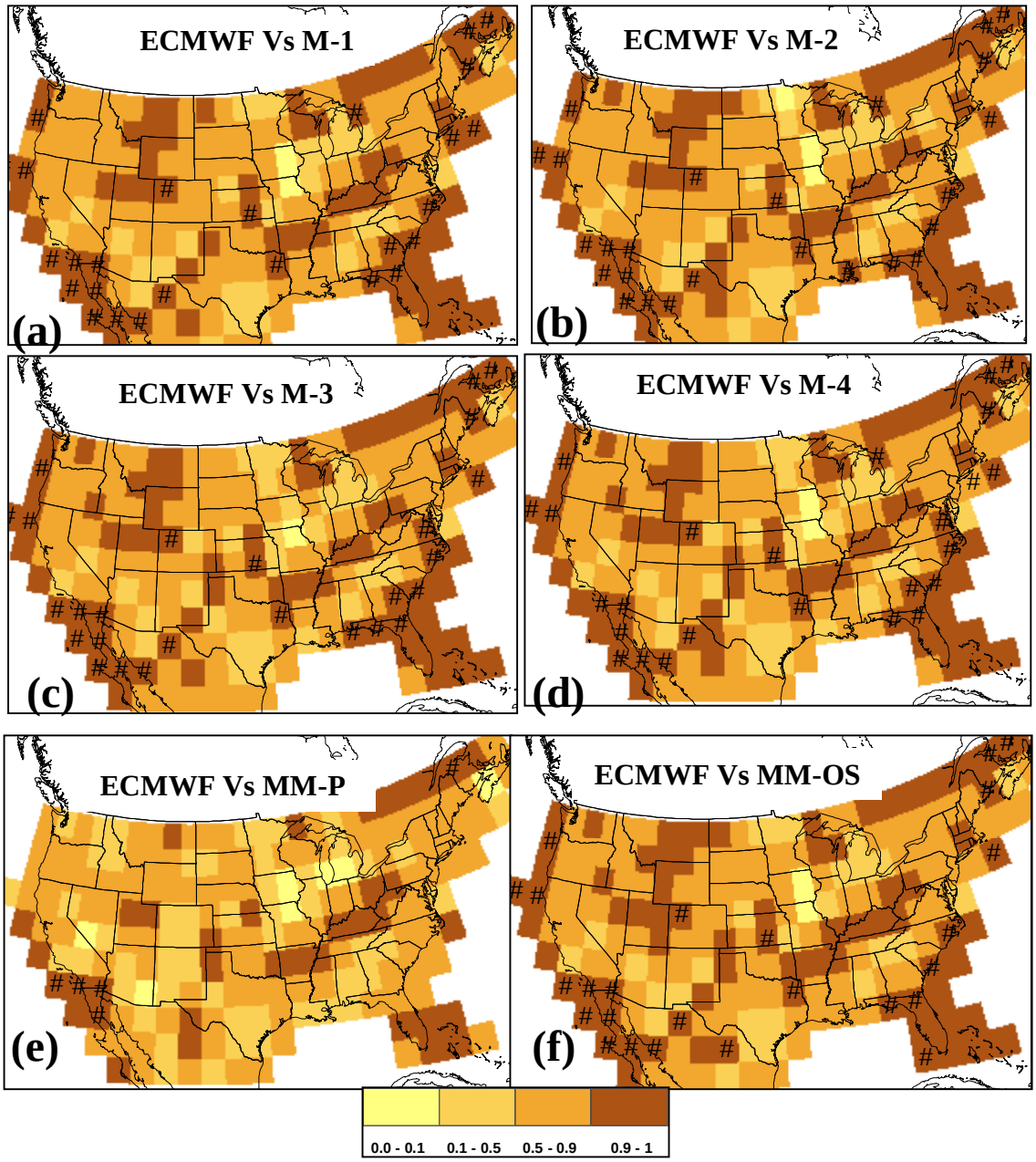

Figure 3.3: Performance comparison of multimodels with the best individual model, ECMWF, in forecasting U.S winter precipitation. The background color indicates the percentile of the test statistic ($\overline{RPS}^{ECMWF} - \overline{RPS}^{MM}$) obtained from the resampled null distribution that represents $\overline{RPS}^{ECHMWF} = \overline{RPS}^{MM}$. A lower (higher) value of the percentiles from the test statistic indicates ECMWF (multimodel) performs better than multimodel (ECMWF). A triangle sign indicates that $\overline{RPSS}$ from multimodel is positive at that grid point.

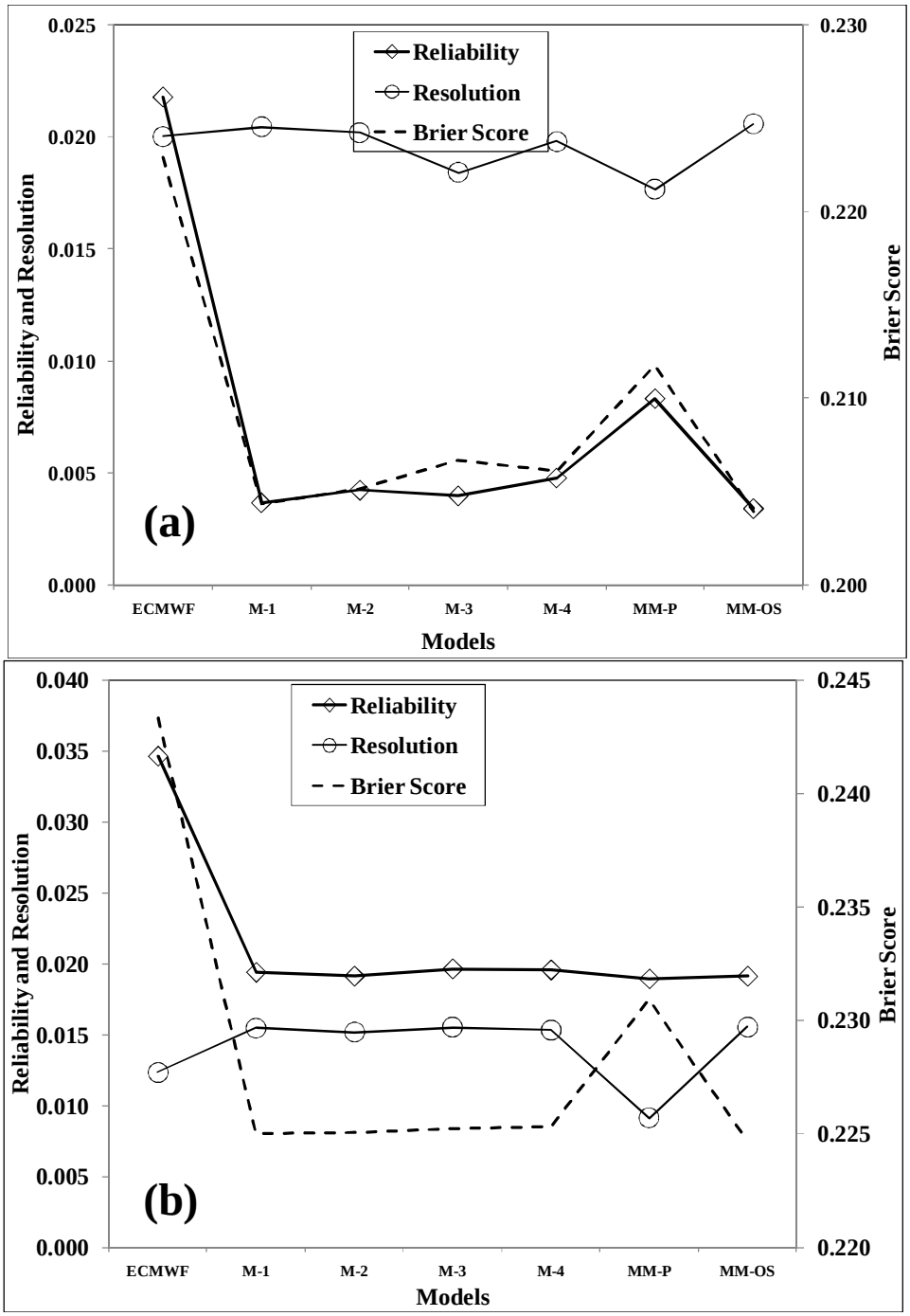

Figure 3.4: Performance comparison of individual model, ECMWF with various multimodels based on Brier Score and its components – reliability and resolution – in predicting below normal precipitation (Figure 3.4a) and above normal precipitation (Figure 3.4b) events.

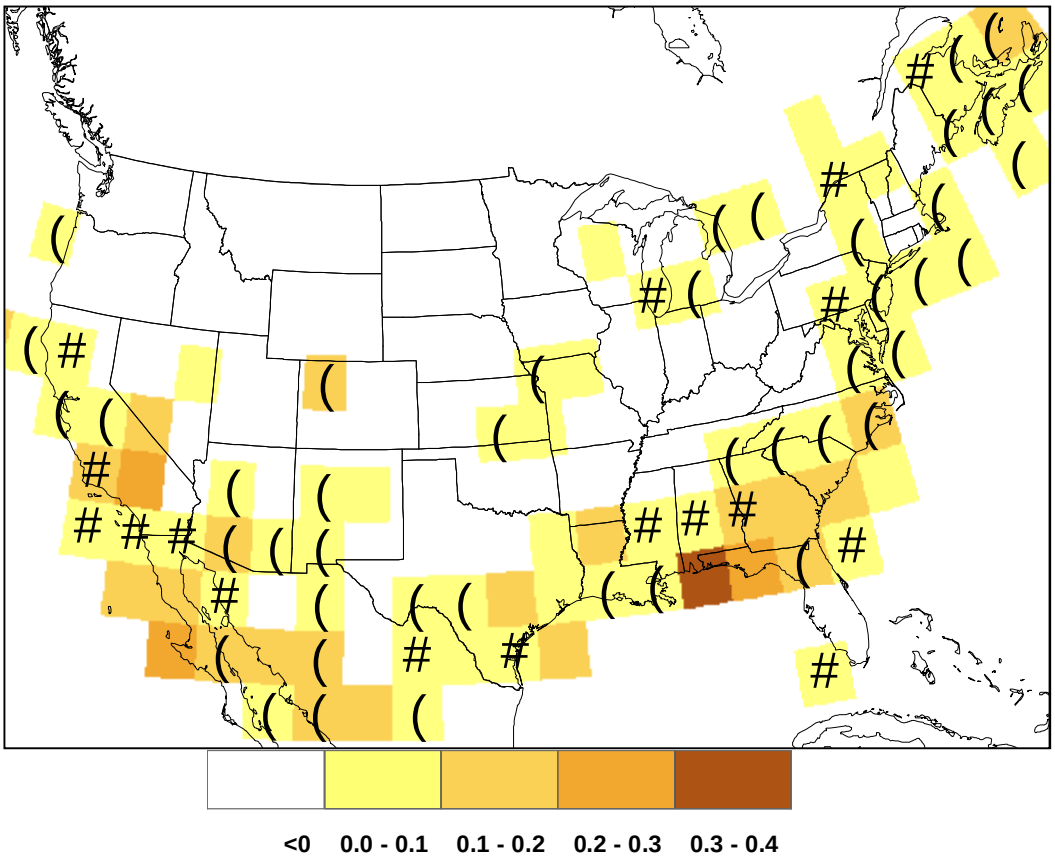

Figure 3.5: Performance of multimodels and individual models, expressed as $\overline{RPSS}$, in forecasting NDJF winter precipitation. Grid points with no markers, open circle and triangle indicate the best performing model (having the highest $\overline{RPSS}$) at that grid point being individual CGCMs (all the five models in Table 3.1), multimodels proposed in this study (M-1, M-2, M-3 and M-4) and existing multimodel techniques (MM-P and MM-OS) respectively.

CHAPTER 4

Improved Drought Management for Falls Lake Reservoir: Role of Multimodel Streamflow Forecasts in Setting up Restrictions

4.1 Introduction

The multiyear drought during 1998–2002 caused severe hardship and economic losses across most of North Carolina [Weaver 2005]. Several local and statewide water supply systems experienced record shortages and many communities operated under mandatory water restrictions from 2001 to 2003 [Weaver 2005]. A similar situation existed during the 2005 and 2007 droughts throughout the state [http://www.ncdrought.org]. Economic losses in North Carolina for the year 2002 were estimated to be $398 million for agriculture and $15–$20 million for municipalities [Weaver 2005]. Unless closely monitored using various sector-specific indicators, the impacts of droughts are progressive, persistent, and pervasive over a large area. Thus, updating drought management plans not only requires monitoring but also needs to include prognostic information about the streamflow potential in the upcoming seasons to develop proactive management measures such as restrictions and hedging. This study combines 3-month ahead climate information based multimodel streamflow forecasts with a reservoir management model that can take ensembles of reservoir inflows to invoke prescribed levels of restriction for water supply.

Droughts experienced by regional water supply systems often result from reduced streamflow/precipitation potential, which could occur due to varying exogenous climatic conditions such as tropical sea surface temperature (SST) [Ropelewski and Halpert 1987; Piechota and Dracup 1996; Barlow et al. 2001]. As water supply systems experience shortages in supply owing to (inflows) natural variability, resulting deficits are further exacerbated by increased demand resulting from urbanization and population growth in the region [Lyon et al. 2005; Vorosmarty et al. 2000]. For instance, in the Triangle Area in North

Carolina, the demand has grown by about 20–62% from 1995 to 2000 [Weaver 2005] resulting in three severe droughts (summers of 2002, 2005, and 2007) in the past 5 years. Given that most of the water supply systems are multipurpose, operating these systems to meet the increased demand under reduced streamflow availability could be very challenging. The main intent of this study is to apply climate information based streamflow forecasts from three models— parametric regression, semiparametric resampling, and multimodel (obtained by combining the former two models)—for setting up restrictions on water supply releases from the Falls Lake Reservoir in North Carolina. By performing retrospective reservoir analyses, the study basically compares the forecasted end of the season target storage probabilities with the climatological probabilities to set up restrictions on water supply releases.

A brief overview of the importance of climate forecasts in reservoir management is discussed in section 4.2. Following that, the reservoir simulation model developed by Arumugam et al. [2003] is detailed (section 4.3), which uses seasonal streamflow forecasts in the form of ensembles, to quantify the reliability of meeting desired releases and the end of the season target storage. Next, the climate information based seasonal streamflow forecasts developed by Devineni et al. [2008] are briefly presented in section 4.4 and are utilized to perform retrospective reservoir analyses to set up restrictions during drought conditions. The following sections 4.5 and 4.6 discuss the potential utility of multimodel forecasts and the proposed reservoir simulation framework for other systems. Finally, the findings of the study are summarized in section 4.7 along with conclusions.

4.2 Background

The National Weather Service River Forecasting System (NWSRFS) issues 3-month lead probabilistic forecasts of streamflow for many river basins in the contiguous United States from 12 river forecasting centers. The Ensemble Streamflow Prediction system from NWSRFS uses conceptual hydrologic models to issue streamflow forecasts based on the current soil moisture, river, and reservoir conditions by assuming that past meteorological events will recur in the future with historical probabilities [Schaake and Larsen 1998]. Recent investigations focusing on the teleconnection between conditions in SSTs and regional/ continental hydroclimatology show that interannual and interdecadal variability in exogenous climatic indices modulate the continental scale rainfall patterns [Ropelewski and Halpert 1987] and streamflow patterns at both global and hemispheric scales [e.g., Dettinger and Diaz 2000] as well as at regional scales [e.g., Piechota and Dracup 1996; Guetter and Georgakakos 1996].

Seasonal streamflow forecasts based on exogenous climatic indices can be obtained using both dynamic and statistical modeling approaches. The dynamic modeling involves coupling a hydrological model with a regional climate model that preserves the boundary conditions specified by the general circulation model (GCM) outputs by considering the topography of the region [e.g., Leung et al. 1999]. However, uncertainty propagation from the coupling of these models [Kyriakidis et al. 2001], representation of physical processes, and low predictive skills of GCM outputs at longer lead time (12–18 months) severely limits the utility of these forecasts for water management. The alternate approach— developing statistical models—focuses on the estimation of conditional distribution of streamflow based

on current conditions of snow pack, streamflow volume, and SST anomalies to issue seasonal and long-lead streamflow forecasts. Various statistical techniques have been employed for this purpose ranging from simple parametric regression models [e.g., Hamlet and Lettenmaier 1999], to complex methods such as linear discriminant analysis [Piechota et al. 2001], spatial pattern analysis [Sicard et al. 2002], and semiparametric resampling strategies [Souza and Lall 2003].

Efforts to develop seasonal streamflow forecasts using tropical/extra-tropical climatic conditions and catchment state have resulted in improved management of water supply systems [Hamlet and Lettenmaier 1999; Yao and Georgakakos 2001; Hamlet et al. 2002]. Using retrospective streamflow forecasts for the Columbia River [Hamlet and Lettenmaier 1999; Hamlet et al. 2002], studies have shown that long-lead streamflow forecasts can be effectively utilized in operating reservoirs to obtain increased annual average hydropower. Similarly, coupled hydraulic–hydrologic prediction models with robust forecast-control methodologies could also result in increased resiliency of reservoir systems to climate variability and change [Georgakakos et al. 1998]. As seasonal streamflow potential changes depending on climatic and land surface conditions, policy instruments and operational rule curves could also be developed to support adaptive water management [Arumugam et al. 2003].

Though the utility of climate forecasts in improving water management has been shown in the literature, it is widely acknowledged that numerous challenges/gaps exist in the real-time application of climate forecasts by water managers [Pagano et al. 2001; Hartmann et al. 2002; Steinmann 2006]. The primary challenges are: (1) low confidence on the skill of

the forecast; (2) communication of probabilistic forecasts; and (3) non availability of decision framework and policy instruments for application. Further, even if the skill of the climate forecasts is significant in a given region, public perception of forecasts particularly goes down due to false alarms (forecast suggests drought, but no drought occurs) and missed targets (forecast suggests normal, but drought occurs) [Steinmann 2006]. For a detailed discussion on the use of climate forecasts in the context of drought management, see Steinman [2006]. It has been widely shown in the literature that multimodel climate/streamflow forecasts ensure better correspondence between the forecasted probabilities and their observed relative frequencies [Barnston et al. 2003; Doblas-Reyes et al. 2000; Devineni et al. 2008] resulting in reduced false alarms and model uncertainty. The main goal of this study is to understand whether application of multimodel streamflow forecasts results in invoking better management decisions (in comparison to individual model forecasts) such as restrictions for improving water allocation during droughts.

4.3 Falls Lake System Details and Management Model Development

Falls Lake is a man-made reservoir in the upper Neuse River, N.C. (Figure 4.1) operated by the U.S. Army Corps of Engineers (USACE) since December 1983 to serve five purposes: (1) flood control; (2) water supply; (3) water quality; (4) recreation; and (5) fish and wildlife. The lake is long and narrow in shape and extends 28 miles up the Neuse River. Three rivers—the Eno, Flat, and Little Rivers—provide the majority of inflows. As a water supply reservoir, Falls Lake provides Raleigh, by contract, with up to 100 million gallons of water a day. Due to the population growth in the city of Raleigh and in the suburbs served by

Falls Lake over the last decade, storage conditions in Falls Lake have been increasingly stressed recently resulting in three severe droughts (2002, 2005, and 2007) over the last 5 years. Current drought management and monitoring activities are coordinated by the North Carolina Drought Management Advisory Council (NCDMAC) in coordination with various state and federal agencies in North Carolina.

4.3.1 Data and Operational Constraints

For operational purposes, reservoir storages of Falls Lake are divided into various pools: (1) flood control pool (controlled storage, 251.5–264.8 ft and uncontrolled storage, 264.8–289.2 ft); (2) Conservation pool (251.5–236.5 ft) with two separate storage accounts for water quality and water supply; and (3) sediment dead storage (236.5–200 ft). All elevations (in feet) are based on the North America Vertical Datum of 1927 (NAVD27). Both water supply and water quality releases are met based on the storages in conservation pool by devoting 39% of the conservation pool storage volume to water supply and the remaining 61% to water quality purposes.

The USACE uses 251.5 ft (131.395 acre-ft) as the operational rule curve or the target pool level, which is obtained based on the average monthly flows recorded at Falls Lake [http://epec.saw.usace.army.mil\Falls_WC_Plan.pdf]. Thus, the USACE tries to ensure the reservoir level at operational rule curve at the beginning (July 1) and end (September 30) of the summer season. During wet summer years (e.g., 1996 and 1999), the above-normal inflows force the reservoir level above 251.5 ft posing operational constraints on flood control and recreation. Under such situations, the USACE releases additional water to

maintain the operational rule curve to reduce the downstream flood risk. Normal outflows for protecting downstream water quality in the Neuse River are 254 cubic feet per second (cfs). However, reservoir outflows during below-normal storage conditions could be reduced to 100 cfs (April to October) and 60 cfs (November to March) after consultation with all stakeholders. Additional information such as monthly releases, stage–storage, and stage–water spread area relationships was obtained from USACE to develop the Falls Lake simulation model, which is described in detail in the following section.

4.3.2 Falls Lake Reservoir Model Formulation

Given seasonal (T-month lead) ensemble inflow forecasts q_j^k and initial reservoir storage, S_0^*, at the beginning of the allocation period (for these analyses, July 1) with j=1,2, . . . ,N denoting the forecast years (N=total number of years of retrospective forecasts), and k=1,2,..., K index representing a particular member out of K ensembles the Falls Lake simulation model determines the seasonal releases R_1 and R_2 representing water supply and water quality allocations, respectively, with specified reliabilities of (1-p_{f1}) and (1-p_{f2}), where p_f implies failure probability. In addition, the water allocation model incorporates an end of the season target storage, S_T^* (T denoting the forecast lead time in months) that is associated with a failure probability p_s. For instance, in the case of Falls Lake, S_T^* corresponds to the storage of the reservoir at 251.5 ft operational rule curve. The simulation model could also estimate the probabilistic constraints [in Eqs. (4.7) and (4.8)], reliability of supply for each use [(1-p_{f1}) and (1-p_{f1})], and p_s given the specified demand R_1^* and R_2^*, for each use along with

S_T^* and S_0^*. Using the basic continuity equation, the seasonal storage equations for each ensemble member *k* are updated for the forecasting year *j*

$$S_{T,j}^k = S_{0,j}^* + q_j^k - E_j^k - (R_{1,j} + R_{2,j}) \quad \ldots (4.1)$$

where seasonal storage equations are constrained so that the storage is between the minimum and maximum possible storage, $S_{\min}$ and $S_{\max}$, respectively

$$S_T^k = \min(S_T^k, S_{\max}), \qquad S_T = \max(S_T^k, S_{\min}) \quad \ldots (4.2)$$

In the event, the end of season storage falling below the minimum possible storage, $S_{\min}$, we encounter deficits, SD_j^k, which needs to be distributed among the users as restrictions

$$SD_j^k = (S_{\min} - S_{T,j}^k) \Big| S_{T,j}^k < S_{\min} \quad \ldots (4.3)$$

$$SD_j^k = \sum_{i=1}^{2} w_{i,j}^k, \quad w_{i,j} = a_i R_{i,j} \quad \ldots (4.4)$$

The restrictions, w_i, for each user could be specified exogenously as a fraction, α_i, of the target release, R_i. The restriction fraction, α_i, could also be allowed to vary depending on

the restriction level. Evaporation, E_j^k is computed as a function of average storage during the season using the water spread area and storage information of the reservoir

$$E_j^k = y_{\ j} d_1((S_0^* + S_T^k)/2)^{d_2} \quad \ldots (4.5)$$

where ψ_j = seasonal evaporation rate and δ_1 and δ_2 = coefficients describing the area–storage relationship. Spline interpolation was employed for obtaining the water spread area corresponding to the average season storage computed for each ensemble. It is important to note that the evaporation is evaluated implicitly for each streamflow member in the ensemble. The estimated average lake evaporation rate (ψ_j) = 0.996 ft/season (after adjusting with the pan coefficient of 0.7) for the summer, which is obtained from the monthly pan evaporation recorded at Chapel Hill, N.C.

The objective is to determine R_i, such that the releases for ith use is bound by the minimum and maximum demand for the season

$$R_{i,\min} \le R_i \le R_{i,\max} \quad \ldots (4.6)$$

Similarly, the study also enforces the probability of having the end of the season storage, S_T, less than the target storage, S_T^*, to be small represented by its failure probability (Prob), p_s, using

$$\text{Prob}(S_T \le S_T^*) \le p_s \quad \ldots (4.7)$$

To ensure the obtained release, R_i, being met with high reliability, (1-p_{fi}), the model includes

$$\text{Prob}(w_i \le w_i^*) \le p_{fi} \quad \ldots (4.8)$$

where w_i^*, specified by the user, denotes the maximum restriction that could be enforced for each user as part of the restrictions. This constraint basically accounts for the uncertainty in releases. Thus, the obtained seasonal release may be between the desired bounds $R_{i,\min}$ and $R_{i,\max}$, but the specified release R_i has a small probability, p_{fi}, of facing restrictions being less than w_i^*. The restriction w_i is calculated for each ensemble member k using the restriction fraction, α_i, based on Eq. (4.4). For all the analyses, w_i^* =0 is assumed.

Looking across all the traces in the ensemble, the model computes the following probabilities to evaluate Eqs. (4.7) and (4.8):

1. $\text{Prob}(w_i \le w_i^*)$ is estimated from the number of traces in which $(w_i \le w_i^*)$ out of total number of traces, N. This includes the calculation of the failure to meet the two specified demands, water supply and water quality.

2. $\text{Prob}(S_T < S_T^*)$ is obtained from the number of traces in which $(S_T < S_T^*)$ out of total number of traces, N.

N=500 ensembles are considered that represent the average seasonal streamflow during the summer [July, August, and September, (JAS)]. In this study, instead of obtaining R_1 and R_2

for the specified constraints, the water supply release (in Figure 4.2), $R^{*}_{1,j}$, and water quality release, $R^{*}_{2,j}$, (with the average flow being equal to 254 cfs or 100 cfs) are specified to estimate the probabilistic constraints in Eqs. (4.7) and (4.8). The above-mentioned probabilities are then computed across the ensembles to evaluate the above listed constraints. Though the model is presented in a simulation framework, it could be extended into an optimization-simulation model by including compensations under restrictions along with a detailed contract structure [Arumugam et al. 2003].

4.3.3 Reservoir Model Verification

Prior to performing the retrospective reservoir analyses using the streamflow forecasts, model verification was performed from 1991 to 2005 by comparing the reservoir model's ability to simulate the observed end of September storages. The model simulations were performed by forcing the model with the observed flows during JAS and initial storages in July to determine the end of the September storages by allocating the reported releases for water quality and water supply. Basically, this verification provides a check on the mass balance of the reservoir model as well as in its ability to model the conservation storage pool into two separate accounts (i.e., water supply and water quality storages). Figure 4.2 shows the observed and model predicted stages at the end September—the end of the season stage. The observed and modeled storages obtained from the reservoir model were converted into stages using the available stage–storage relationship for Falls Lake. Figure 4.2 clearly shows that the developed model is quite reasonable in predicting the observed September storages

upon simulation with observed flows and reported releases. This gives the confidence in employing the simulation model presented here for further analyses that utilize the seasonal streamflow forecasts from three models for invoking restrictions.

4.4 Seasonal Streamflow Forecasts for Falls Lake

This section briefly describes the development of streamflow forecasts for Falls Lake during the summer season. For additional details on the streamflow forecasting model, predictor identification, and the skill of cross-validated forecasts, see the forecasting paper [Devineni et al. 2008] and the technical report [Sankarasubramanian et al. 2006] (available online: http://www.stat.ncsu.edu/library/papers/mimeo2595.pdf). Seasonal streamflow forecasts were developed for the summer season based on April, May, and June (AMJ) climatic information, denoted by anomalous SST conditions in the tropical Pacific, tropical North Atlantic, and over the North Carolina coast.

Predictor identification using Spearman rank correlation was performed on the International Research Institute for Climate and Society (IRI) data library between the global SSTs [http:// iridl.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.v2/ .ssta/] and the seasonal streamflows. Grid points of SSTs (black rectangles) in Figure 4.3(a) that have significant correlation with the predictand were considered as predictors in developing the forecasts. The correlations shown in Figure 4.3(a) are for 78 years of flows. Thus, if the absolute value of correlation is greater than 0.22, then one expects the correlation between the predictor and predictand to be statistically significant (at 5% confidence level). As the SST grid points were spatially correlated, principal component analysis was performed and

the first two principal components (explained 73% of total variance exhibited in SSTs) were retained for model development. Spring season (April–June) streamflow and the previous month's streamflow (June alone) were also considered as surrogate predictors to incorporate land surface conditions such as soil moisture. But, the correlations between the previous month/seasonal flows and the summer flows are statistically not significant.

Two nonlinear models, parametric regression (with the predictand being cube-root of the flows) and semiparametric resampling models [Souza and Lall 2003], were considered in developing multimodel forecasts. With regard to individual model selection, one can even consider the land surface model in developing streamflow forecasts. As the skewness of the recorded summer flows is 1.9, cube-root transformation was applied for developing the parametric regression model. With regard to individual model selection, one can even consider the land surface model in developing streamflow forecasts. Studies have considered objective criterion along with stepwise regression to select the best combination of nonlinear models in developing multimodel forecasts [Regonda et al. 2006]. In this study, the resulting seasonal streamflow forecasts from parametric regression and semiparametric resampling models were combined using a multimodel combination algorithm to develop improved seasonal streamflow forecasts [Devineni et al. 2008, Sankarasubramanian et al. 2006].

This study employed seasonal streamflow forecasts from three models—regression, resampling, and multimodel—for improving the drought management of Falls Lake. The adaptive forecasts for the period 1976–2005 were developed by training the model using the observed flows and predictors available from 1928 to 1975. The correlations between the observed flows and the ensemble mean of the seasonal streamflow forecasts are 0.44, 0.49,

and 0.51 for resampling, regression, and multimodel, respectively, which are statistically significant for the 30 years of validation. Figure 4.3(b) shows the adaptive forecasts from the three models for the period 1991–2005. The forecasts (in Figure 4.3(b)) are shown as conditional mean, which is obtained from the 500 ensembles of the conditional distribution of streamflows developed for each year. Representing the conditional distribution with large ensembles will only lead to better estimates of probability constraints [Eqs. (4.7) and (4.8)] without improving the skill of the probabilistic forecasts. For instance, with regard to the parametric regression model, the actual information content in the forecasts is purely determined by its conditional mean and conditional variance. The null forecast, the climatological ensembles, whose ensembles was also considered were developed by simple bootstrapping of JAS flows. This approach is reasonable, as there is no year-to-year correlation between summer flows at Falls Lake. These streamflow forecasts and the initial storages observed on July 1 were provided as inputs to the reservoir management model to estimate the reliability of meeting the water supply releases (in Figure 4.2) and minimum water quality releases as well as to estimate the probability of end of September storage being below the target storage (corresponding to target stage 251.5 ft) [$\text{Prob}(S_T < S_T^*)$].

4.5 Results and Analyses

The following analyses show the utility of streamflow forecasts in predicting the below-normal storage conditions that could result by releasing the required water supply and water quality releases. This is different from identifying the streamflow forecasts as below normal, as future reservoir storages need to account for the initial conditions in July as well

as lake evaporation, which in turn depend on unknown future storages. Thus, the analyses presented here utilized the streamflow forecasts from three models [Figure 4.3(b)] to obtain: (1) the reliability of supplying the seasonal demand for water quality and water supply uses and (2) probability of having end of September storage less than the target storage [Prob $(S_T < S_T^*)$]. Based on the obtained target releases from each model, the performance of streamflow forecasts from each model is validated in predicting future storage conditions by combining reservoir releases with the observed flows. The study also identified possibilities for imposing restrictions based on the end of the season target storage probabilities estimated by each forecasting model. Predicting below-normal storage conditions well in advance would help in imposing restrictions before the summer season, which could improve the resilience of the system during prolonged droughts.

4.5.1 Reliability of Meeting the Target Releases

The proposed simulation model [Eqs. (4.1) – (4.8)] could be employed in one of the following two ways for a given use: (1) obtain the release for the specified reliability; or (2) obtain the reliability for the specified target release. The first approach is more useful when the seasonal demand is more than the initial storage and the forecasted inflows, whereas the latter is more useful when the initial storage and the forecasted inflows are more than the net seasonal demand. On the other hand, if the initial storage itself could meet the net seasonal demand including lake evaporation, then the reliability of meeting target releases is 100%. Under those situations, there is limited/no use of forecasts, as the available storage itself ensures the total seasonal demand. The required seasonal release for water quality use is

43560 acre-ft (254 cfs). Under severe droughts, this could be reduced to 18,248 acre-ft (100 cfs) with the approval of NCDMAC.

Without constraining the end of the season target storage [Eq. (4.7)], the initial storage itself was able to supply the required water quality release of 254 cfs and the required water supply releases for all 15 years considered for analyses. Figure 4.4 shows the modeled storages and stages for the two scenarios of water quality releases. The modeled storages and stages were obtained by combining the observed streamflows during JAS with the specified water quality and water supply releases. As seen in Figure 4.4, the modeled end of September storages were greater than 23,073 acre-ft (236.5 ft) — the storage at the bottom of conservation pool. Thus, the initial storage itself was able to supply the entire water quality and water supply releases each year.

All the streamflow forecasting models guaranteed the target release with 100% reliability, thereby limiting the use of forecasts when the end of the season target storage is not constrained. In other words, it is important that the total water demand (water supply and water quality releases along with evaporation losses) over the forecast lead time needs to be constrained by the available storage. Under those conditions, the forecasts are useful in assigning the reliability for various uses. If the initial storages do not constrain the net demand, then there is limited/no use of forecasts in quantifying the future storage scenarios. Thus, results shown in Figure 4.4 correspond to 100% reliability of supplying the target releases with all the streamflow forecasting models suggesting the same release scenarios which leaves the forecasts redundant. On the other hand, by enforcing the end of the season target storage constraint with p_s=0.5, water quality and water supply releases are reduced

considerably for a specified reliability of 90%. This shows that the end of the season target storage is the binding constraint, which could be effectively used to invoke restrictions even though the initial storage may ensure 100% reliability of supplying the total seasonal demand. The following analyses shows in detail on how restrictions could be employed based on the estimates of $\text{Prob}(S_T < S_T^*)$ obtained from the reservoir simulation model.

4.5.2 End of the Season Target Storage Probabilities

Given the streamflow forecast ensembles and the initial storage conditions in July, one can estimate the $\text{Prob}(S_T < S_T^*)$ [Eq. (4.7)] that would result upon releasing any of the two water quality release scenarios and the corresponding required water supply releases shown in Figure 4.2. Figure 4.5 shows the estimates of $\text{Prob}(S_T < S_T^*)$ where S_T^*=131,395 acre-ft for releasing the water supply demand (in Figure 4.2) and for two water quality release scenarios. Thus, the probability estimates shown were obtained from each streamflow forecasting model and from climatological ensembles, which were constructed by simply bootstrapping JAS stream-flows. Figure 4.5 also shows the observed streamflows in each year suggesting their tercile category (Q_t < 0.33 percentile — below normal; Q_t < 0.66 percentile — above normal; otherwise — normal). Both Figure 4.5(a) (normal water quality releases=254 cfs) and Figure 4.5(b) (drought conditions water quality releases =100 cfs) demonstrate that the estimates of $\text{Prob}(S_T < S_T^*)$ vary depending on the forecasted streamflow potential by each model. In comparison, estimates of $\text{Prob}(S_T < S_T^*)$ from climatological ensembles do not vary much, with the values hovering between 0.6 and 0.7 for

the normal water quality releases and from 0.5 to 0.6 for the drought conditions water quality release. The small variations that are seen under climatological ensembles are primarily due to differences in initial storage conditions. Similarly, for the year 1995, the estimates of Prob $(S_T < S_T^*)$ are very low due to increased initial storage with the reservoir's initial stage at 255.98 ft, which is almost 4.5 ft above the operational rule curve of 251.5 ft. It is important to note that Figure 4.5 does not use the observed streamflows to estimate Prob$(S_T < S_T^*)$, as the forecasts were developed based on the climatic information available during April–May–June.

Figures 4.5(a and b) also show clearly that the estimates of Prob$(S_T < S_T^*)$ from streamflow forecasts are above the estimates of Prob$(S_T < S_T^*)$ from climatological ensembles during below normal inflow conditions and vice-versa during above normal inflow years indicating the variability in predicted summer flows. This is perfectly in line with the expectation that the probability of attaining the target storage will be low (high) during below-normal (above-normal) inflow conditions. Thus, the initial storage may ensure 100% reliable supply during the season, but the estimates of Prob$(S_T < S_T^*)$ could be utilized to invoke restrictions for improving the end of September storage conditions during below-normal years. For instance in 2002, estimates of Prob$(S_T < S_T^*)$ are between 0.7 and 0.85 from the three forecasting models with drought condition water quality releases, which suggests the need for invoking restrictions even before the beginning of summer season.

For further analysis on invoking restriction levels, only water quality releases of 100 cfs, were considered, as that corresponds to minimum acceptable release under drought

conditions for meeting the downstream flow requirements. Even with drought condition water quality releases, the streamflow forecasts suggest significant risk of falling below the target storage in comparison to the climatology, which suggest the need for restrictions on water supply releases during below-normal inflow years. Naturally, the amount of restriction on water supply could be determined based on the estimates of $\text{Prob}(S_T < S_T^*)$ suggested by each model, whose performance (in terms of reduced releases) could be validated by simulation with the observed flows. Issues related to identifying appropriate restriction levels and varying restrictions based on the estimates of $\text{Prob}(S_T < S_T^*)$ are discussed in the following sections.

4.5.3 Comparison between Multimodel Forecasts and Individual Model Forecasts

The estimates of $\text{Prob}(S_T < S_T^*)$ in Figures 4.5(a and b) differ for each streamflow forecast, as the conditional distribution in the form of ensembles exhibit different skill. The multimodel forecasts were obtained by combining forecasts from regression, semiparametric resampling models along with climatological ensembles by evaluating each forecasting model's skill from the predictor state. The main argument behind multimodel combinations based on predictor state space is that if the prediction from a particular model (including climatological ensembles) is poor during particular conditions, then it chooses the best performing model under those conditions. For additional details, see Sankarasubramanian et al. [2006] and Devineni et al. [2008].

From Figures 4.5(a and b), it is clearly seen that the estimates of Prob$(S_T < S_T^*)$ of multimodel forecasts are much closer to the estimates of Prob$(S_T < S_T^*)$ from climatology, which indicates reduced risk of falling below the target storage. But, the utility of multimodel forecasts is more apparent in years 2004 and 2005 with the observed streamflows being above normal and below normal, respectively. The estimates of Prob$(S_T < S_T^*)$ from the regression model for both years suggest a higher risk of not meeting the target storage in comparison to the climatological probabilities. On the other hand, the estimates of Prob $(S_T < S_T^*)$ from the resampling model suggest a lower risk of not meeting the target storage in comparison to the climatological estimate of Prob$(S_T < S_T^*)$ during both years. From Figure 4.4, it is clear that the modeled storages in 2004 and 2005 are above and below the target storage (131,395 acre-ft), respectively. Thus, if one invoked restriction measures based on regression forecasts in 2004 (false alarm) and ended up not invoking any restriction based on resampling forecasts in 2005 (missed target), then the individual model forecasts would have consequently advocated incorrect management measures. However, in both Figures 4.5(a and b) multimodel forecasts seem to predict the outcome correctly suggesting that the Prob$(S_T < S_T^*)$ is lower than the climatology in 2004 and the probability of $(S_T < S_T^*)$ is higher than the climatological risk in 2005. This suggests that the improved predictability of multimodel streamflow forecasts results in improved analyses of future reservoir storage conditions.

4.5.4 Enforcing Restriction Based on the Estimates of Prob $(S_T < S_T^*)$

The restriction analyses presented in this section are primarily focused on below-normal years (1991, 1992, 1993, 1994, 1997, 1998, 2002, and 2005) shown in Figure 4.5. From reservoir management perspective, as the observed flows are not realized in July, one could use the estimates of Prob $(S_T < S_T^*)$ to forecast the end of September storage conditions. For instance, if Prob $(S_T < S_T^*)$ estimated from forecasts is greater than the Prob $(S_T < S_T^*)$ estimated from climatology, then restrictions on water supply flows might be required to increase the probability of achieving the target storage. Two different approaches are suggested to invoke restrictions to improve the end of the season target storage conditions: (1) specify the restriction percent to quantify the reduced risk of not meeting the target storage (Figure 4.6) or (2) specify the desired reduction in the risk of not meeting the target storage to obtain the corresponding restriction percent (Figure 4.7).

The analyses presented here consider restrictions that could be invoked on water supply releases with the water quality flows corresponding to drought-condition release (100 cfs). Restrictions on water supply releases could be specified if forecast-based estimates of Prob $(S_T < S_T^*)$ are greater than that of climatological estimates of Prob $(S_T < S_T^*)$. Thus, if Prob $(S_T < S_T^*)$ from a given streamflow forecasting model (Figure 4.5(b)) is between 0.5 and 0.6, 0.6 and 0.7, and 0.7 and 0.8, then a restriction fraction of 10, 20, and 30% was applied on the quantified water supply release in Figure 4.2. Figure 4.6(a) shows the reduction in the estimates of Prob $(S_T < S_T^*)$ after applying restrictions based on the estimates of Prob $(S_T < S_T^*)$ in Figure 4.5(b) (no restrictions). Based on the restricted water supply

target, the difference in September storage are obtained (Figure 4.6(b)) by simulating the restricted water supply releases and water quality releases with the observed flows.

From Figure 4.6(a), it can clearly be seen that resampled flows being bootstrapped from observed flows do not show any appreciable decrease in the estimates of Prob $(S_T < S_T^*)$. It is important to note that, for year 2005, the estimate of Prob $(S_T < S_T^*)$ from the resampling model is less than that of climatology and hence the model suggests no restriction. This is indicated by the limited difference in the end of September storage, as restriction could not be applied based on the estimated Prob $(S_T < S_T^*)$, which indicates a missed target by the resampling model. As expected, differences in the end of September reservoir storage with and without restricted water supply releases clearly show that more water is stored by invoking restrictions, which should in turn improve the resilience of the system during the fall season. *For instance, the improved storage in September 2002 due to restrictions is sufficient enough to supply water for an additional 45 days in the fall season.*

Comparing the performance of different streamflow forecasts in improving the end of September storage conditions, multimodel forecasts seem to perform consistently better because of its ability to predict the below-normal storage conditions better. From Figure 4.6(b), regression suggests invoking restrictions in 2005, which is in line with the flows being below normal, but regression model suggests invoking restriction in year 2004 (see Figure 4.5(b)), which is incorrect as the observed flows belong to above-normal category. Thus, we would have invoked restrictions in year 2004, resulting in end of September storage over the 251.5 ft rule curve target level. Based on the difference in estimates of Prob

$(S_T < S_T^*)$ and the end of September storage conditions obtained by invoking different restriction levels, multimodel streamflow forecasts appear to reduce false alarms by suggesting restrictions only during below-normal inflow years.

A different approach to invoke restrictions is by specifying the desired reduction in the estimate of $\text{Prob}(S_T < S_T^*)$. Figure 4.7 provides the suggested restriction percentage from each model (Figure 4.7(a)) for 5% reduction in the estimate of $\text{Prob}(S_T < S_T^*)$. Figure 4.7(b) shows the difference in the end of September storages between restricted and unrestricted water supply releases for each forecasting model upon simulating with the observed flows. The advantage of this approach is that the restriction percentages are actually specified by the model based on the desired level of reduction in the risk of not meeting the target storage. From Figure 4.7(a), we can clearly see that the multimodel streamflow forecasts seem to advocate restrictions that are in between the restriction percentages suggested by resampling and regression models (except 1994 and 2002). For instance, in year 2004, though the flow is below normal, it is very close to the 0.33 percentile (147 cfs), thus the suggested restriction percentage is less than the restriction suggested by regression and resampling models. Figure 4.7(b) illustrates that the difference in end of September storage from the multimodel varies consistently to the change in streamflow potential. For example, in year 2003, the average streamflow is only 1.7 cfs, and the restriction suggested by the multimodel is much closer to the resampling model. Of course, the initial storage available in July also plays an important role in estimating the $\text{Prob}(S_T < S_T^*)$, but multimodel forecasts, obtained by combining forecasts from resampling and regression models reduce the model uncertainty and improve

the confidence on streamflow forecasts based water allocation by reducing the number of false alarms and missed targets.

4.6 Discussion

In developing seasonal water allocation policies, initial storages may ensure 100% reliability of supplying target releases for the intended uses, thereby limiting the utility of climate forecasts. But, ensuring the end of the season target storage (or the operational rule curve) will be met with high probability could offer additional insights for invoking the appropriate level of restrictions during below-normal inflow years. Further, as the water demand increases over the service area (due to urbanization and population growth), the initial storage may no longer ensure 100% reliability, which will necessitate the application of climate forecasts for invoking restrictions. During above-normal inflow years, as the forecasts based $\text{Prob}(S_T < S_T^*)$ will be lower than its climatological probability, forecasts based allocation would avoid unnecessary restrictions if the initial storage is lower than the operational rule curve. On the other hand, if the initial storage is higher than the operational rule curve, then additional release could be considered to reduce the downstream flood risk such that the forecasts based estimates of $\text{Prob}(S_T < S_T^*)$ are equal to its climatological probability.

The retrospective analysis presented in this study could also be utilized to determine the appropriate beginning of the season storage under future increased demand scenarios. Using climatological ensembles, one can estimate the increased beginning of the season

storage, S_0, which needs to ensure the current climatological $\text{Prob}(S_T < S_T^*)$ will remain unchanged even under future release scenarios. Similarly, the proposed formulation also could be utilized to develop rule curves that change according to the inflow potential. For instance, in Figure 4.6 it is shown that by restricting reservoir releases during below-normal years, the $\text{Prob}(S_T < S_T^*)$ could be increased. To develop rule curves for this scenario, one can specify $S_T = S_T^*$ and obtain previous month target storages that will ensure the restricted releases during the season. It is also important that these rule curves need to be updated regularly based on the updated climate information, which is important towards better prediction of intraseasonal variability in streamflows [Sankarasubramanian et al. 2008].

The main advantage in utilizing multimodel forecasts is in reducing model uncertainty by constituting ensembles from multiple models. In the multimodel combination scheme of Devineni et al. [2008], higher weight is given to the individual model that performs well under similar predictor conditions. For instance, if an individual model performs better during El Nino conditions, then higher number of ensembles is drawn from that particular model under similar predictor conditions. By combining individual models with climatology, one can reduce the overconfidence in individual model forecasts to develop multimodel forecasts that have reduced false alarms and missed targets [Devineni et al. 2008]. This study clearly shows that employing such multimodel forecasts for season-ahead water allocation provides a more reliable way to develop appropriate management strategies such as invoking (or not invoking) restrictions during below-normal (above-normal) years. Future studies on climate forecasts application will focus on better

management of water supply systems under increased demand potential without resorting to capacity expansion and investments on new systems by considering alternate water uses (e.g., reclaimed water) and trading.

4.7 Summary and Conclusions

A reservoir simulation model that uses ensembles of streamflow forecasts is presented and applied for allocating water during the summer season (JAS) from the Falls Lake Reservoir in the Neuse River Basin, N.C. Given the initial storage at the beginning of the season and ensembles of seasonal streamflow forecasts, the simulation model can estimate the reliability of the specified target releases and the end of the season target storage probability.

The customized simulation model for Falls Lake was analyzed using JAS seasonal streamflow forecasts from three models: *parametric regression, semiparametric resampling, and multimodel forecasts (obtained from the former two models*). The performance of these three models in estimating $\text{Prob}(S_T < S_T^*)$ was evaluated by comparing with the estimates of $\text{Prob}(S_T < S_T^*)$ from climatological ensembles to predict below-normal storage conditions, which could help in invoking restrictions for improving storage conditions at the end of the summer season.

Analyses of Falls Lake using the simulation model without constraining the end of season target storage showed 100% reliability of meeting target releases, implying that the entire seasonal demand could be met purely based on initial storage. This invalidated the utility of streamflow forecasts available for the summer season. However, by constraining the

system to meet the end of the season target storage, it is clearly shown that the estimates of $\text{Prob}(S_T < S_T^*)$ from forecasts are higher than the climatology estimates during below-normal summer inflow years and vice versa during above-normal inflow years, thereby indicating the utility of forecasts in invoking restrictions. By invoking restrictions during JAS based on the predicted estimates of $\text{Prob}(S_T < S_T^*)$, the study shows that, upon validating with JAS observed flows, increased storage conditions result in September. Among the three streamflow forecasting models, multimodel streamflow forecasts seem to better predict the change in streamflow potential, thus resulting in reduced false alarms and missed targets in predicting below-normal storage conditions at the end of September. Thus, applying multimodel forecasts would reduce uncertainty from individual models which could lead to better decisions and also could improve public confidence in utilizing seasonal streamflow forecasts for water management application.

4.8 References

Arumugam, S., Sharma, A., and Lall, U. (2003). "Water allocation for multiple uses based on probabilistic reservoir inflow forecasts." *Proc., Int. Association of Hydrological Sciences*, International Association of Hydrological Science, Wallingford, U.K.

Barlow, M., Nigam, S., and Berbery, E. H. (2001). "ENSO, Pacific decadal variability, and US summertime precipitation, drought, and stream flow." *J. Clim.*, 14(9), 2105–2128.

Barnston, A. G., Mason, S. J., Goddard, L., DeWitt, D. G., and Zebiak, S. E. (2003). "Multimodel ensembling in seasonal climate forecasting at IRI." *Bull. Am. Meteorol. Soc.*, 84, 1783–1796.

Dettinger, M. D., and Diaz, H. F. (2000). "Global characteristics of stream flow seasonality and variability." *J Hydrometeorol.*, 1(4), 289– 310.

Devineni, D., Sankarasubramanian, A., and Ghosh, S. (2008). "Multimodel ensembling of streamflow forecasts: Role of predictor state in developing optimal combinations." *Water Resour. Res.*, 44(9), 1–22.

Doblas-Reyes, F. J., Deque, M., and Piedelievre, J. P. (2000). "Multimodel spread and robabilistic seasonal forecasts in PROVOST." *Q. J. R. Meteorol. Soc.*, 126(567), 2069–2087.

Georgakakos, A. P., Yao, H. M., Mullusky, M. G., and Georgakakos, K. P. (1998). “Impacts of climate variability on the operational forecast and management of the upper Des Moines River basin.” *Water Resour. Res.*, 34(4), 799–821.

Guetter, A. K., and Georgakakos, K. P. (1996). “Are the El Niño and La Niña predictors of the Iowa River seasonal flow?” *J. Appl. Meteorol.*,35(5), 690–705.

Hamlet, A. F., and Lettenmaier, D. P. (1999). “Columbia River streamflow forecasting based on ENSO and PDO climate signals.” *J. Water Resour. Plann. Manage.*, 125(6), 333–341.

Hamlet, A. H., Huppert, D., and Lettenmaier, D. P. (2002). “Economic value of long-lead streamflow forecasts for Columbia River hydropower.” *J. Water Resour. Plann. Manage.*, 128(2), 91–101.

Hartmann, H. C., Pagano, T. C., Sorooshian, S., and Bales, R. (2002). “Confidence builders—Evaluating seasonal climate forecasts from user perspectives.” *Bull. Am. Meteorol. Soc.*, 83(5), 683–698.

Kyriakidis, P. C., Miller, N. L., and Kim, J. (2001). “Uncertainty propagation of regional climate model precipitation forecasts to hydrologic impact assessment.” *J Hydrometeorol.*, 2(2), 140–160.

Leung, L. R., Hamlet, A. F., Lettenmaier, D. P., and Kumar, A. (1999). "Simulations of the ENSO hydroclimate signals in the Pacific Northwest Columbia River basin." *Bull. Am. Meteorol. Soc.*, 80(11), 2313– 2329.

Lyon, B., Christie-Blick, N., and Gluzberg, Y. (2005). "Water shortages, development, and drought in Rockland County, New York." *J. Am. Water Resour. Assoc.*, 41(6), 1457–1469.

Pagano, T. C., Hartmann, H. C., and Sorooshian, S. (2001). "Using climate forecasts for water management: Arizona and the 1997–1998 El Nino." *J. Am. Water Resour. Assoc.*, 37(5), 1139–1153.

Piechota, T. C., Chiew, F. H. S., Dracup, J. A., and McMahon, T. A. (2001). "Development of exceedance probability streamflow forecast." *J. Hydrol. Eng.*, 6(1), 20–28.

Piechota, T. C., and Dracup, J. A. (1996). "Drought and regional hydrologic variation in the United States: Associations with the El Nino Southern Oscillation." *Water Resour. Res.*, 32(5), 1359–1373.

Regonda, S. K., Rajagopalan, B., Clark, M., and Zagona, E. (2006). "A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin." *Water Resour. Res.*, 42(9), 1–14.

Ropelewski, C. F., and Halpert, M. S. (1987). “Global and regional scale precipitation patterns associated with the El-Nino southern oscillation.” *Mon. Weather Rev.*, 115(8), 1606–1626.

Sankarasubramanian, A., Devineni, N., and Ghosh, S. (2006). “Multi model ensembling based predictor state space: Seasonal forecasts and climate relation.” *Miniseries*, Institute of Statistics North Carolina State Univ., Raleigh, N.C.

Sankarasubramanian, A., Lall, U., and Espuneva, S. (2008). “Role of retrospective forecasts of GCMs forced with persisted SST anomalies in operational streamflow forecasts development.” *J Hydrometeorol.*, 9(2), 212–227.

Schaake, J., and Larson, L. (1998). “Ensemble streamflow prediction (ESP): Progress and research needs.” *Special Symp. on Hydrology*, American Meteorological Society, Boston, J19–J24, 1998.

Sicard, E., Sabatier, R., Niel, H., and Cadier, E. (2002). “A new approach in space-time analysis of multivariate hydrological data: Application to Brazil’s Nordeste region rainfall.” *Water Resour. Res.*, 38(12),1319.

Souza, F., and Lall, U. (2003). “Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate.” *Water Resour. Res.*, 39(11), 1–13.

Steinemann, A. C. (2006). “Using climate forecasts for drought management.” *J. Appl. Meteorol.*, 45(10), 1353–1361.

Vorosmarty, C. J., Green, P., Salisbury, J., and Lammers, R. B. (2000). “Global water resources: Vulnerability from climate change aci population growth.” *Science*, 289(5477), 284–288.

Weaver, C. J. (2005). “The drought of 1998–2002 in North Carolina– Precipitation and hydrologic conditions.” *USGS Scientific InvestigationsRep.*, Washington, D.C.

Yao, H., and Georgakakos, A. (2001). “Assessment of Folsom Lake response to historical and potential future climate scenarios. 2—Reservoir management.” *J. Hydrol.*, 249(1–4), 176–196.

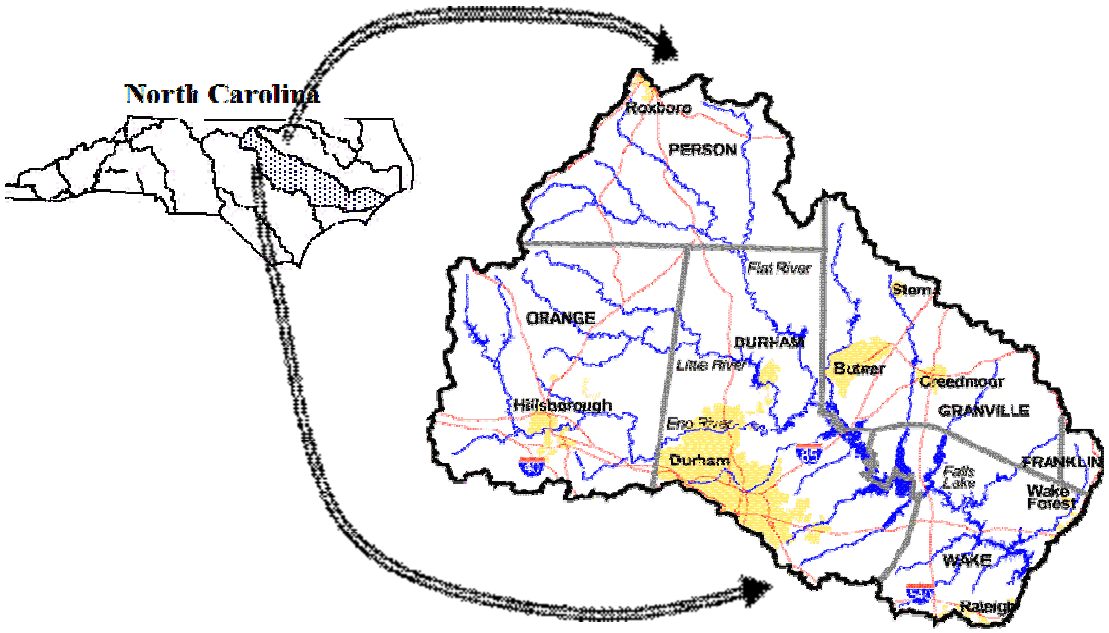

Figure 4.1: Location of Neuse River basin and Falls Lake Reservoir in the upper Neuse river basin.

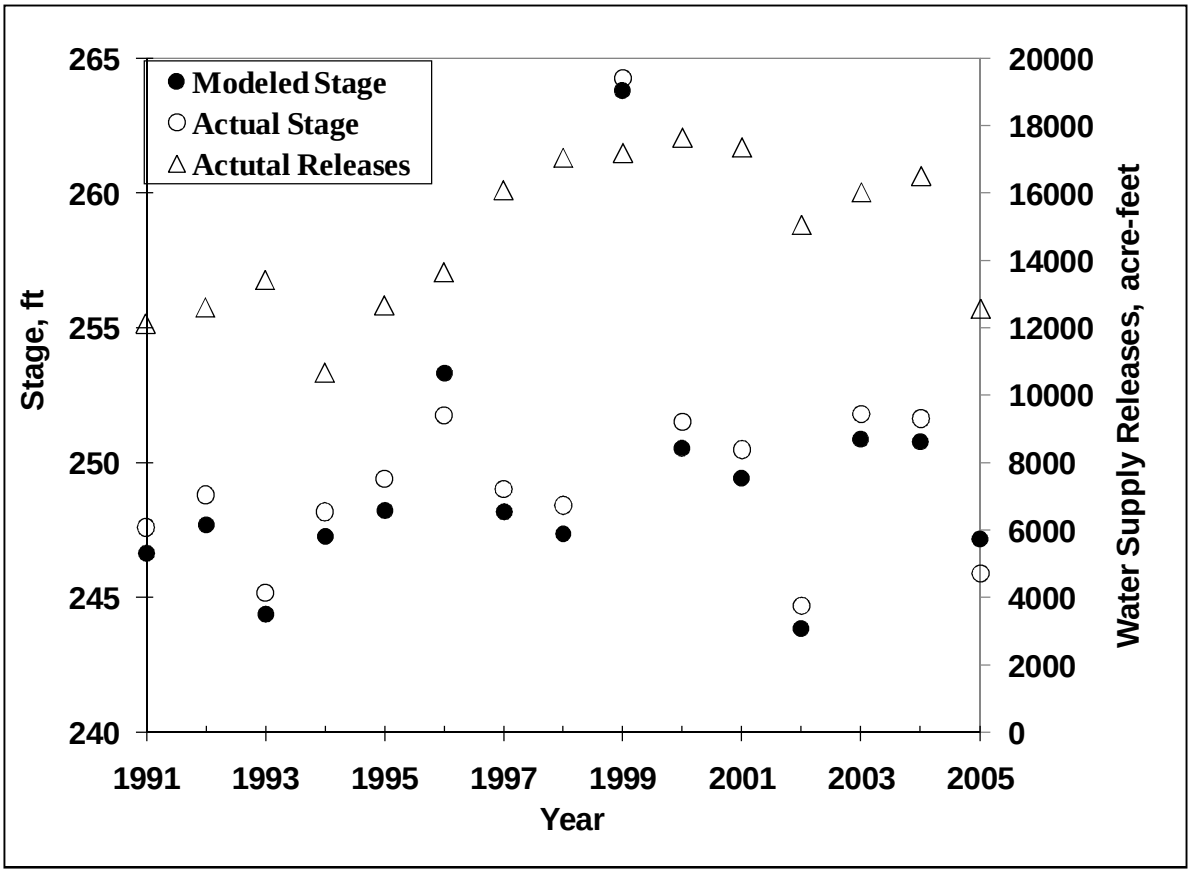

Figure 4.2: Comparison of modeled stages with the observed stages in September for the period 1991-2005. Figure also shows the reported water supply releases during JAS from Falls Lake. Modeled stages are obtained upon simulating the model with observed flows, releases and by forcing the model with the initial storage recorded each year on July 1.

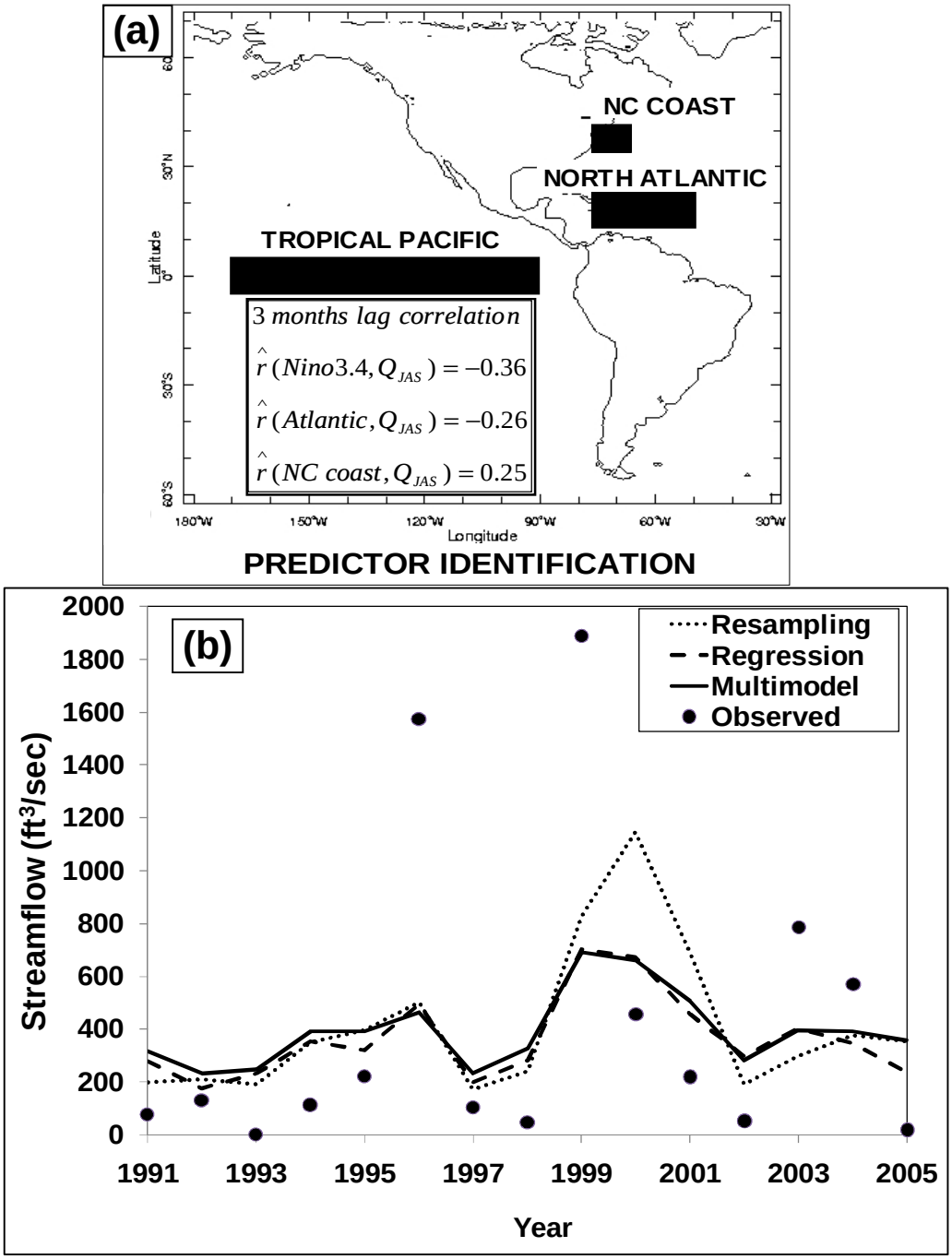

Figure 4.3: Leave-one out Cross-validated seasonal (JAS) streamflow forecasts for the Falls Lake from three forecasting models (Figure 4.3b -Regression, Resampling and Multimodel) along with the with the employed predictors (Figure 4.3a). Principal components of the three SST regions (shown as rectangles in Figure 4.3a) were performed and the dominant two components were employed for developing the streamflow forecasts represented in the form of ensembles.

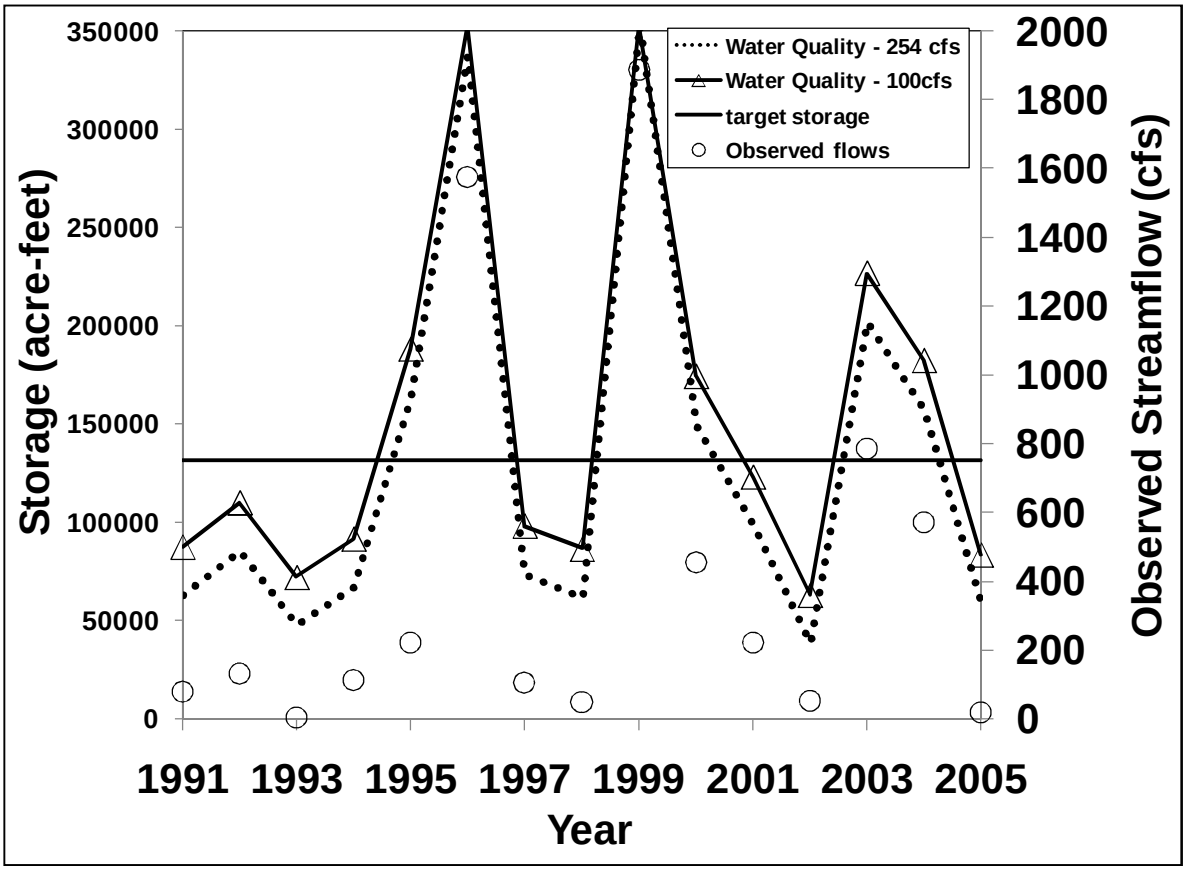

Figure 4.4: Modeled storages for two water quality release scenarios: Normal (254 cfs) and Drought (100 cfs) conditions along with the observed streamflows and the target storage (solid horizontal line). The storages shown are obtained by combining the observed streamflows with the chosen water quality release and the corresponding year's summer water supply release in Figure 4.2.

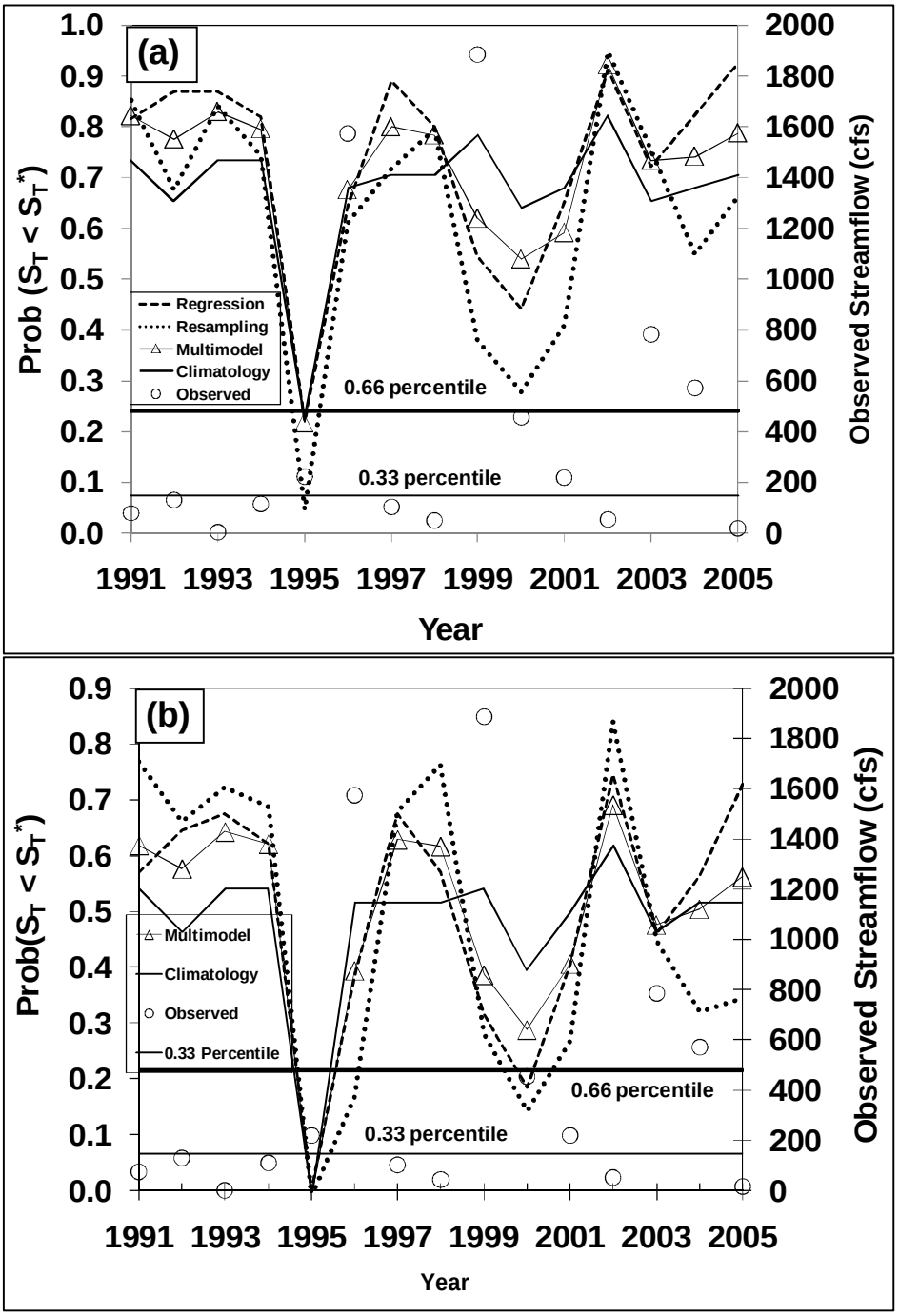

Figure 4.5: Role of Streamflow Forecasts in predicting the end of the season target storage that corresponds to the stage of 251.5 feet, m.s.l. (a) Normal water quality release of 254 cfs (b) Restricted Water quality release of 100 cfs. Note that the multimodel forecasts suggest clearly increased risk of not meeting the target storage in comparison to the risks suggested by the climatology during below-normal years.

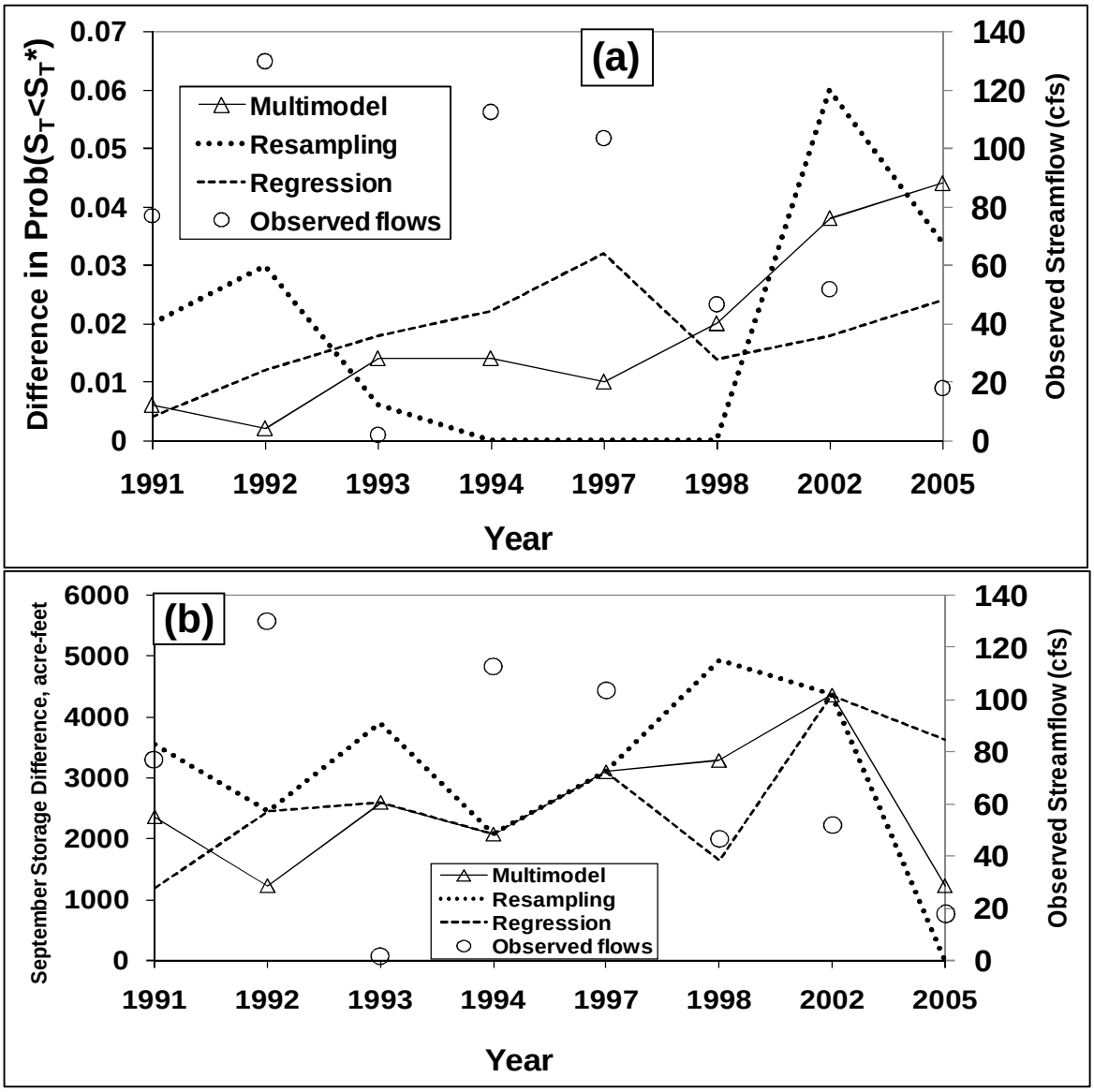

Figure 4.6: Performance of streamflow forecasts in reducing the risk of not attaining end of the season target storage (Figure 4.6a) under different restriction levels and in improving the end of the season target stage (upon validating with observed flows) (Figure 4.6b). September storage difference is obtained from the simulated additional September storage that would have occurred under restricted water supply flows by simulating with the observed flows. Restrictions are obtained based on the estimates of Prob($S_T \leq S_T^*$) with 10% , 20% and 30% restriction if the estimates of Prob($S_T \leq S_T^*$) are between 0.5-0.6, 0.6-0.7 and 0.7-0.8 respectively. All the years shown here are below-normal years.

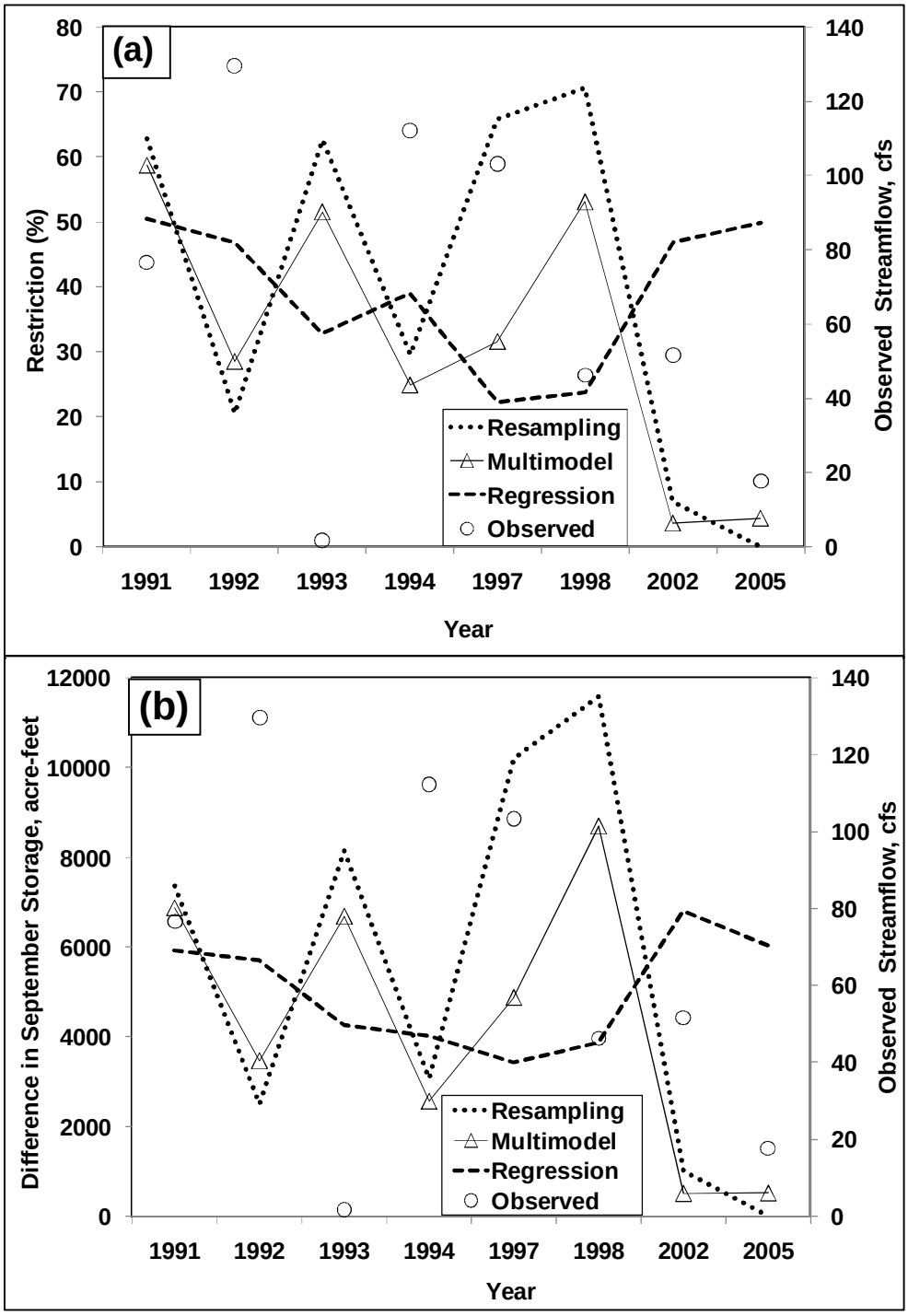

Figure 4.7: Performance of streamflow forecasts in suggesting restriction for the prescribed level of reduction in the risk (5%) of not attaining end of the season target storage (Figure 4.7a) and in improving the end of the season target stage (Figure 4.7b) (upon validating with observed flows) for the obtained restriction in Figure 4.7a.

CHAPTER 5

Climatology of Monthly Runoff: Causality and Relations to Seasonality in Precipitation and Temperature

5.1 Introduction

The partitioning of precipitation into evaporation and streamflow is a key aspect for understanding the hydroclimatology of the region. For instance, partitioning of precipitation into runoff and evaporation based on the climatic index of the basin is critical in understanding the role of annual moisture and energy balance over the basin. Shifts in the seasonality of precipitation and temperature can also lead to significant changes in the time of occurrence of peak streamflow in the basin. Another attribute related to the regional hydroclimatology, interannual variability in precipitation and runoff, determine the type of reservoir systems in the region. For instance, most of the reservoirs in the eastern U.S. are within-year reservoir systems, since the interannual variability in precipitation and runoff is smaller. On the other hand, the reservoir systems in the western U.S are over-year systems (i.e., carry deficit/excess supply from one year to another) due to large interannual variability.Therefore, understanding the controls of mean monthly/annual runoff rate plays a significant role in water resources planning and management.

Numerous studies have focused on developing physical relationship to explain the annual hydroclimatology of a region [Budyko, 1974; Milly, 1994a, 1994b; Zang et al. 2001; Schreiber, 1904; Ol' dekop, 1911; Wolock and McCabe, 1999]. Milly [1994a] incorporated the seasonality in precipitation and potential evapotranspiration on the annual water balance. Woods [2003] has developed analytical models for seasonality and annual water balance of a catchment. Potter et al. [2005] and Hickel and Zang [2006] have also analyzed the impact of rainfall seasonality on mean annual water balance. However, limited effort was made in understanding the effect of seasonality of climate on mean monthly runoff. Farmer et al.

[2003] explored the role of climate and landscape interactions in explaining the annual, monthly and daily water balance of several arid basins. Yokoo et al. [2008] explored the effects of seasonal variability of climate on mean annual and monthly water balances based on a physically based hill-slope model.

This study is primarily focused on understanding the influence of seasonality of precipitation and temperature in explaining the monthly runoff which is heavily dependent on the interplay between the climatic inputs and the storage characteristics of the basin. A simple water balance model is used for this purpose without adopting any automatic calibration methods. Results from the model analysis are used to explain the seasonality of observed streamflow over 1373 watersheds in the continental United States (U.S.). Section 5.2 describes the hydroclimatic database employed in the study and the motivation behind the work. The methodology and preliminary results are presented in sections 5.3 and section 5.4 respectively. Finally, the summary and findings from the study are presented in section 5.5.

5.2 Data Description and Motivation

5.2.1 Streamflow Database

The hydro-climatologic data network (HCDN) developed by Slack et al. [1993], available from the U.S. Geological Survey is used for this study. This data set contains records of average daily streamflow from 1659 locations in the United States and its territories. The streamflow records range from 1874 to 1988, with an average record length of 44 years. Streamflow from the HCDN database are least affected by anthropogenic influences such as land use changes and ground water retrieval and the accuracy ratings were

at least "good" for all the records. In this study, we used 1373 basins in the 48 contiguous states from the continental U.S. Figure 5.1 shows the aridity index R, which is defined as the ratio of mean annual potential evapotranspiration to precipitation $\overline{PET}/\overline{P}$ for the HCDN basins. Five broad regions are identified based on their moisture (precipitation) and energy (temperature) availability (Moisture and Energy Availability Scenarios, MEAS). Details of the scenarios and the analyses are presented in section 5.4.

5.2.2 Precipitation and Potential Evapotranspiration Database

Sankarasubramanian and Vogel [2005] developed a precipitation and temperature database for the HCDN sites. The precipitation-elevation regressions on independent slopes model (PRISM) climate analysis system described by Daly et al. [1994] was used to obtain 37-year time series of monthly precipitation and average minimum and average maximum daily temperature using 0.5° time series grids. Sankarasubramanian and Vogel [2005] also obtained the monthly potential evapotranspiration using the monthly temperature and extraterrestrial solar radiation based on Hargreaves method [Hargreaves and Samani, 1982].

5.2.3 Seasonality in Observed Precipitation and Streamflow – Motivation

Seasonality Index (SI), a measure used to quantify the seasonality in a chosen attribute, is used to investigate the relationship between the seasonality in precipitation and the seasonality in observed streamflow. SI quantifies the average time of occurrence as well as the degree to which the event tends to be concentrated in a given month [Markham, 1970]. SI of 0 indicates that there is no seasonal characteristic for that attribute with its occurrence

being uniform throughout the year. Alternately, if the SI is close to 1, it indicates that the attribute is highly seasonal and its occurrence is concentrated in a particular month of the year. Details on estimating the SI for a hydroclimatic variable is given in the Appendix C.

From figure 5.2a, we can observe that the basins in the western U.S. have high SI for both precipitation and streamflow indicating that both these attributes are highly seasonal. While the average time of occurrence for precipitation (Figure 5.2b) in these basins is during January, streamflow has an average time of occurrence during February and March, which indicates the seasonality in streamflow being influenced by the seasonality in winter precipitation. However, this phenomenon is not observed in the basins over the eastern U.S. The SI of precipitation across the eastern U.S is small (0 – 0.2) indicating that the precipitation is uniformly distributed across all the months. Observed streamflow has a SI of 0.2 – 0.4 indicating that streamflow in these regions is exhibiting a seasonal behavior with a peak around February - April for most of the stations. It is important to note that temperature always has strong seasonality. High temperature during the summer not only shows increased evaporation, but also indicates that the evaporation is lesser during other months. Other factors such as the basin storage capacity and form of precipitation could also contribute to the amount and the timing of streamflow.

5.3 Methodology

The main objective of this paper is to explain the role of seasonality in precipitation and temperature in influencing the seasonality in streamflows. For this purpose, we employed Thornthwaite monthly water balance model [Thornthwaite and Mather, 1955] and forced it

with climatic inputs, precipitation and potential evapotranspiration, which are obtained from a parametric form to give us the flexibility to observe the basin responses under various scenarios of seasonal climatic variations.

5.3.1 Climate Seasonality

Seasonally varying climatic inputs are synthetically generated by assuming that both P and PET are periodic. In other words, the temporal variations in the moisture and energy supplies are assumed to be sinusoidal with certain random variability in the components. They are represented by the following equations.

$$P(t) = \bar{P}(1 + d_p \sin(w_p t + a_p)) + e_t \quad \ldots (5.1)$$

$$PET(t) = \overline{PET}(1 + d_{pet} \sin(w_{pet} t + a_{pet})) + e_t \quad \ldots (5.2)$$

Where $\bar{P}$ = average precipitation (mm/month), $\overline{PET}$ = average PET (mm/month)

R = Aridity Index, δ_p=Amplitude of P, δ_{pet}=Amplitude of PET

with $0 \leq \delta_p \leq 1$; $0 \leq \delta_{pet} \leq 1$;

$\omega_p = \omega_{pet} = 2\pi/T$, for T = 12 months, is the angular frequency

α_p *and* α_{pet} phase angle used to determining the extreme month in P and PET

$\varepsilon_t \sim N(0, 0.25\bar{P})$ follow normal distribution.

In addition to this, we define $\Delta\alpha = \alpha_p - \alpha_{pet}$ as the phase difference between P and PET. If $\Delta\alpha = 0$, P and PET are in phase. Alternately, if $\Delta\alpha = \pi$, P and PET are out of phase. Table 5.1 shows the assumed values of mean annual precipitation and standard deviation of annual precipitation for various types of basins depending on the aridity index. The annual potential evapotranspiration for a basin is obtained by multiplying the annual precipitation with its aridity index. To remove the effect of arbitrary initial conditions, we generated 50 years of monthly P and PET data for a given basin using equations (5.1) and (5.2).

5.3.2 Moisture and Energy Availability Scenarios (MEAS) Considered

Based on the seasonality of observed P and PET, we developed five different scenarios (MEAS 1 – MEAS 5) that encompass the behaviors of most of the basins from the HCDN database (Figure 5.1). Table 5.2 has the list of these scenarios and the regions they represent. MEAS 1 (MEAS 2) corresponds to humid (arid) basins where precipitation exhibits strong seasonality and is out of phase with temperature. In other words, for these basins one can expect wet and cold winters followed by dry and hot summers. MEAS 3 (MEAS 4) corresponds to humid (arid) basins where precipitation is uniformly distributed throughout the year. MEAS 5 corresponds to arid basins where P and PET are in phase with a strong seasonality in P. Dry winters are followed by hot and wet summers in these regions.

Seasonality in the simulated precipitation and temperature can be described by their respective amplitudes (δ_p and δ_{pet}). For example, high seasonality in precipitation can be achieved by assuming δ_p equal to 1. Alternately, one can assume δ_p equal to 0 to simulate a case where the precipitation has no seasonality. Timing, or the month of occurrence for the

climatic inputs are achieved through the phase angles α_p and α_{pet}. The underlying assumption in our simulations is that PET will exhibit strong seasonality with a peak during the month of July. For MEAS 1 and MEAS 2 the time of occurrence of peak precipitation is chosen in a way that it reflects the time of occurrence of observed precipitation of the basins in that region (January in this case). The simulated P and PET are forced with the water balance model to estimate the monthly streamflow, evaporation and basin storage. The next section discusses the interactions of P, PET and basin storage in explaining the seasonality of monthly streamflow based on the outcomes of the simulations.

5.4 Results and Analysis

Scenarios of climatic inputs are simulated and forced with the Thornthwaite monthly water balance model. This section presents the results from the model simulations in order to illustrate the effects of underlying climate interactions and catchment storage relationships with monthly streamflow.

5.4.1 MEAS 1 and MEAS 2: P and PET are out of phase with strong seasonality in P

Figure 5.3 shows the SI of estimated streamflow and catchment storage for scenarios MEAS 1 and MEAS 2. While R < 1 represents the outcomes for humid and semi-humid basins, R > 1 shows the results for temperate basins. Climate inputs are simulated such that P and PET are out of phase, and that P exhibits a strong seasonality (achieved by assuming the value of δ_p as 1) with the peak occurring in January. We assume that small (large) amplitude in the sinusoidal function can simulate cases where the variability in monthly precipitation is

low (high). Since SI is related to the intra-annual variability of a hydrologic quantity, δ_p close to zero (one) indicates low (high) seasonality. For example, the SI of simulated P with a δ_p of 1 is around 0.5 and the SI with a δ_p of 0.1 is around 0.09, indicating a very high and low seasonality respectively.

From Figure 5.3, we can see that for humid basins ($R \leq 0.25$), when P and PET are out of phase, the seasonality in storage is not significant. This behavior can be attributes to the soil moisture holding capacity of the basin. Since the annual precipitation is significantly greater than the annual potential evapotranspiration of the basin, the soil will be saturated or full to capacity throughout the year. Hence the basin does not experience any significant variability in storage, leading to a low value of SI. We can see that the seasonality of storage increases as the aridity index of the basin increases. In other words, there is a pronounced seasonality in storage for semi humid to arid basins. As the aridity index increases, the energy supply for the basin increases causing an increase in the evaporation. Since the PET has strong seasonality, summer evaporation will be high. Further, high seasonality in precipitation causes winter dominant rainfall and dry summers. The combined effect of no rainfall and maximum evaporation during summer leads to depletion of soil storage, thereby introducing a seasonal variability in storage.

We can also see that there is a strong seasonality in the streamflow from the basins. Streamflow occurs when the basin is full to capacity, or when it can no longer hold water. Seasonality in streamflow can be explained by understanding the interplay between winter precipitation and summer evaporation. Seasonality in P causes maximum rainfall during winter and less rainfall during summer. Further, a strong seasonality in PET causes more

evaporation from the basin in summer compared to winter. This phenomenon introduces a variability or large difference in streamflows of the two seasons, causing a high SI in streamflow. As the basin gets less humid, the seasonality of streamflow becomes pronounced. We can also understand that peak streamflow occurs during the winter season, as the summers experience very little flows. Hence we can say that the seasonality of PET or the amount of evaporation during summer plays a vital role in governing the streamflow of the basin.

5.4.2 MEAS 3 and MEAS 4: P has no seasonality

Figure 5.4 shows the SI of estimated streamflow and storage for scenarios MEAS 3 and MEAS 4. Climate inputs are simulated such that P has no seasonality or is uniform throughout the year (achieved by assuming the value of δ_p as 0). We can see that the seasonality of streamflow is not significant when the P has no seasonality for very humid basins. When rainfall occurs uniformly throughout the year in basins with soil moisture full to capacity most of the times, the basin yields excess runoff during the rainfall events. Hence there is no significant difference between winter and summer streamflows, causing the seasonality of streamflow for very humid basins to be negligible. However as the aridity index of the basin increases, we observe that the streamflow and storage exhibits a strong seasonality. The energy supply for the basin increases with increasing aridity index, causing an increase in the evaporation during summer. This will cause all the precipitation in summer to leave the basin as evaporation. Hence the basin experiences high amount of runoff during the winter season and less runoff and a depletion of soil store during the summer season,

leading to a high variability in the runoff. Therefore, we can see that the SI of streamflow is high (with a peak occurring during the winter season), though the SI of P is low.

Based on the analysis presented above, we try to understand the behavior of the 1373 HCDN basins in the continental U.S. We categorized basins with aridity index less than 1 as humid basins. These can be further divided as humid basin ($R \leq 0.5$) and semi-humid basins ($0.5 \leq R \leq 1$). Basins over the north western U.S. particularly in the states of Washington and Oregon are humid basins (Figure 5.1). Similarly, the basins over North Eastern U.S., i.e. the states of Vermont, Maine, New York and some basin over the west of Rocky Mountains, northern California and over the Appalachians are examples of semi-humid basins. Basins with aridity index greater than 1 are considered arid basins or dry basins. Potential evapotranspiration exceeds the precipitation in such basins and actual evaporation approaches precipitation. Most of the basins over Mid-Atlantic, South Atlantic, Great Lakes, Ohio, Tennessee, Upper and Lower Mississippi are arid basins.

From, figure 5.2, we can see that precipitation has no pronounced seasonality (SI between 0 – 0.2) for most of the New England region. We can identify this to be a case with no seasonality in P in semi humid basins ($0.5 < R < 1$), MEAS 3. Similarly, basins over the Mid-Atlantic, Ohio, and Tennessee are arid basins with no seasonality in precipitation (MEAS 4). From figure 5.4 we can understand that the streamflow will exhibit a strong seasonality with dominance around winter. Figure 5.2b shows us the seasonality and time of occurrence for observed streamflow for these regions. We can see that there is a strong seasonality in streamflow with the peak streamflows occurring in winter. There are few basins over the North East and over the Appalachian where the streamflow peak occurs

around April. The delay in the occurrence of streamflow in these basins can be attributed to melt season, since most of these basins are driven by snowpack. Similarly basins over the Pacific Northwest, particularly the states of Washington, Oregon are very humid basins ($R < 0.5$) with strong seasonality in winter precipitation. Hence these basins can be categorized as scenario MEAS1. From figure 5.3, we can say that the streamflow will exhibit a strong seasonality with a peak around winter.

Similar analyses are currently being conducted on MEAS 5, i.e. the peninsular Florida and some very arid basins over the Mid-continent, i.e. the Texas Gulf and Arizona. When P and PET are in phase, the basin typically experiences dominant rainfall during summer and less rainfall during winter. Strong seasonality in P indicates that all the rainfall is occurring during summer months with only a small amount of rainfall during the winter months. Hence, the basin would experience runoff only during the summer months, causing a high seasonality in streamflow. Further, with a strong seasonality in P, we can expect the time of occurrence of streamflow to be shifted towards summer seasons. Such model based analyses will provide insights into understanding and reliably predicting the runoff from basins without any calibration. Hence, it is important to understand the role of climate and catchment characteristics in governing the basin responses.

5.5 Summary and Findings

The current study investigates the role of seasonality of precipitation and temperature in explaining the climatology of monthly streamflow. This is achieved by forcing simulated climate inputs with a simple water balance model that requires no calibration. Observed

differences in monthly climatology of precipitation and streamflow for the humid and arid basins can be attributed to variability in precipitation, evaporation and basin storage.

When precipitation and temperature are out of phase, we understand that the runoff primarily dependents on the seasonality of precipitation. There is a strong seasonality in streamflow with dominance during winter. In other words, there will be a distinct wet and dry season. As the aridity index of the basin increases, the winters are dominated by accumulation of soil moisture storage and significant winter runoff. Summers are very dry leading to no flows and emptying of basin. When precipitation has no seasonality, the runoff will be dependent on the aridity index of the basin. Very humid basins will experience runoff in all the seasons. Hence, the SI of the streamflow is not significant. As the dryness increases, the evaporation will increase during summer leading to a winter dominant runoff with strong seasonality in streamflow for semi humid, temperate and arid basins. Based on these findings, our future work will explore developing relationships to estimate mean monthly runoff purely based on the monthly climatology of precipitation, potential evapotranspiration and storage capacity of the basin.

5.6 References

Budyko, M. I. (1974), Climate and Life, translated from Russian by D. H. Miller, Academic, San Diego, Calif.

Daly, C., R. P. Neilson, and D. L. Philips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, 33(2), 140–158.

Duffie, J. A., and W. A. Beckman (1980), Solar Engineering of Thermal Processes, 109 pp., John Wiley, New York.

Farmer, D., M. Sivapalan and C. Jothityangkoon (2003). Climate, soil and vegetation controls upon the variability of water balance in temperate and semi-arid landscapes: Downward approach to hydrological prediction. Water Resources Research, Vol. 39, No. 2, 1035, doi: 10.1029/2001WR000328.

Hargreaves, G. H., and Z. A. Samani (1982), Estimating potential evapotranspiration, *J. Irrig. Drainage Eng.*, 108(3), 225–230.

Hickel, K. and L. Zhang (2006) Estimating the impact of rainfall seasonality on mean annual water balance using a top-down approach. *Journal of Hydrology* **331**(3-4): 409-424.

Jensen, M .E., R. D. Burman, and R. G. Allen (1990), Evapotranspiration and irrigation water requirements, Manuals and Rep. on Eng. Practice 70, 350 pp., Am. Soc. of Civ. Eng., New York.

Markham, C. G. (1970), Seasonality of Precipitation in United-States, *Annals of the Association of American Geographers* **60**(3): 593-597.

Milly, P. C. D. (1994a), Climate, interseasonal storage of soil water and the annual water balance, *Adv. Water Resour.*, 17, 19–24.

Milly, P. C. D. (1994b), Climate, soil water storage, and the average annual water balance, *Water Resour. Res.*, 30(7), 2143–2156.

Ol'dekop, E. M. (1911), On evaporation from the surface of river basins, *Transactions on Meteorological Observations*, Lur-evskogo, Univ. of Tartu, Tartu, Estonia.

Pike, J. G. (1964), The estimation of annual runoff from meteorological data in a tropical climate, *J. Hydrol.*, 2, 116–123.

Potter, N. J., L. Zhang, et al. (2005), Effects of rainfall seasonality and soil moisture capacity on mean annual water balance for Australian catchments, *Water Resources Research* **41**(6).

Schreiber, P. (1904), Ueber die Beziehungen zwischen dem Niederschlag und der Wasseruhrung der Flusse in Mitteleuropa, Meteorol. Z., 21, 441–452.

Slack, J. R., A. M. Lumb, and J. M. Landwehr (1993), Hydroclimatic data network (HCDN): A U.S. Geological Survey streamflow data set for the United States for the study of climate variation, 1874–1988, *Water Resour. Invest*. Rep., 93-4076.

Thornthwaite, C. W., and J. R. Mather (1955), The water balance, Publ. Climatol. Lab. Climatol. Drexel Inst. Technol., 8(1), 1-104.

Vogel, R. M., and A.Sankarasubramanian (2005), .USGS Hydro-Climatic Data Network (HCDN): Monthly Climate Database, 1951-1990.Data set available on-line from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, U.S.A.

Wolock, D. M., and G. M. McCabe (1999), Explaining spatial variability in mean annual runoff in the conterminous United States, *Clim. Res.*, 11, 149–159.

Woods, R. (2003), The relative roles of climate, soil, vegetation and topography in determining seasonal and long-term catchment dynamics. *Advances in Water Resources* **26**(3): 295-309.

Y. Yokoo, M. Sivapalan and T. Oki, Investigating the roles of climate seasonality and landscape characteristics on mean annual and monthly water balances, *Journal of Hydrology* **357** (2008), pp. 255–269.

Zhang, L., W. R. Dawes, and G. R. Walker (2001), Response of mean annual evapotranspiration to vegetation changes at catchment scale, *Water Resour. Res.*, 37(3), 701–708.

Table 5.1: Mean and standard deviation for different types of basins in the study.

Aridity Index	Type	Mean (mm/year)	Standard deviation (mm/year)
$0 \leq R \leq 0.5$	Humid	1200	150
$0.5 \leq R \leq 1$	Semi-Humid	1000	200
$1 \leq R \leq 2$	Temperate	800	100
$R \geq 2$	Arid	500	100

Table 5.2: Moisture and Energy Availability Scenarios considered for the study.

Scenario	Aridity Index	Regions	Seasonality in P	$\Delta\alpha$
MEAS 1	$0 \leq R \leq 1$	Pacific Northwest (Washington, Western Oregon, Northern Idaho)	Strong	π (Out of Phase)
MEAS 2	$1 \leq R \leq 2$	California	Strong	π (Out of Phase)
MEAS 3	$0 \leq R \leq 1$	New England	No seasonality	0 (No Phase)
MEAS 4	$1 \leq R \leq 2$	Mid-Atlantic, South Atlantic, Great Lakes, Ohio, Tennessee, Upper and Lower Mississippi	No seasonality	0 (No Phase)
MEAS 5	$1 \leq R \leq 2$	Florida	Strong	0 (In Phase)

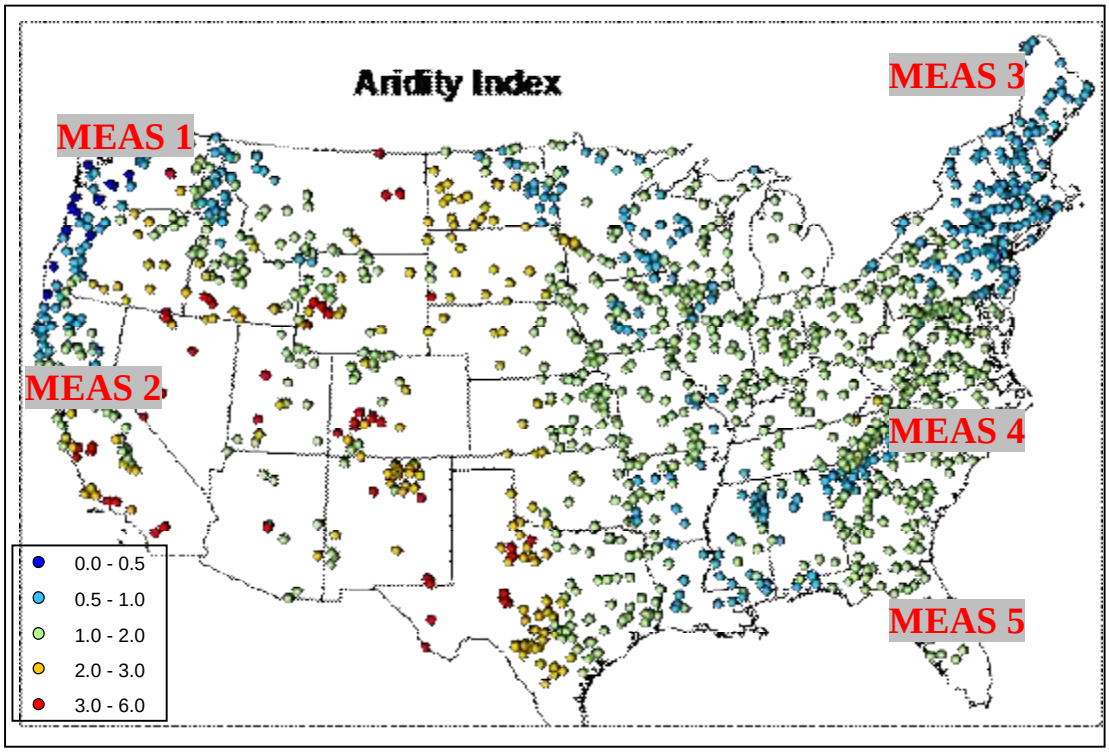

Figure 5.1: Aridity index for all the 1373 HCDN basins.

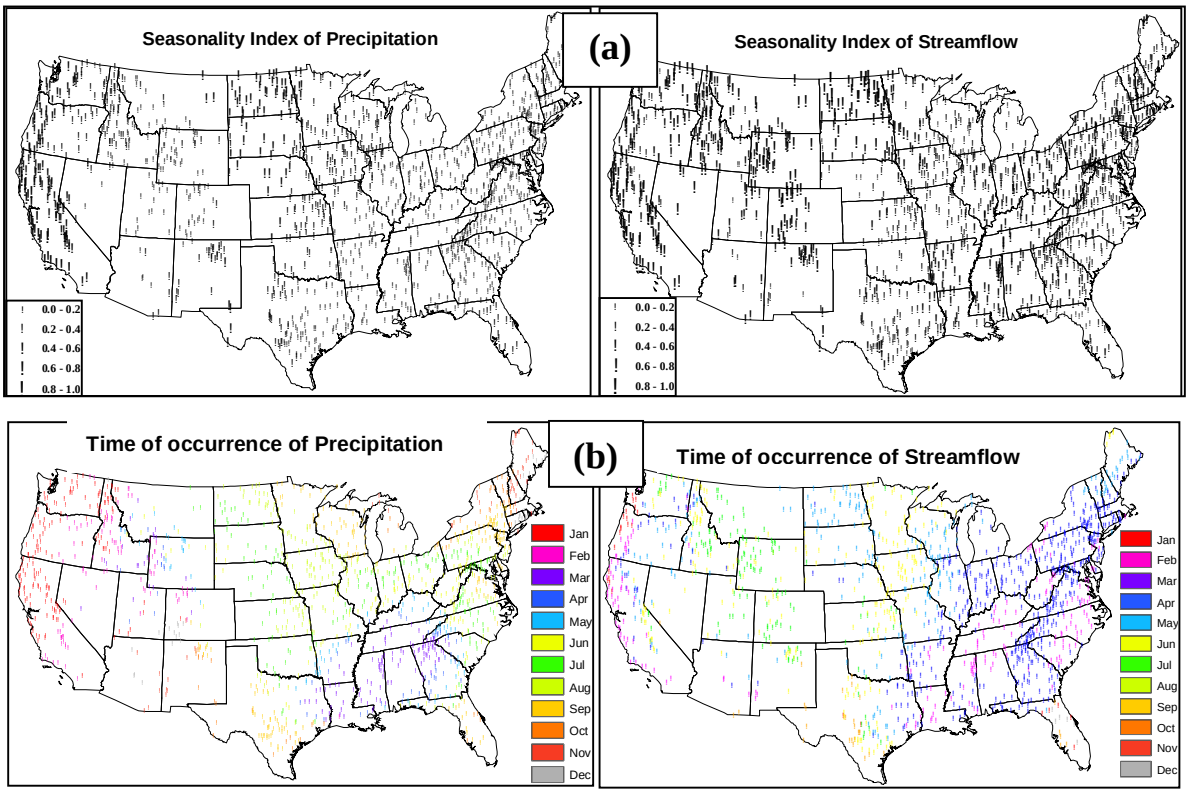

Figure 5.2: Seasonality of observed precipitation and streamflow for the 1373 HCDN basins.

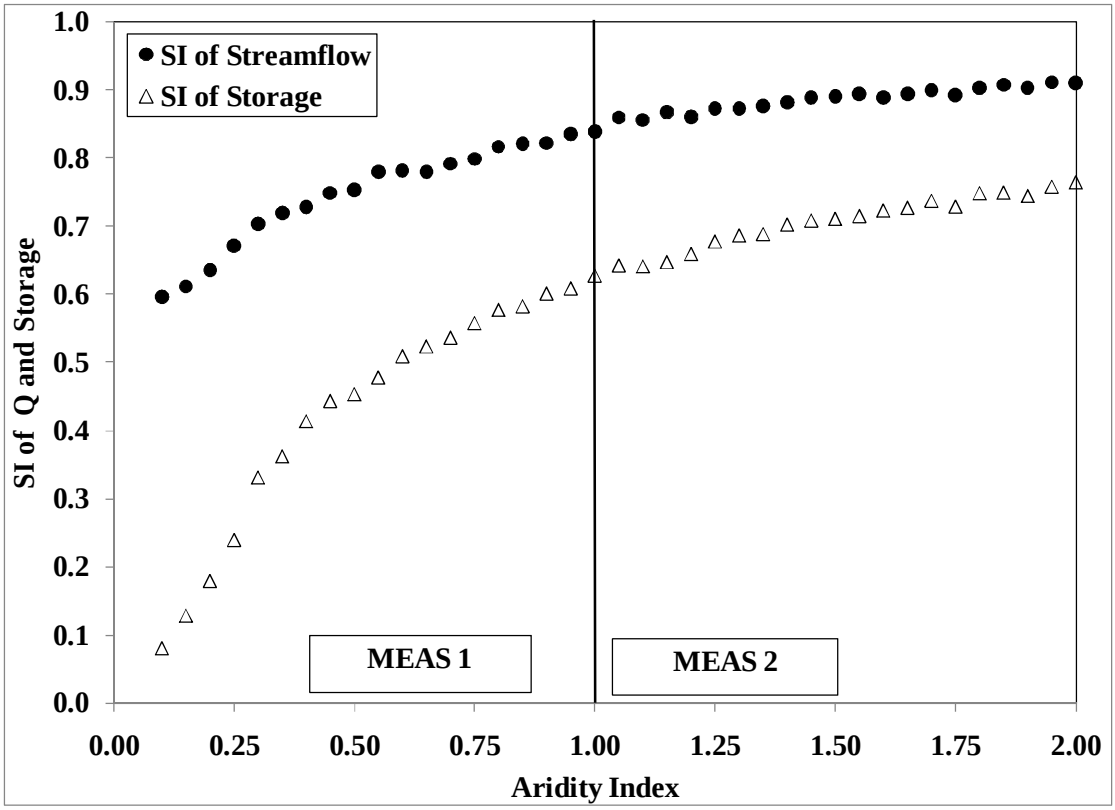

Figure 5.3: SI of estimated streamflow and basin storage for scenarios MEAS 1 and MEAS 2 for different basins. P has a strong seasonality and is out of phase with PET.

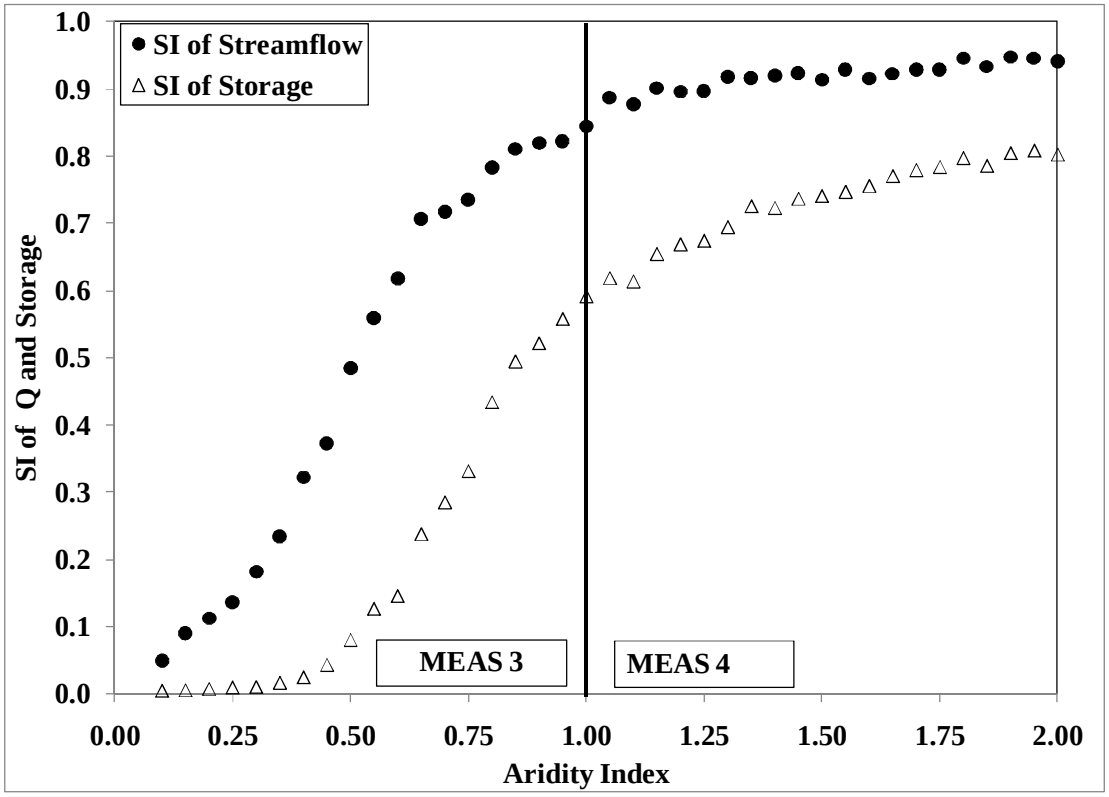

Figure 5.4: SI of estimated streamflow and basin storage for scenarios MEAS 3 and MEAS 4 for different basins.

CHAPTER 6

Summary, Conclusions and Scope for Future Work

6.1 Improvising the Multimodel Combination schemes for Seasonal Climate Forecasting

The research presented in this dissertation is motivated by the fact that the skill of the GCMs primarily depend on the predictor conditions and hence the model uncertainties could be better reduced by combining the GCMs based on their ability to predict under a given predictor state. Recent 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]. On the basis of this, we proposed a methodology to combine multiple GCMs that evaluates the performance of the GCMs contingent on the dominant predictor state. The methodology assigns weights for each GCM by evaluating their skill, quantified by Mean Square Error, over similar predictor conditions. We consider Nino3.4 as the primary predictor influencing the winter precipitation and temperature [Quan et al. 2006] over the United States.

The study presented in Chapter 2 combines seven atmospheric GCMs with climatological ensembles to develop multimodel predictions of precipitation and temperature for the continental United States. The performance of the developed multimodel schemes are compared with individual models based on various verification measures such as rank probability skill score, reliability and resolution scores and brier score. The improvements resulting from multimodel combination over individual models and over the current state of the art multimodel schemes are verified through a rigorous nonparametric hypothesis testing. The study also showed that the proposed multimodel combination methodology improves the reliability and resolution of tercile probabilities resulting with reduced Brier scores. The

improved reliability results with reduced number of false alarms and missed targets in categorical forecasts. Analysis of weights also showed that the proposed methodology assigns higher (lower) weights for GCMs and lesser (higher) weights for climatology during anomalous (neutral) ENSO conditions in grid points.

The study presented in Chapter 3 evaluates the proposed multimodel combination methodology in a forecasting context. This study considered forecasted Nino3.4 as the primary predictor and combined five coupled GCMs with climatological ensembles to develop real-time multimodel precipitation forecasts over the continental United States. Six different multimodel schemes are developed based on an adaptive forecasting scheme and the skill measures show that the developed multimodel forecasts perform better than individual model forecasts and over multimodel forecasts based on pooling and long-term skill. The study clearly showed that the proposed multimodel combination algorithm is beneficial in developing real-time multimodel climate forecasts.

6.2 Role of Multimodel Forecasts in improving Water Management

In Chapter 4, we explored the utility of the multimodel forecasts in developing various drought management strategies for the Falls Lake Reservoir in the Neuse River Basin, NC. In this study, we applied the climate information based multimodel streamflow forecasts for invoking restrictions on the water supply releases and for improving storage conditions at the end of the season for the Falls Lake Reservoir. The performance of summer (JAS) forecasts from three models (parametric regression, semiparametric resampling, and the multimodel forecasts obtained from the two single models) in estimating $\text{Prob}(S_T < S_T^*)$

was evaluated by comparing with the estimates of $\text{Prob}(S_T < S_T^*)$ from climatological ensembles to predict below-normal storage conditions. Invoking restrictions during the summer based on the predicted estimates of $\text{Prob}(S_T < S_T^*)$, lead to increased storage conditions in September. In addition to this, the study also shows that among the three streamflow forecasting models, multimodel streamflow forecasts seem to better predict the change in streamflow potential, thus resulting in reduced false alarms and missed targets in predicting below-normal storage conditions at the end of September. Hence, these analyses show that the combining multimodels contingent on the dominant predictor state is an attractive strategy in improving the skill of multimodel forecasts. Further, applying multimodel forecasts would reduce uncertainty from individual models which could lead to better decisions and also improve public confidence in utilizing seasonal forecasts for water management application.

6.3 Seasonal Hydroclimatology of the Continental U.S.

Moisture and energy availability and their temporal distribution could significantly influence the seasonality of land surface response. Under Chapter 5, in order to understand the role of precipitation and temperature seasonality in influencing the streamflow seasonality, we presented a systematic analysis using simple water balance model that require no calibration and showed that the observed differences in monthly climatology of precipitation and streamflow for the humid and semi arid basins can be attributed to variability in precipitation, evaporation and basin storage.

In conclusion, a new methodology for developing optimal multimodel schemes is presented and demonstrated by developing improved multimodel climate forecasts over the continental United States. The application of the proposed methodology in short-term water management is demonstrated by utilizing the multimodel streamflow forecasts with a reservoir model to invoke restrictions for the Falls Lake Reservoir, NC. The role of seasonality of precipitation and temperature in explaining the climatology of monthly runoff is investigated using a physical model with no calibration.

6.4 Scope for future work

The proposed multimodel schemes are developed by exchanging the climate forecasts in the temporal domain (i.e. by identifying neighbors in the predictor state space). This work could be extended with a Bayesian hierarchical model. Given that Bayesian hierarchical modeling facilitates multi-level modeling, we could extend the proposed multimodel combination scheme to take into account variability in forecasting skill that occur primarily due to variability in location, time and state of the predictor. By looking at the spatial and temporal organized modes exhibited by climate forecasts one can employ a Bayesian hierarchical framework to develop multimodel climate forecasts.

We can also explore the benefits of multimodel climate forecasts in developing streamflow forecasts. Given that the climate forecasts for GCMs are available at large spatial resolutions, one can apply various downscaling techniques like coupling GCMs with Regional Climate Model and using the outputs with large scale watershed model, statistical downscaling or Model Output Statistics etc. However, most of these techniques develop a

statistical relationship between ensemble mean of the climate forecasts and the observed streamflows at the site with the assumption that most of the information carried in the forecasts in represented by the mean. Given that the multimodel forecasts improved the reliability of the forecasts (better correspondence between forecast probability and its observed relative frequency), we could investigate various techniques that can map the conditional distribution of the multimodel climate forecasts to the conditional distribution of the streamflow forecasts. Utilizing the multimodel climate forecasts will lead to better representation of the conditional distribution of the statistically downscaled streamflow forecasts, thereby improving the reliability of meeting the seasonal demands.

6.5 Journal Publications from the Research

- Devineni, N., and A. Sankarasubramanian (2010), Improved Prediction of Winter Precipitation and Temperature over the continental United States: Role of ENSO State in Developing Multimodel Combinations, *Monthly Weather Review.*, in press.

- Devineni, N., A. Sankarasubramanian, and S. Ghosh (2008), Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations, *Water Resour. Res.*, 44, W09404, doi:10.1029/2006WR005855.

- Golembesky, K., A. Sankarasubramanian and N. Devineni (2009), Improved Drought Management of Falls Lake Reservoir: Role of Multimodel Streamflow Forecasts in Setting up Restrictions, *Journal of Water Resources Planning and Management,* 135, 188, DOI:10.1061/(ASCE)0733-9496(2009)135:3(188).

- Sankarasubramanian, A., U. Lall, N. Devineni and S. Espuneva (2009), The Role of Monthly Updated Climate Forecasts in Improving Intraseasonal Water Allocation, *Journal of Applied Meteorology and Climatology*, 48(7), 1464–1482.

- Devineni, N., and A. Sankarasubramanian (2010), Climatology of Monthly Runoff: Causality and Relations to Seasonality in Precipitation and Temperature, *to be submitted to Journal of Hydrology*.

6.6 References

Branković Č., and T. N. Palmer, 2000: Seasonal skill and predictability of ECMWF PROVOST ensembles. *Quart. J. Roy. Meteor. Soc.*, **126**, 2035–2067.

Quan X., M. Hoerling, J. Whitaker, G. Bates, and T. Xu, 2006: Diagnosing sources of US seasonal forecast skill, *J. Climate*, **19,** 3279-3293.

Shukla J., J. Anderson, D. Baumhefner, C. Brankovic, Y. Chang, E. Kalnay, L. Marx, T. Palmer, D. Paolino, J. Ploshay, S. Schubert, D. Straus, M. Suarez, and J. Tribbia, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593-2606.

APPENDICES

Appendix A: The Hotelling – Williams Correlation Hypothesis Test

Two correlations can have dependence if they are computed across the same individuals. Hence, we have to test whether these two correlations are equal. For example, let the correlation between ECHAM4.5 predicted precipitation and observed precipitation be ρ_{12} and the correlation between CCM3v6 predicted precipitation and observed precipitation be ρ_{13}. ρ_{12} can be equal to ρ_{13} implying that the correlations are equal or ρ_{12} can be greater or lesser than ρ_{13} implying that one correlation can be significant over the other correlation. Hence, we can assume the following null hypothesis to test for the equality of two correlations.

H_0: $\rho_{12} = \rho_{13}$ … (A-1)

Hotelling and William suggested that the test statistic has a Student's t-distribution with (*n-3*) degrees of freedom and can be calculated as follows.

$$t_{(n-3)} \sim (\rho_{12}-\rho_{13}) \sqrt{\frac{(n-1)(1+\rho_{23})}{\frac{2(n-1)}{(n-3)}|R|+\bar{r}^2(1-\rho_{23})^3}} \quad \text{… (A-2)}$$

Where, $\bar{r}=\frac{\rho_{12}+\rho_{13}}{2}$ … (A-3)

and $|R| = 1-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2+2\rho_{12}\rho_{13}\rho_{23}$ … (A-4)

We can find the critical value for t with (n-3) degrees of freedom and either accept or reject the null hypothesis. For instance, if the percentile of test statistic is greater than or equal to 0.9 at 10% significance level, the null hypothesis that the correlations are equal can be

rejected. ECHAM4.5 predicted precipitation will be more significantly correlated with observed precipitation than CCM3v6 predicted precipitation.

Appendix B: Rank Probability Score and Rank Probability Skill Score

Given that seasonal forecasts are 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 precipitation/temperature, 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 precipitation/temperature 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 precipitation/temperature forecasts at time't' from m^{th} model with i=1, 2... N ensembles, $Q_{i,t}^{m}$, then the forecast probabilities for the d^{th} category could be expressed as $FP_{d,t}^{m}(Q) = n_{d,t}^{m} / N$ by computing the number of ensembles between $Q_d \le Q_{i,t}^{m} \le Q_{d+1}$.To compute RPSS, the first step is to compute Rank Probability Score (RPS). Given D categories and $FP_{d,t}^{m}(Q)$ for a forecast, we can express the RPS for a particular year't' from m^{th} model as

$$RPS_t^m = \sum_{d=1}^{D} \left[CF_{d,t}^m - CO_d \right]^2 \quad \text{... (B-1)}$$

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 precipitation/temperature falls in the d^{th} category, $CO_q = 0$ for $1 \le q \le d\text{-}1$ 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 $FP_{d,t}^{c\lim}(Q) = 1/D$.

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

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., 2003]. One can use RPSS to produce maps showing the spatial characteristics of the forecast skill [Goddard et al., 2003]. 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 a given 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. 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

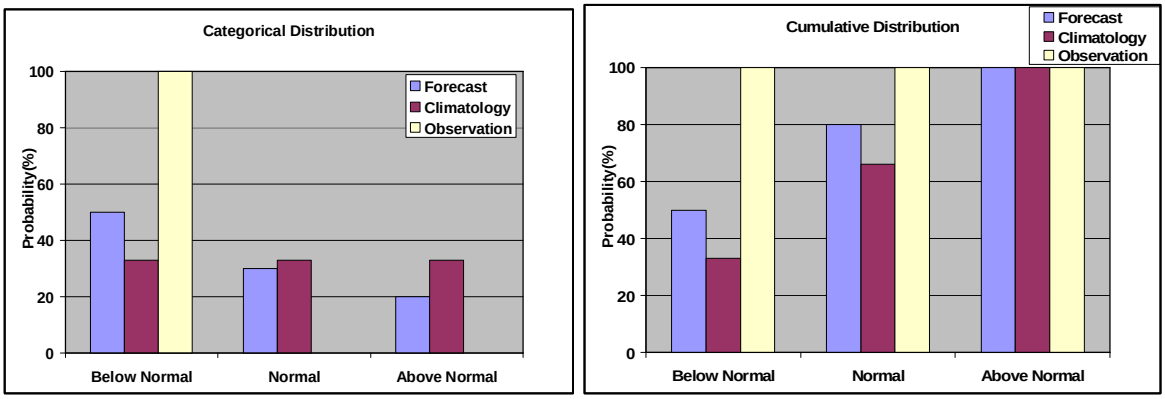

Figure B-1: Observed Category falling in Below Normal

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

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

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

$1-(0.29/0.5578) = 0.48$

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

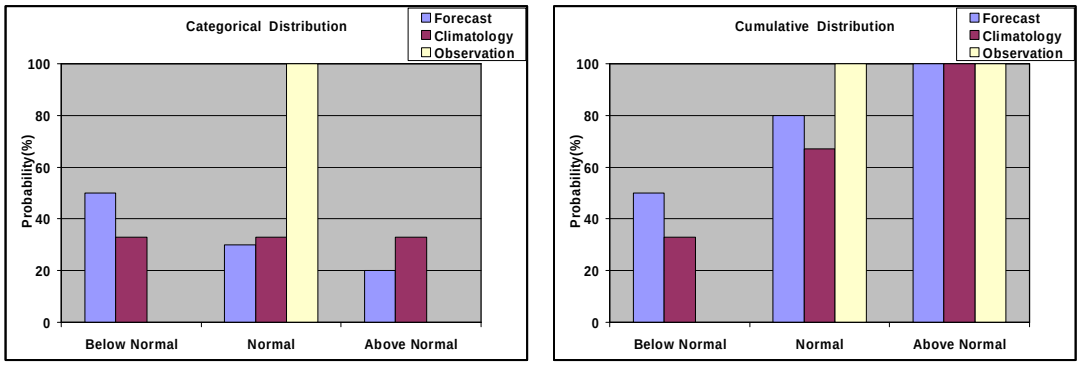

Figure B-2: Observed Category falling in Normal

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

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

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

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

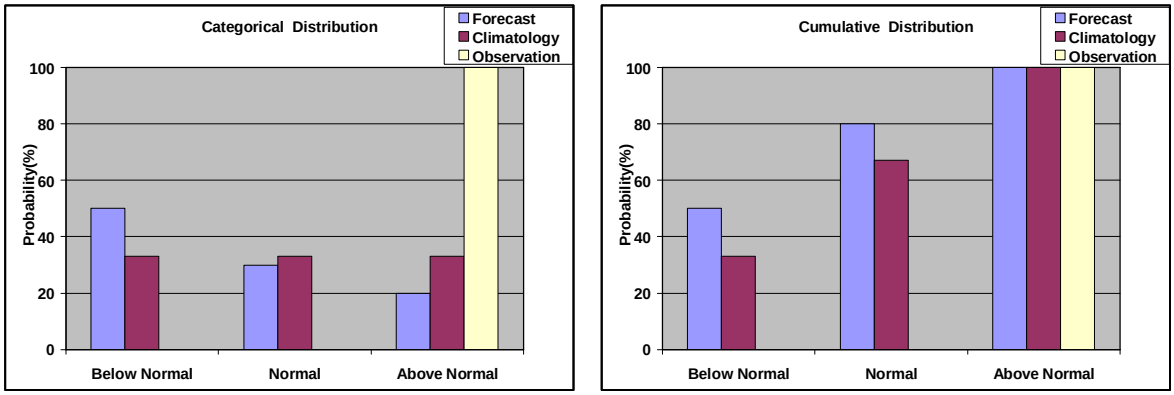

Figure B-3: Observed Category falling in the Above Normal

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.

APPENDIX C: Seasonality of Hydrologic Variables

Seasonality of a hydrologic variable can be measured using circular or directional statistics [Markham 1970]. Circular statistics is useful for quantifying the time of occurrence of an event when time is measured on a circle. Seasonality can be quantified by calculating two components, the average time of occurrence, and the degree to which the event tend to be concentrated in time, called the seasonality index. Under this method, the time through the year is represented on a circle and each month is assigned an angle, measured from 1st January as shown in Table C1. The hydrologic variable of interest is assumed to be a vector quantity with magnitude and direction. The 12 months vectors can be added vectorially to get a resultant vector with a direction ϕ_R and magnitude P_R.

$$\phi'_R = tan^{-1}\left(\frac{S}{C}\right) \quad \ldots(C\text{-}1)$$

$$P_R = \left(S^2 + C^2\right)^{1/2} \quad \ldots(C\text{-}2)$$

$$S = \sum_{m=1}^{12} P_m \sin \phi_m \quad \ldots(C\text{-}3)$$

$$C = \sum_{m=1}^{12} P_m \cos \phi_m \quad \ldots(C\text{-}4)$$

P_m are the 12 monthly magnitudes and ϕ_m are the 12 monthly time angles (Table C1).

The average time of occurrence, ϕ_R, is given by the value of ϕ'_R and the signs of S and C. This decides the quadrant of the circle the resultant vector falls under.

$$\phi_R = \phi'_R \; if \; S > 0 \; and \; C > 0; \quad \ldots(C\text{-}5)$$

$$\phi_R = \phi'_R + 180^0 \; if \; C < 0; \quad \ldots (C\text{-}6)$$

$$\phi_R = \phi'_R + 360^0 if \; S < 0 \; and \; C > 0; \quad \ldots (C\text{-}7)$$

ϕ_R can be converted to the month of the year in which it occurs using Table C1. The seasonality index is the ratio of the resultant vector P_R to the total magnitude and the ratio ranges from 0 to 1. For example, the seasonality index of precipitation is 0 if the magnitude of precipitation is equal in all the months, and is equal to 1 if all the precipitation occurs in a single month.

Table C1: Angles for computing seasonality index and average time of occurrence for monthly data.

Mid-Month					First of Month	
Month	Day of Year	Angle (degrees)	Sine	Cosine	Day of year	Angle(degrees)
Jan	16	15.8	0.272	0.962	1	1.0
Feb	45.5	44.9	0.705	0.709	32	31.6
Mar	75	74	0.961	0.276	60	59.2
Apr	105.5	104.1	0.970	-0.243	91	89.8
May	136	134.1	0.718	-0.696	121	119.3
Jun	166.5	164.2	0.272	-0.962	152	149.9
Jul	197	194.3	-0.246	-0.969	182	179.5
Aug	228	224.9	-0.705	-0.709	213	210.1
Sep	258.5	255	-0.966	-0.26	244	240.7
Oct	289	285	-0.966	0.259	274	270.2
Nov	319.5	315.1	-0.706	0.708	305	300.8
Dec	350	345.2	-0.256	0.967	335	330.4