Mitigating Predictive Uncertainty in Hydroclimatic Forecasts: Impact of uncertain inputs and model structural form

Shahadat Hossain Chowdhury

A thesis submitted for the degree of Doctor of Philosophy at the University of New South Wales, Sydney, Australia

April 2009

(blank page)

SUMMARY

Hydrologic and climate models predict variables through a simplification of the underlying complex natural processes. Model development involves minimising predictive uncertainty. Predictive uncertainty arises from three broad sources which are measurement error in observed responses, uncertainty of input variables and model structural error. This thesis introduces ways to improve predictive accuracy of hydroclimatic models by considering input and structural uncertainties. The specific methods developed to reduce the uncertainty because of erroneous inputs and model structural errors are outlined below.

The uncertainty in hydrological model inputs, if ignored, introduces systematic biases in the parameters estimated. This thesis presents a method, known as simulation extrapolation (SIMEX), to ascertain the extent of parameter bias. SIMEX starts by generating a series of alternate inputs by artificially adding white noise in increasing multiples of the known input error variance. The resulting alternate parameter sets allow formulation of an empirical relationship between their values and the level of noise present. SIMEX is based on the theory that the trend in alternate parameters can be extrapolated back to the notional error free zone. The case study relates to erroneous sea surface temperature anomaly (SSTA) records used as input variables of a linear model to predict the Southern Oscillation Index (SOI). SIMEX achieves a reduction in residual errors from the SOI prediction. Besides, a hydrologic application of SIMEX is demonstrated by a synthetic simulation within a three-parameter conceptual rainfall runoff model.

This thesis next advocates reductions of structural uncertainty of any single model by combining multiple alternative model responses. Current approaches for combining hydroclimatic forecasts are generally limited to using combination weights that remain static over time. This research develops a methodology for combining forecasts from multiple models in a dynamic setting as an improvement of over static weight combination. The model responses are mixed on a pair wise basis using mixing weights that vary in time reflecting the persistence of individual model skills. The concept is referred here as the pair wise dynamic weight combination. Two approaches for forecasting the dynamic weights are developed. The first of the two approaches uses a mixture of two basis distributions which are three category ordered logistic regression model and a generalised linear autoregressive model. The second approach uses a modified nearest neighbour approach to forecast the future weights. These alternatives are used to first combine a univariate response forecast, the NINO3.4 SSTA index. This is followed by a similar combination, but for the entire global gridded SSTA forecast field. Results from these applications show significant improvements being achieved due to the dynamic model combination approach. The last application demonstrating the dynamic combination logic, uses the dynamically combined multivariate SSTA forecast field as the basis of developing multi-site flow forecasts in the Namoi River catchment in eastern Australia. To further reduce structural uncertainty in the flow forecasts, three forecast models are formulated and the dynamic combination approach applied again. The study demonstrates that improved SSTA forecast (due to dynamic combination) in turn improves all three flow forecasts, while the dynamic combination of the three flow forecasts results in further improvements.

CERTIFICATE OF ORIGINALLITY

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project’s design and conception or in style, presentation and linguistic expression is acknowledged.’

S Chowdhury

ACKNOWLEDGEMENT

I would like to first and foremost acknowledge my supervisor, Associate Professor Ashish Sharma, who has always been there to provide valuable advice, guidance and critical feedback on my research. He motivated me to return to undertake this research challenge after an absence of 15 years from the universities. It was one of the best decisions of my career, thank you Ashish! My supervisory team included Dr Ian Cordery and Dr James Ball. This research was funded by the Australian Research Council and the Sydney Catchment Authority.

My gratitude goes to Professor Upmanu Lall of Columbia University who during my four months stay in New York in 2005-2006 provided me with excellent research supervision. Thank you Manu for the opportunity of being a visiting scholar of your university! It was a wonderful learning experience. During that time I was also benefited from discussions with Dr Andrew Robertson of International Research Institute of Climate and Society, New York.

My thanks are also due to administrative staffs of School of Civil and Environmental Engineering, special mention goes to friendly Les Brown and helpful Patricia McLaughlin, who tolerated my numerous requests. I enjoyed representing students in the Faculty Board of the Faculty of Engineering. In addition, I recognise the cooperation of my team mates of the Post Graduate Research Students Association of the School during my tenure as the vice president of the Association.

Last but not the least; I would like to mention my family. My academic journey started by learning my first alphabet from my parents. My grand mother has been excitedly waiting to hear the news of my attaining the highest academic degree. To my lovely daughter and son Fatiha and Furqan: I missed your many fine hours in last four years. My wonderful wife Jintana has been my inspiration and my light in dark hours: thanks for tirelessly managing the home and outside alone. Jintana, this work would be impossible without you!

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

1 INTRODUCTION

1.1 Motivation

Water resource management and planning plays a vital role in shaping human habitat and its surrounding environment. Good hydrological knowledge facilitates superior water resource planning and implementation. Hydrologists rely on models to simulate the natural process that are not readily known, such as flood levels, irrigation demand, storage volumes or the next seasons flow. For example, a seasonal flow forecast model is important for irrigation, hydropower generation, flood mitigation or managing failure risk of drinking water supply. Any model development involves minimising the predictive uncertainty (residual error) in order to improve the output reliability.

This research offers robust tools to improve predictive accuracy of hydroclimatic models. Firstly, it demonstrates a statistical tool to reduce the uncertainty caused due to imprecise input variables. Secondly, it introduces a robust multimodel combination approach to reduce structural uncertainty of any single model. Consequently the understanding of the potential sources of uncertainty, as explained next, is a prerequisite of exploring any methods for reducing this uncertainty.

1.2 Background and Context

It is worthwhile to flag early in this manuscript that no consensus definition of uncertainty exists in scientific literature (Montanari, 2007). This thesis refers to uncertainty as measurements of departure of estimates from respective true values. The terms noise, error (common in statistics) and uncertainty (used in hydrology) are used synonymously here.

There has been considerable interest in analysing uncertainties in environmental models (Beven, 2006; Beven and Freer, 2001; Butts et al., 2004; Montanari, 2007; Montanari and Grossi, 2008). The models vary from hydraulic routing of hourly flood hydrograph to statistical models of seasonal sea surface temperature. Figure 1.1 shows sources of uncertainty introduced at various stages of the modelling process. Total error in hydrological modelling arises from three broad sources (Huard and Mailhot, 2006; Kavetski et al., 2002; Vrugt et al., 2005). These are (a) measurement error in observed responses, (b) uncertain input variables and (c) model structural error.

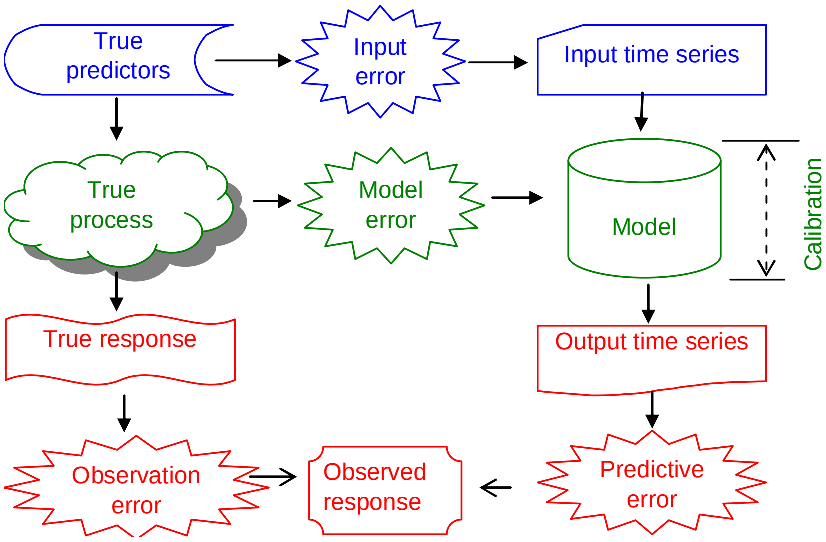


Figure 1.1. Conceptual diagram of categories of error in time series models.

1.2.1 Measurement error in observed responses

The observed response is used to calibrate model parameters. Erroneous observation leads to weak parameter estimates contributing adversely to the total simulation error. The uncertainty in response variables due to unsteady error structure, non stationary systems or measurement anomalies are discussed in a number of hydrological studies (Cordery et al., 2004; Khadam and Kaluarachchi, 2004; Viney and Bates, 2004). Studies on model calibration and design of likelihood function generally consider unseen true population based on an error prone sample of response data. There have been continual research contributions on various aspects of calibration of hydrological models (Huang and Liang, 2006; Madsen, 2000; Madsen, 2003; Sorooshian et al., 1993; Tolson and Shoemaker, 2008; V´azquez et al., 2008; Yapo et al., 1996; Yapo et al., 1998).

In a perfect model with true input data, the ordinary least square method provides robust estimates of parameters against response variables with white measurement errors. Standard statistical text books contain sufficient details on how to deal with various error distributions of the response variable while calibrating the model parameters (Agresti, 1996; Chambers, 1992; Hastie and Tibshirani, 1986; Hastie et al., 2000). Note that calibration is alternatively known as the inverse problem in time series statistics. The research on robust calibration spreads across the entire science discipline and is beyond the scope of this PhD research.

1.2.2 Uncertain input variables

Uncertainty of input variables may range from errors in the observed inputs to a coarse density of observation stations. The associated errors in the input can be a result of instrumentation, interpolation or extrapolation of data in space and time or conversion of point measurements into areal values. For example inadequate representation of catchment rainfall in a rainfall runoff model or designing predictors based on insufficient climate indices in flow forecast models. The uncertainty in hydrological model inputs, if ignored, introduces systematic bias in the parameters estimated. There have been many studies that address the effect of input uncertainty of a hydrological model (Andreassian et al., 2001; Huard and Mailhot, 2006; Kavetski et al., 2002; Kavetski et al., 2006a; Kavetski et al., 2006b; Troutman, 1982). However, Huard and Mailhot (2006) conceded the limited application of various frameworks (including their own proposed Bayesian method) due to high level of non linearity; they depicted the reduction of input error effect of widely used models as an open research problem. The reduction of the bias caused by input error forms a minor part of this research.

1.2.3 Inadequate model structure

Hydrologic models are simplifications of complex natural phenomena. These simplifications lead to structural uncertainty. Competing hydrological models have been developed with an aim to minimise model uncertainty for various applications (Boughton, 2005). Different model structures are being studied to reduce uncertainty including the changes in uncertainty as a function of temporal and spatial scaling, model complexity versus accuracy, model robustness to different climate zones etc. (Huang and Liang, 2006; Liden and Harlin, 2000; Wagener and Wheater, 2005). The model structural error is specific to the model of choice and the scope of improvement is often limited.

The differential improvement in prediction, achieved by adjusting a single model structure in isolation, diminishes asymptotically. For example, the progression of an autoregressive (AR) model into an autoregressive integrated moving average (ARIMA) results from the quest of shrinking structural error. However the progression is still constrained by the basic premise of a regression model conforming to a given distribution. Another example from a conceptual modelling point of view may be the evolution from a simple 3 parameter ABC rainfall runoff model (Feiring, 1967; Huard and Mailhot, 2006) to a 16 parameter Sacramento model (Burnash, 1975; Burnash, 1973). In this example, the same basic premise post progression remains that the state variables are not readily observable. Hence it is reasonable that the expedition for structural perfection should look beyond a single modelling platform. The concept of model combination arises to exploit strength of different modelling platform. The concept is as old as time series statistics, as alluded to in the following quote in Clemen (1989):

> "*In combining the results of these two methods, one can obtain a result whose probability law of error will be more rapidly decreasing*" (Laplace, 1818)

Clemen (1989) drew conclusions based on two thousand relevant journal pages and 11 books that combining forecasts leads to increased forecast accuracy. He declared that "we no longer need to justify this methodology". The forecasting community agreed that use of ensemble forecasts is common practice now (Armstrong, 2001; Hoeting, 1999; Kim et al., 2006; Menezes et al., 2000; Zou and Yang, 2004). In the field of hydrology, model combination is sometime known as ensemble forecasting (Armstrong, 2001, pp 418), hence the term ensemble appears synonymously with the term combination in this thesis. Note that the earliest ensemble prediction in meteorology may be attributed to Sanders (1963).

The general theory of uncertainty and its propagation can be found in various statistical text books (Barry, 1978; Dieck, 2002; Jaech, 1985; Rabinovich, 1993). The method relevant to this thesis is the weighted combination of two results where the weights reflect the respective uncertainty. The weights are proportional to the precision (inverse of error variance) of the respective results pp 144-148 (Dieck, 2002). Besides, Dieck (2002) documented prime equations for uncertainty analysis and analytical equations of uncertainty (error) propagation. These theories are based on the assumption of stationary error which is not always true for time series variable. The thesis extends the above concept into non stationary error variance. Accordingly, it develops a time varying combination method as an improvement over the existing static combination methods applicable to hydroclimatic forecasts.

1.3 Scope of the Research

This research focuses on hydrological and climate modelling using the broader discipline of time series forecasting and applied statistics to develop model combination approaches. It explores ways to improve predictive accuracy of hydroclimatic models after the scope of improvement of an individual model has been exhausted. The reduction in uncertainty is achieved by a two pronged statistical approach. Firstly it proposes shrinking the parameter bias of individual models (Chapter 2 and Chapter 3) by modelling the input error distribution. Secondly, it advocates combining a number of bias corrected individual models (Chapter 4 to Chapter 6) as a way to reduce overall uncertainty of any single modelling structure. The second topic of combining models forms the bulk of the research contribution. The two topics are briefly introduced below while the relevant details of any particular theory are included at later chapters.

1.3.1 Mitigation of input error

The first part of this thesis introduces a method termed ‘Simulation Extrapolation’ or SIMEX (Cook and Stefanski, 1994) for use in parameter estimation where input error is significant. SIMEX mitigates parameter bias caused by inaccurate input variable values. This thesis presents the first application of SIMEX to hydrology and climate modelling. The thesis first assesses the rationale behind SIMEX in a linear regression (climate model) setting in Chapter 2 and then in a non linear conceptual case study (rainfall runoff model) in Chapter 3.

1.3.2 Mitigation of model error

This research advocates model combination as a way to reduce model structural error (Butts et al., 2004). Static weight combination is the current practice to combine multiple hydroclimatic models as a means of reducing structural error (Barnston et al., 2003; Colman and Davey, 2003; Greene et al., 2005; Peng et al., 2002; Raftery et al., 2005; Robertson et al., 2004; Sharma and Lall, 2004). Static combination ignores any non stationarity of individual model skill (or, increased accuracy associated with individual models for defined conditions or periods of time), a likely attribute of any hydroclimatic model. The consideration of persistence of the component model skills while combining forecasts is introduced here. The method is named the Pair wise Dynamic Weight (PDW) combination method. The term dynamic is used here to denote the fact that the mode of combination varies with time.

The usefulness of PDW method is demonstrated using cases with increasing complexity. The first case study involves combining univariate sea surface temperature anomaly (SSTA) forecasts (Chapter 4) which is extended to the case of a multivariate SSTA forecast next (Chapter 5). Finally PDW is applied to a multivariate hydrological forecast where the underlying data exhibits significant asymmetry in its probability distribution attributes (Chapter 6).

1.4 Outline of the Thesis

This thesis is presented as a collation of papers. The chapters of this manuscript are either published or submitted for publication in scientific journals. Each chapter can be read as a stand alone document. The notations are chapter specific. Readers may note certain amount of duplication of information that is presented among the chapters. No dedicated literature review chapter exists. The literature reviews specific to each topic are included at the beginning of the relevant chapters. The overall conclusion and a summary of original contributions are included in Chapter 7. A comprehensive and updated bibliography is attached at the end for ease of the readership.

The two main contributions of the research presented here related to input and structural errors (Sections 1.3.1 and 1.3.2) are presented in five hydroclimatic modelling scenarios as follows:

Chapter 2. Application of SIMEX to reduce the effect of input error on a linear model. The case study presented here relates to erroneous sea surface temperatures used as input variables of a linear model to predict the Southern Oscillation Index (SOI).

Chapter 3. Application of SIMEX in a typical non linear hydrologic problem where the associated rainfall inputs are uncertain. The application studies the impact on model parameter values when the input uncertainty is taken into consideration.

Chapter 4. Application of the forecast combination procedure based on PDW to a univariate response case, which in this case is the forecast of the NINO3.4 sea surface temperature index.

Chapter 5. Application of forecast combination based on PDW to a multivariate response case, a globally grided sea surface temperature anomaly medium term forecast.

Chapter 6. Using the forecast combination logic based on PDW for forecasting highly variable and skewed multi-site river flows using predicted sea surface temperature anomaly fields.

Chapter 7. Conclusion

1.5 Summary

This thesis developed an overall structure for mitigating predictive uncertainty. As has been mentioned above, each chapter can be read as a stand alone document. The case studies presented are selected to best illustrate the logic developed in each chapter. Consequently the applications do not use the same case study from the beginning to the end. The rest of this section summarises the overall structure of the thesis, as illustrated in Figure 1.2.

This introductory chapter motivates the research as a means to aid water resources planning by improving accuracy of the hydroclimatic model simulations. The improvements are sought by reducing total predictive uncertainty. The term ‘uncertainty’ is explained along with the conceptualisation of uncertainties under three broad sources. These are measurement error of responses, imprecise input variables and uncertain model structure. The thesis limits its scope into mitigating impacts of the later two sources of uncertainty.

The second and third chapters introduce a statistical tool (SIMEX) to mitigate the effect (parameter bias) of imprecise input variables in a model. This research encourages application of SIMEX to improve prediction using hydroclimatic models.

The manuscript, from the fourth chapter onwards, focuses on model structural uncertainty, without taking into consideration the input uncertainty addressed using SIMEX in chapters 2 and 3. The impact of structural uncertainty is reduced by combining multiple model responses. Chapter four underpins the primary theory of dynamic combination of univariate responses using a mixture regression method.

The dynamic weight method is further developed in chapter five. The procedure is extended to multivariate responses. Accordingly, the mixture regression models developed are now applicable for more generic modelling scenarios where interest is not on a single response, but a collection of outcomes. Besides, a nearest neighbour based non parametric method of constructing multivariate dynamic weight is presented.

Chapter six aims at assessing the impact of the model combination algorithms presented in chapters 4 and 5 to a practical hydrological application. This chapter allows a critical evaluation of effectiveness of the proposed dynamic combination in reducing forecast uncertainty. The discussion of the chapter six provides helpful insight into strength and weakness of the method in improving hydroclimatic predictions.

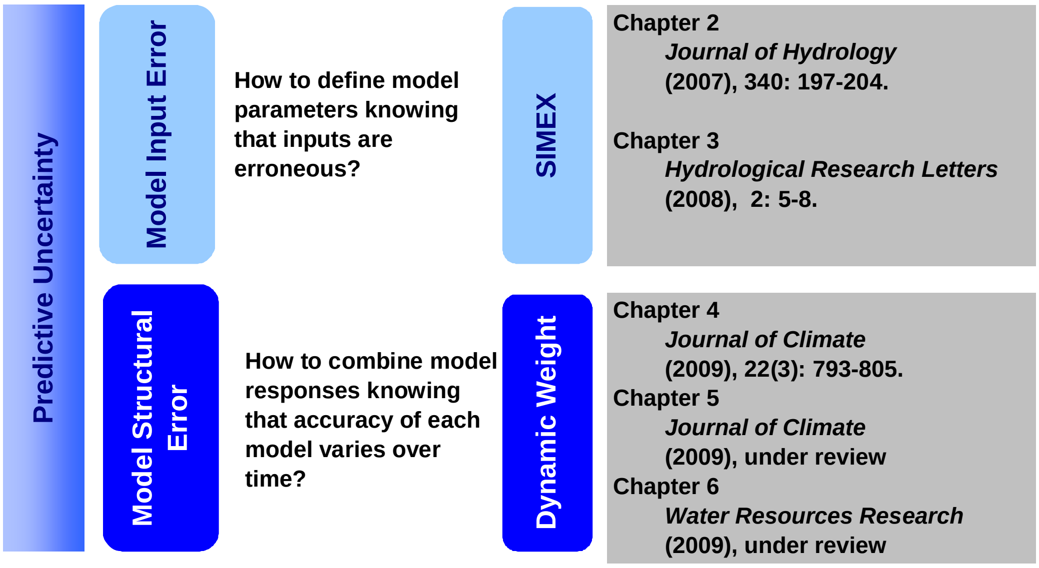


Figure 1.2. The overall narrative of the thesis.

1.6 Reference

Agresti, A., 1996. An Introduction to Categorical Data Analysis. Wiley Series in Probability and Statistics, New York.

Andreassian, V., Perrin, C., Michel, C., Usart-Sanchez, I. and Lavabre, J., 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. Journal of Hydrology, 250(1– 4): 206– 223.

Armstrong, J.S., 2001. Combining forecasts. In: J.S. Armstrong (Editor), Principles of Forecasting: A Handbook for Researchers and Practitioners. Kluwer Academic Publishers, Norwell, pp. 417-439.

Barnston, A.G., Mason, S.J., Goddard, L., Dewitt, D.G. and Zebiak, S.E., 2003. Multimodel Ensembling in Seasonal Climate Forecasting at IRI. Bulletin of the American Meteorological Society, 84(12): 1783-1796.

Barry, B.A., 1978. Errors in practical measurement in science, engineering, and technology. Wiley, New York.

Beven, K., 2006. On undermining the science? Hydrological Processes, HPToday, 20: 3141-3146.

Beven, K. and Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1-4): 11-29.

Boughton, W., 2005. Catchment water balance modelling in Australia 1960–2004. Agricultural Water Management, 77: 91-116.

Burnash, R.J., 1975. The NWS river forecast system–catchment modelling. In: V.P. Singh (Editor), Computer Models of Watershed Hydrology. Water Resources Publications, Littleton, Colo.

Burnash, R.J., Ferral, R.L., McGuire, R.A., 1973. A Generalized Streamflow Simulation System–Conceptual Modelling for Digital Computers., U.S. Department of Commerce, National Weather Service, and State of California, Department of Water Resources, Silver Springs, MD/Sacramento, California.

Butts, M.B., Payne, J.T., Kristensen, M. and Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. Journal of Hydrology, 298(1-4): 242-266.

Chambers, J.M., 1992. Linear Models. In: J.M. Chambers and T. Hastie (Editors), Statistical Models in S. Wadsworth & Brooks, Cole, pp. 95-144.

Clemen, R.T., 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5(4): 559-583.

Colman, A.W. and Davey, M.K., 2003. Statistical prediction of global sea-surface temperature anomalies. International Journal of Climatology, 23(14): 1677-1697.

Cook, J.R. and Stefanski, L.A., 1994. Simulation-Extrapolation Estimation in Parametric Measurement Error Models. Journal of the American Statistical Association, 89(428): 1314-1328.

Cordery, I., Mehrotra, R., Sharma, A. and Nazemosadat, M.J., 2004. Hydrological models in a non stationary environment. In: B. Webb (Editor), Hydrology, Science and practice for the 21st century. British Hydrological Society, Imperial College, London, pp. 103-107.

Dieck, R.H., 2002. Measurement Uncertainty: Methods and Application. Instrumentation, Systems and Automation Society (ISA), United States.

Feiring, M.B., 1967. Streamflow Synthesis. Harvard University Press, Cambridge, Mass.

Greene, A.M., Goddard, L. and Lall, U., 2005. Probabilistic multimodel climate change scenarios for the evolution of regional temperatures, The International Research Institute for Climate and Society, The Earth Institute at Columbia University, New York.

Hastie, T. and Tibshirani, R., 1986. Generalized Additive Models. Statistical Science, 1(3): 297-310.

Hastie, T., Tibshirani, R. and Friedman, J., 2000. The Elements of Statistical Learning, Data Mining, Inference and Prediction. Springer Series in Statistics.

Hoeting, J.A., Madigan, D., Raftery, A. E. and Volinsky, C. T., 1999. Bayesian Model Averaging: A Tutorial (with discussion). Statistical Science, 14(4): 382-417.

Huang, M. and Liang, X., 2006. On the assessment of the impact of reducing parameters and identification of parameter uncertainties for a hydrologic model with applications to ungauged basins. Journal of Hydrology, 320(1-2): 37-61.

Huard, D. and Mailhot, A., 2006. A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc". Water Resources Research, 42: WR004661.

Jaech, J.L., 1985. Statistical analysis of measurement errors. Wiley, New York.

Kavetski, D., Franks, S. and Kuczera, G., 2002. Confronting Input Uncertainty in Environmental Modelling. In: H.V. Gupta, Sorooshian, S., Rousseau, A. N.

and R. Turcotte (Editor), Calibration of Watershed Models. AGU Water Science and Applications Series, Duan, pp. 49-68.

Kavetski, D., Kuczera, G. and Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research, 42(W03407).

Kavetski, D., Kuczera, G. and Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resources Research, 42(W03408).

Khadam, I.M. and Kaluarachchi, J.J., 2004. Use of soft information to describe the relative uncertainty of calibration data in hydrologic models. Water Resources Research, 40: W11505.

Kim, Y.O., Jeong, D. and Ko, I.H., 2006. Combining rainfall-runoff model outputs for improving ensemble streamflow prediction. Journal of Hydrologic Engineering, 11(6): 578-588.

Laplace, P.S.d., 1818. Deuxieme Supplement a la Theorie Analytique des Probabilities (Courcier Paris). reprinted (1847) in Oeuvres Completes de Laplace, 7(Paris, Gauthier-Villars): 531-580.

Liden, R. and Harlin, J., 2000. Analysis of conceptual rainfall-runoff modelling performance in different climates. Journal of Hydrology, 238(3-4): 231-247.

Madsen, H., 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. Journal of Hydrology, 235(3-4): 276-288.

Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in Water Resources, 26(2): 205-216.

Menezes, L.M.d., Bunn, D.W. and Taylor, J.W., 2000. Review of guidelines for the use of combined forecasts. European Journal of Operational Research, 120: 190-204.

Montanari, A., 2007. What do we mean by ‘uncertainty’? The need for a consistent wording about uncertainty assessment in hydrology. Hydrological Processes, HPToday, 21: 841-845.

Montanari, A. and Grossi, G., 2008. Estimating the uncertainty of hydrologic forecasts: a statistical approach. Water Resources Research, In press.

Peng, P., Kumar, A. and van den Dool, H., 2002. An analysis of multimodel ensemble predictions for seasonal climate anomalies. Journal of Geophysical Research, 107(D23): 4710.

Rabinovich, S.G., 1993. Measurement errors : theory and practice. American Institute of Physics.

Raftery, A.E., Geneiting, T., Balabdaoui, F. and Polakowski, M., 2005. Using Bayesian Model Averaging to Calibrate Forecast Ensembles. Monthly Weather Review, 133: 1155-1174.

Robertson, A.W., Lall, U., Zebiak, S.E. and Goddard, L., 2004. Improved Combination of Multiple Atmospheric GCM Ensembles for Seasonal Prediction. Monthly Weather Review, 132(12): 2732-2744.

Sanders, F., 1963. On subjective probability forecasting. Journal of Applied Meteorology, 2: 191-201.

Sharma, A. and Lall, U., 2004. Model averaging and its use in probabilistic forecasting of hydrologic variables. In: B. Webb (Editor), Hydrology, Science and practice for the 21st century. British Hydrological Society, Imperial College, London, pp. 372-378.

Sorooshian, S., Duan, Q. and Gupta, V.K., 1993. Calibration of rainfall runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. Water Resources Research, 29(4): 1185-1194.

Tolson, B.A. and Shoemaker, C.A., 2008. Efficient prediction uncertainty approximation in the calibration of environmental simulation models. Water Resources Research, 44: W04411.

Troutman, B.M., 1982. An analysis of input errors in precipitation-runoff models using regression with errors in independent variables. Water Resources Research, 18(4): 947–964.

V´azquez, R.F., Willems, P. and Feyen, J., 2008. Improving the predictions of a MIKE SHE catchment-scale application by using a multi-criteriapproach. Hydrological Processes, 22: 2159-2179.

Viney, N.R. and Bates, B.C., 2004. It Never Rains on Sunday: The prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. International Journal of Climatology, 24: 1171-1192.

Vrugt, J.A., Diks, C.G.H., Gupta, H.V., Bouten, W. and Verstraten, J.M., 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research, 41(1): W01017.

Wagener, T. and Wheater, H.S., 2005. Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. Journal of Hydrology, In Press, Corrected Proof.

Yapo, P.O., Gupta, H.V. and Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology, 181(1-4): 23-48.

Yapo, P.O., Gupta, H.V. and Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. Journal of Hydrology, 204(1-4): 83-97.

Zou, H. and Yang, Y., 2004. Combining time series models for forecasting. International Journal of Forecasting, 20(1): 69-84.

CHAPTER TWO

This chapter reprinted the following article:

Chowdhury, S. and Sharma, A. (2007), Mitigating Parameter Bias in Hydrological Modelling due to Uncertainty in Covariates. *Journal of Hydrology*, 340: 197-204. DOI:10.1016/j.jhydrol.2007.04.010.

Reprinted with permission from Elsevier Ltd, Oxford, UK.

Reprint license no: 2007930313675. dated: 14 August 2008

Type of reprint: Thesis.

Media: Print and electronic.

Reprinted with altered style and minor corrections.

2 MITIGATING PARAMETER BIAS DUE TO UNCERTAINTY IN COVARIATES

ABSTRACT

The uncertainty in hydrological model covariates, if ignored, introduces systematic bias in the parameters estimated. We introduce here a method to determine the true value of parameters given uncertainty in model inputs. This method, known as SIMulation EXtrapolation (SIMEX) operates on the basis of an empirical relationship between parameters and the level of input noise (or uncertainty). The method starts by generating a series of alternate model inputs by artificially adding white noise in increasing multiples of the known error variance. The resulting parameter sets allow us to formulate an empirical relationship between their values and the level of noise present. SIMEX is based on theory that the trend in alternate parameters can be extrapolated back to the notional error free zone.

We illustrate the strength of SIMEX in improving skills of predictive models that use uncertain sea surface temperature anomaly (SSTA) data over the NINO3 region as predictor to the Southern Oscillation Index (SOI), an alternate measure of the strength of the El Nino Southern Oscillation. Our hypothesis is that the higher magnitude of noise in the pre-1960 data period introduces bias to model parameters where SSTA is the input variable. The relatively error invariant Southern Oscillation Index (SOI) is regressed over SSTA and calibrated using a subset of the series from 1900 to 1960. We validate the resulting models using the less erroneous 1960 to 2003 data period. Overall the application of SIMEX is found to reduce the residual predictive errors during the validation period.

2.1 Introduction

Hydrological modelling involves estimating model parameters that describe the relationship between one or more response variables and associated covariates. In many instances the covariates are either derived or erroneous representations of the variables they seek to represent. The associated errors in the covariates can be a result of instrumentation, interpolation or extrapolation of data in space and time or conversion of point measurement into areal values. In some instances it becomes possible to formulate a process whereby their nature and distributional representation can be specified, and its impact on the assessment of model and parameter uncertainty established. This paper discusses a mechanism for assessing the impact of input errors on hydrological model specification, presenting a methodology whereby parameter estimation procedures can be cognizant of the nature of uncertainty associated with erroneous model inputs.

Total residual error in hydrological modelling and parameter uncertainty arise from three broad sources. These are: (a) model structural error, (b) error in observed responses, and (c) input error or error in observed model covariates. Uncertainty in the model structure is an important source of error in the modelled responses (Butts et al., 2004), requiring development of alternatives whereby outputs from multiple models are pooled together so as to generate an ensemble of hydrographs that are able to represent the uncertainty present (Marshall et al, 2007). Errors in responses have a significant effect on the specification of the model, especially so if the methods used to assess model parameterisation exhibit sensitivity to outliers (Yapo et al., 1996). Errors in input data introduce systematic bias in model parameterisation during calibration. Kavetski et al. (2002) demonstrate the extent of this bias in parameter specifications of a distributed hydrologic model by artificially corrupting the rainfall. A conventional least square calibration disregards the presence of input error in the data and has been shown to result in parameters and associated uncertainty estimates that result in significant bias in the resulting model outputs.

Consider a situation where one is in a position to specify the probability distribution governing the additive errors associated with the model inputs. What is the extent of bias that is introduced as a result of these errors in a hydrological modelling study? What is the effect likely to be when the magnitude or variance of the additive errors is increased in a controlled manner? Is it possible to develop a relationship between the error distribution and the resulting parameter estimates? If so, is it possible to use this relationship as a basis for estimating parameters that would apply were the input variables free of additive errors? This paper seeks to answer the above questions in the context of hydroclimatic modelling applications, introducing a method termed ‘Simulation Extrapolation’ or SIMEX (Cook and Stefanski, 1994) for use in parameter estimation where input data is not error free.

The paper is organised as follows. The next section outlines the logic behind the SIMEX approach. In Section 2.2, the method is applied to estimate parameters of a conceptual rainfall runoff model in a synthetic setting designed to illustrate its applicability and limitations. The later sections present an application of the approach to ascertain the parameters of a simple regression model designed to estimate seasonal values of the Southern Oscillation Index based on reconstructed sea surface temperature data with known error characteristics. The paper concludes by summarising the main issues raised and suggestions for further work.

2.2 Simulation Extrapolation (SIMEX)

Measurement error and its implication in linear and non-linear models have been studied at length in the field of applied statistics (Brown and Mariano, 1993; Carroll, 1995; Fuller, 1987; Liang and Liu, 1991; Linder et al., 1993). Simulation Extrapolation (SIMEX) is one aspect of this area that proposes an algorithm for estimation of optimal model parameters given prior knowledge about the additive error distribution linked to the model covariates. SIMEX was introduced by Cook and Stefanski (1994) and Stefanski and Cook (1995) with the aim of providing an intuitive approach for assessing the ‘true’ parameter set in situations where covariate data is corrupt. SIMEX expects the error in input data to be of an additive form, often requiring modification of the basic model structure (such as a log-transform in case of multiplicative errors) to enforce the additive assumption. While SIMEX has been popular in the applied statistical area since its inception (Berry et al., 2002; Holcomb, 1999; Kim et al., 2000; Kim and Gleser, 2000; Marcus and Elias, 1998; Stefanski, 2000); this study is the first application of the method in hydrology or hydrometeorology to our knowledge

SIMEX may be applied in a linear model, an exponential model, a logistic regression (Carroll et al., 1995) and in certain non linear settings (Staudenmayer and Ruppert, 2004). This method has also been extended to non-parametric models in recent years (Carroll et al., 1999; Lin and Carroll, 2000; Staudenmayer and Ruppert, 2004). We illustrate the rationale behind SIMEX using a simple linear regression model. Consider the following linear regression model:

$$Y = \beta_x X + \varepsilon_x \qquad [1]$$

where, X represents a zero mean independent variable or covariate, Y represents a zero-mean response, β_x is the slope parameter and ε_x is the associated error. The error term ε_x has a zero mean and reflects the uncertainty associated with the response Y and the prescribed model structure.

Consider a case where instead of observing the covariate X, the observed variable is inclusive of an additive error term, the erroneous covariate (denoted W) being expressed as:

$$W = X + U \qquad [2]$$

where U is an independent and Normally distributed additive error having a zero mean and variance of σ_u^2, denoted as $U \equiv N(0, \sigma_u^2)$. Hence Equation [1] changes to the following:

$$Y = \beta_w W + \varepsilon_w \qquad [3]$$

The error in W introduces a bias in parameter estimate β_w with respect to the coefficient β_x that defines the true model. Consequently, the residual error (ε_w)

not only includes ε_x but also an error introduced because of the bias in β_w. In statistical literature β_w is defined as naïve estimate and β_x as the true estimate of the regression model.

A numerical example of Equation [1] and [2] is illustrated in Figure 2.1 where {X= x_1, x_2, x_3… x_{400}} are drawn from uniform random numbers bounded by [-1, +1]. The response {Y= y_1, y_2, y_3… y_{400}} is synthetically generated as, Y= 0.8* X + N (0, 0.1^2)

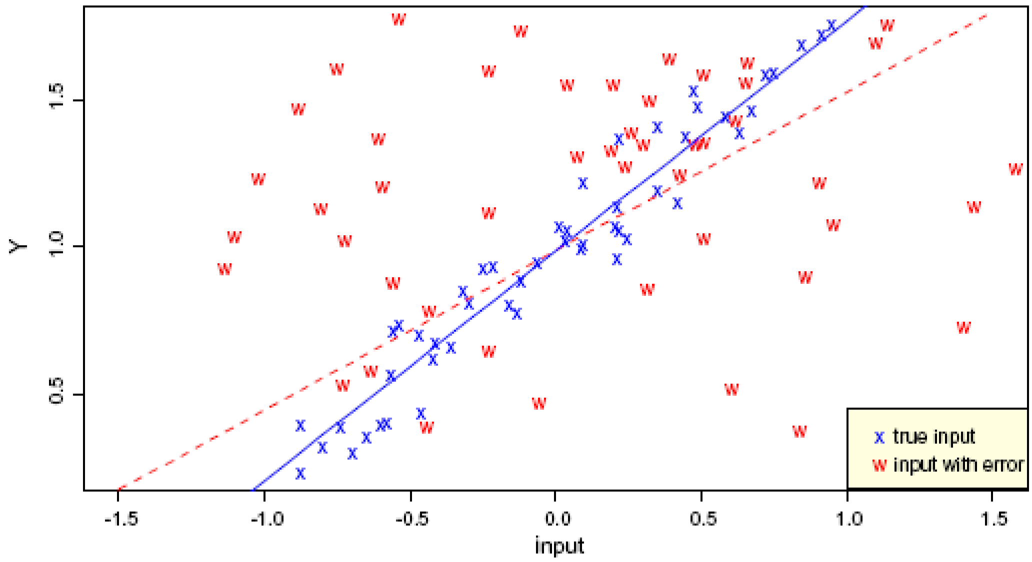


Figure 2.1 The response Y plotted against true covariate X and the covariate measured with error W. The solid blue line is Y~X and the broken red line is Y~W model. Note the reduction in parameter value (slope) due to error in the covariate.

Here the true parameter value β_x = 0.8. The notional observed covariate with additive error {W= w_1, w_2, w_3… w_{400}}, is produced by adding $N(0, \sigma_u^2)$ to X where , σ_u^2 = $\{0.30 \text{ to } 0\}_1^{400}$ decreasing linearly as a function of the sequence length. In this paper the capital italics, eg. N(.) denote functions, the curly brackets and straight uppercase fonts indicate a set of discrete vectors if not mentioned otherwise. Only 50 points out of 400 are drawn for clarity in Figure 2.1. Note how the naïve estimate due to error in observation yields β_w= 0.55. The error in covariate biases the parameter towards zero. The SIMEX method attempts to remove the bias from β_w through the use of a simulation based procedure described below.

The parametric SIMEX algorithm starts with a given model structure G(X; β) and known Y, W and σ_u^2. Estimation of the SIMEX estimate of β_x (denoted β_{simex} in the text below) proceeds as follows:

1. Initialise i =1.

2. Specify the variance inflation factor {λ_i} such that 0< λ_{i-1} <λ_i < λ_{i+1}

3. Generate a series of random normal deviates $U^*_{(i)} \equiv N(0, \lambda_i\, \sigma_u^2)$.

4. Estimate the synthetic covariate, $W^*_{(i)} = W + U^*_{(i)}$, the resulting series representing an increased error variance equal to (1+ λ_i) σ_u^2

5. Estimate $b^*_{(i)}$ from $Y = G(W^*_{(i)}; b^*_{(i)}) + \varepsilon_w$

6. Repeat the steps from 3 to 5 B times where B denotes the number of trials to be used and is chosen subjectively, and accept the average of all $b^*_{(i)}$ as its expected value $\beta^*_{(i)}$.

7. Repeat the steps from 2 to 6 for i =2,3,..n, where n is subjectively specified

8. Construct a relationship $\beta^*_{(i)} \sim \lambda_i$, denoted as $\beta^* = S(\lambda)$. The function used to relate these variables is usually chosen based on the data formulated in step 7. A straight line or a quadratic relationship are some possible choices of S(.).

9. Extrapolate to SIMEX estimate of the parameter $\beta_{simex} = S(\lambda=-1)$. Given the variance of the inflated error in step 4 equals $(1+\lambda)\sigma_u^2$ the case where (λ=-1) represents a scenario where no additive error is present in the model.

The SIMEX algorithm repeatedly adds noise with variance $\lambda\sigma_u^2$ to W (where λ>0) and computes the slope, which, in case of a linear model, consistently estimates $S(\lambda) = \beta_x\, \sigma_x^2 / [\sigma_x^2 + \sigma_u^2\, (1+ \lambda)]$, where σ_x^2 is the variance of the true covariate X. The SIMEX theory states that the solution $S(\lambda = -1)$ represents an asymptotically unbiased parameter estimate for β_x (Carroll et al., 1995).

In the example illustrated in Figure 2.1 we select λ = {0.0, 0.50, 1.00, … 5.00}. Note that λ = 0.0 represents the case where the estimated parameter value is the naïve estimate. The number of trials used for estimating $\beta^*_{(i)}$ was specified as 500 (B=500, see step 6). Figure 2.2 illustrates the relationship between β^* and λ. The SIMEX estimate of the parameter β_{simex} is 0.79. It is apparent that the naïve estimate (β_w= 0.55) offers a poor representation of the true parameter value (β_x= 0.80), which is ascertained accurately through the SIMEX extrapolation (β_{simex} = 0.79) described above. SIMEX estimates are closer to the true parameter values than the naïve estimates.

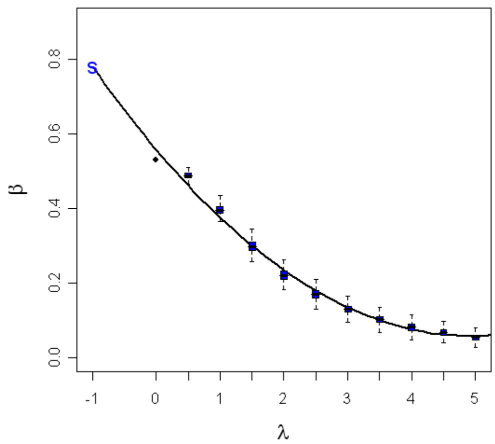


Figure 2.2 The regression coefficient β^* is plotted against λ. Note the bias due to increase in the error variance (λ). The point (β^*,λ = 0) is the naïve estimate and S(λ = -1) is the SIMEX corrected estimate, denoted by the letter 's'. The whiskers indicate the ± 5% confidence limit. The extrapolation is based on a quadratic function.

It should also be noted that the use of algorithm presented above may be constrained when there is limited knowledge on the extent of the error variance σ_u^2. Replicate measurements are often available in modelling in place of an estimate of σ_u^2. An example of such a situation is the error variance associated with the catchment averaged rainfall versus known point rainfall measurements at multiple locations in and around the catchment. There exists a non-parametric alternative of SIMEX where the multiple replicates are used as the basis of specifying the additive error structure in steps 2 to 4 of the above algorithm (Devanarayan and Stefanski, 2002).

We now compare SIMEX to existing hydrological research on the impact of uncertainties on parameter behaviour (Beven and Freer, 2001; Carpenter and Georgakakos, 2004; Jonsdottir et al., 2006; Lee et al., 1990; Rajaram and Georgakakos, 1989). The above referenced works present methods that are formulated by breaking down the functional form being modelled into distinct, separate steps, each of which can have an associated error term that is then subsequently ascertained. Consequently, the rationales used (variations of generalised likelihood measures and stepwise Kalman filter) often impose restrictions on the error structures associated with the processes considered (error usually assumed to be white Gaussian noise) and a simplification of the propagation of the input errors on the model outputs. For example Rajaram and Georgakakos (1989) assumed an error propagation logic whereby the total uncertainty was a linear combination of Gaussian input and Gaussian state uncertainty. The methods (Berry et al., 2002; Carroll et al., 1999; Rajaram and Georgakakos, 1989) that require assumptions about the distribution of unobserved true covariate X can be classed as 'structural estimators' while SIMEX is a 'functional estimator' which needs no such assumption (Staudenmayer and Ruppert, 2004). In addition to the above, most approaches for bias correction use either 'method of moments' or 'orthogonal regression' as their basis (Carroll et al., 1995). SIMEX uses the method of moments as its basis for ascertaining the true model parameters, implying that the variance inflation factor λ is ascertained based solely on the order two moment of the difference between observed and error added data. In contrast, orthogonal regression estimators (Huard and Mailhot, 2006; Kavetski et al., 2002; Kavetski et al., 2006a; Kavetski et al., 2006b) minimize the orthogonal distance (OD) between the points {Y,W} and the function Y=*G*(X; *β*) with unknown parameters {β_x, X}, as stated in Equation [4]:

$$OD(\beta) = \sum_{i=1}^{N} \left\{ \left(y_i - G(x_i;\beta)\right)^2 + \eta\left(w_i - x_i\right)^2 \right\} \qquad [4]$$

where the first difference term inside the summation over the N observations represents the deviation of observed responses from the modelled response using the true model parameters, and the second represents the deviation of the true covariates from the (erroneous) observed ones, η representing the relative uncertainty of the X and Y data, something that needs to be assumed in practice. It should be noted that as X represents the true (unknown) observation set, orthogonal regression estimators consider these as parameters, thereby increasing the effective number of parameters that are being estimated. In addition, specification of η can lead to an over correction for attenuation due to measurement errors (Carrol et al, page 29, 1995).

2.3 Application of SIMEX

The previous section outlined SIMEX using a synthetic setting (Figure 2.2). The form of the errors (probability distribution and variance) was assumed known and stationary. In a general context, the error distribution will seldom be completely known and will need to be ascertained based on multiple observations of the random variable (such as multiple raingauge readings based on which the catchment averaged rainfall is estimated if the method were to be applied for rainfall runoff model parameter estimation) or the mathematics used in its formulation (such as the averaging process in the catchment averaged rainfall). In a general setting, it is also possible that the error distribution is not stationary, but changes as a function of time. Such non-stationarity is common in many hydroclimatic systems, with the error variance reducing with time as instrument precisions have increased.

One such example where the error distribution is markedly non-stationary is the case of Sea Surface Temperature Anomaly (SSTA) data, often used as the basis of specifying General Circulation Model initial conditions, and serving as a basis for many empirical rainfall and streamflow probabilistic forecasting approaches (Chen et al., 2004; Goddard and Mason, 2002; Latif et al., 1998). Reconstructed, gridded, monthly sea surface temperature anomaly (SSTA) data (Kaplan et al., 1998), available from the Climate Data Library of the Lamont-Doherty Earth Observatory of Columbia University, New York, is used in the example discussed here. The SSTA data set was reconstructed based on point measurements of sea surface temperature using the Kaplan Optimal Smoother (OS) interpolation algorithm (Kaplan et al., 1997) . The reconstructed data set and associated error characteristics are available from 1856 onwards, at a resolution of 5° latitude by 5° longitude. The raw data sets used in Kaplan OS SSTA are collected by UK Meteorological Office and referred to as MOHSST5 (Bottomley et al., 1990). The interpolation procedure allows for estimation of the error variance at each time step/grid location. The error variance associated with the NINO3 El Nino Southern Oscillation (ENSO) index expressed as the average of SSTA over 5°N to 5°S and 150°W to 90°W, is illustrated in Fig ure 2.3.

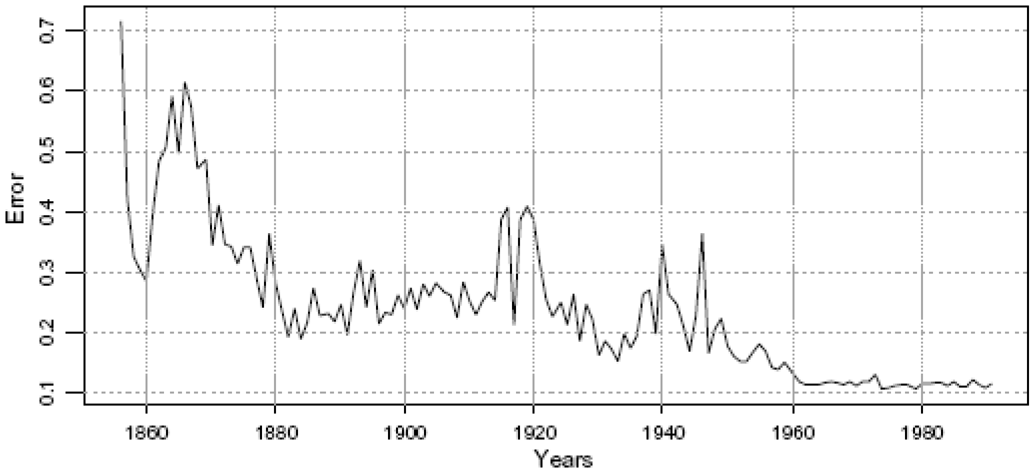


Figure 2.3 The error variance of monthly NINO3 estimates. Note higher of error prior to 1960. The post 1960 data contains low error due to improved instrumentation and more records.

The error variance is relatively small in more recent times (post 1960) due to better instrumentation and higher temporal and spatial recording density. The data during earlier periods has higher error variance (also note the noticeable increase in the error variance over the two world wars). Assuming the error variance to be either insignificant, or stationary in time, is likely to impact on the many applications the NINO3 data has been used for. We describe next an attempt to develop a predictive model that uses the above described SSTA data as the basis for infilling another well studied ENSO indicator, the Southern Oscillation Index (SOI).

2.3.1 Predicting southern oscillation index

The Southern Oscillation Index (SOI) represents the standardised pressure difference between Darwin and Tahiti with records extending back to 1876. The SOI is estimated from two controlled weather stations, and is thereby assumed to have less significant error in the context of the results reported below.

Consider a setting where one needs to estimate SOI data based on a predictive relationship developed using selected SSTA data. Assume that the relationship between the SSTA predictor variables and the SOI can be expressed using a simple linear regression model. We attempt to use the above setting to ascertain the impact the non-stationarity in the error variance associated with the SSTA (as illustrated in Figure 2.3) has on the resulting model. We formulate the regression model using SSTA corresponding to the NINO3 region outlined above, and develop the infilling model using the relatively error prone data from 1900 to 1960, using the later less erroneous data (1961 to 2003) as the basis for validating the relationship.

The lagged correlation of seasonal SSTA to the SOI of next two seasons and the following three months are estimated to serve as the basis for identifying predictors for use in the above mentioned model. These correlations are presented in Table 2.1 separately for the two segments the data has been divided into for development and validation purposes. Note that the seasons mentioned in the table referred to southern hemisphere, for example summer period spans from December to February. A consistent drop in correlation in pre 1960 data is experienced in most cases except when autumn SSTA (not shown here) is used. We discard the results affected by the autumn predictability barrier (Ruiz et al., 2005). The remaining three seasons result in 15 correlations of which 2 have very small values flagging them unsuitable under the linear setting the model is developed in. The resulting 13 correlations serve as the basis for the infilling/predictive models developed. Ten out of the 13 models show reduction in correlation in the 1900-1960 segments of the data, see Table 2.1. Hence our preliminary assumption of higher error in earlier part of the SSTA data appears valid and a likely indicator of loss of skill from any linear model calibrated using this data segment.

Table 2.1 Correlation of lagged sea surface temperature anomaly (SSTA) to southern oscillation index (SOI) for pre and post 1960 period. The values marked "*" represent cases where the correlation coefficients are small, and are not considered in this study (WIN: winter, SPR: spring, SUM: summer).

	SSTA	WIN		SSTA	SPR		SSTA	SUM
SOI	Post'60	Pre'60	SOI	Post'60	Pre'60	SOI	Post'60	Pre'60
SPR	-0.57	-0.56	SUM	0.73	0.72	AUT	-0.57	-0.53
SUM	-0.53	-0.65	AUT	-0.51	-0.42	WIN*	0.05	-0.25
Sep	-0.58	-0.55	Dec	-0.58	-0.58	Mar	-0.63	-0.55
Oct	-0.48	-0.44	Jan	-0.69	-0.58	Apr	-0.58	-0.49
Nov	-0.51	-0.48	Feb	-0.65	-0.68	May*	-0.13	-0.30

The above 13 relationships are expressed through simple linear regression models (without an intercept term), with parameters calibrated using the 1900 to 1960 data record. We validate the models using post 1960 data and measure the mean squares of errors of the prediction to quantify model performance. We next alter the parameter estimates using the SIMEX procedure described in earlier sections, using the known error variance of SSTA (varying as a function of time, unlike the SIMEX application described in the previous section) during the calibration period. We re-compute the residual variance of prediction during the validation period. The results from this analysis are presented in Figure 2.4 and Table 2.2.

Table 2.2 Predictive model performance using Winter SSTA predictors. The values tabled represent the sum of the Mean Square Errors associated with each predictive model, the summation enabling the overall improvement obtained due to SIMEX.

SOI	Before SIMEX	After SIMEX
Spring	9.6	9.4
Summer	9.5	9.5
September	12.0	11.7
October	13.3	13.2
November	15.6	15.3

Table 2.2 presents the improvements obtained using SIMEX for the validation period for the case where Winter SOI is the sole response variable. Figure 2.4 illustrates similar results obtained for all the response variables considered in the study. There is an overall improvement in validation fit with reduction in residual variance as can be inferred from both sets of results. While the improvements are not dramatic, they serve to remind us of the impact ignoring additive error (and error non-stationarity) can have when high dimensional models that have complex non-linear interactions associated with them are built and used for climate modelling applications.

2.4 Discussion and Conclusion

The total residual error in hydroclimatic models consists of errors in the (a) response variable, (b) model structure and (c) input data. The errors in input data introduce bias in parameter estimates. We demonstrate that SIMEX mitigates this bias from the parameters of selected models as long as the error distribution and any associated non-stationarity is known This research focuses on the overall improvement offered due to the SIMEX approach. Note that the extent of improvement of the predictions also depends on structural error and noise in the response data.

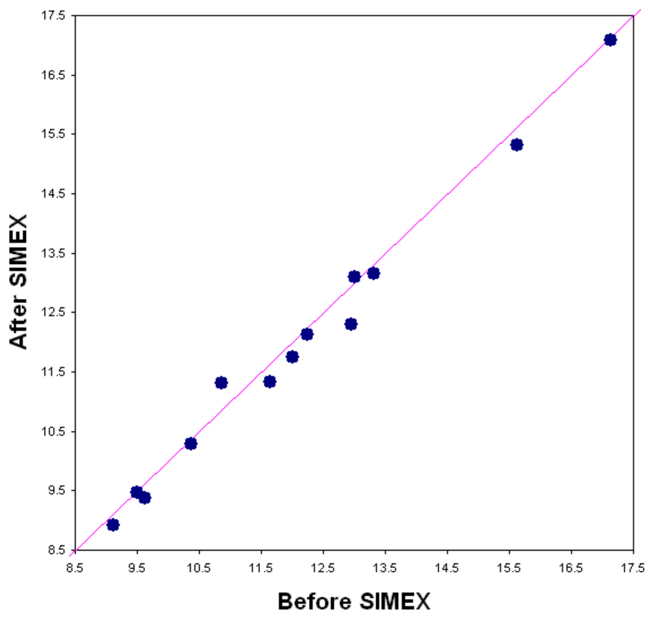


Figure 2.4 Sum of mean squares of errors during validation period before and after applying SIMEX. The points below the 1:1 line shown an improvement due to SIMEX

Parametric SIMEX requires the specification of input error distribution, something that is decided based on the causes that lead to the error being present. We feel the specification of the error distribution should be problem specific, for instance, errors in rainfall are often assumed to be log-normally distributed given that they have a lower bound of zero that cannot be violated. Multiplicative error in rainfall to preserve the number of rain-days (Kavetski et al., 2006b) or a combination of multiplicative and a small additive error (Carpenter and Georgakakos, 2004) have been used in prior rainfall-runoff studies. The statistical inference of spatial and temporal interpolation models of point measurements can be used to estimate the error structure (Kaplan et al., 1997). The general topic of measurement uncertainty is covered in various books (Barry, 1978; Dieck, 2002; Jaech, 1985; Rabinovich, 1993).

Parameter error is assumed independent of time and input error in the works of Rajaram and Georgakakos (1989) and Lee et al. (1990) and in follow up studies such as Jonsdottir (2006). We feel the SIMEX logic has certain advantages over the other methods of assessing uncertainty used more widely in hydrology (Beven and Freer, 2001; Carpenter and Georgakakos, 2004; Jonsdottir et al., 2006; Lee et al., 1990; Rajaram and Georgakakos, 1989). The main advantages is the flexibility SIMEX provides in allowing for a range of error structures (especially pertinent in our examples where the input error varies as a function of time), and the simplicity of the logic used in the method which will enable other researchers to develop it for a wider range of problems than what we have presented in our paper. Structural estimators and orthogonal regression based estimators (defined in the last paragraph of Section 2.2) tend to over parameterise in cases of complex systems (refer to discussion in concluding paragraph of Section 2.2). A functional (eg. SIMEX) estimator has advantages over structural estimators when model structures are complex (such as non-linear hydrological systems) and error distributions nonlinear and nonstationary. The application of Bayesian representations of orthogonal regression schemes (Huard and Mailhot, 2006; Kavetski et al., 2002; Kavetski et al., 2006a; Kavetski et al., 2006b) offer added benefits due to the utilisation of prior knowledge on the parameters involved. Note that SIMEX utilises similar prior knowledge while formulating the extrapolation function $S(\lambda)$ as well in ascertaining sample parameters b^* for each increment of the variance inflation factor as mentioned in the algorithm in Section 2.2. We envisage comparative studies in future to determine the relative strength and weaknesses of the orthogonal regression versus SIMEX method using historical hydrologic time series.

The focus of this research is to study parameter behaviour as a function of input uncertainty. The main example included in this paper refers to the predicting the Southern Oscillation Index using erroneous sea surface temperature anomaly data, where the error structure is markedly heteroskedastic as per the accuracy of observational networks that were available in the previous century. A preliminary synthetic study to test the stability of SIMEX in a typical non-linear hydrological setting of the Sacramento Rainfall Runoff Model was conducted and reported in Chowdhury and Sharma (2005). This preliminary study within a limited three parameter space indicated that SIMEX, in a Sacramento modelling context, is exhibiting a trend with respect to the variance inflation factor λ (and hence can be extrapolated),providing another example of how the SIMEX rationale can be used to ascertain parameters in typical hydroclimatic settings.

While the purpose of this paper was to expose the SIMEX rationale in a simple setting to a hydrological audience, there are several aspects of SIMEX that need further work and investigation. These includes issues related to the number of simulations needed, the number of increments of the variance inflation factor needed, the type of regression model (parametric or non-parametric) to be used to perform the extrapolation, and the confidence intervals associated with SIMEX estimated parameters, all require further study and investigation. Another issue to be investigated in greater detail is the specification of the error distribution for various hydroclimatic variables, and the special case where multiple input error terms (with possible co-dependence) may be needed in order to characterise the response.

While these issues have not been covered in the present study, we hope that our presentation provides readers with an appreciation of the importance of input errors, the problems in ignoring them in model building, the complications that arise when they are non-stationary, and the utility of the SIMEX procedure in specifying models for use in error free conditions. We intend to explore many of the limitations outlined above, along with the sensitivity of the procedures described to a range of factors including data length in future papers that follow on from this work.

2.5 Acknowledgement

We acknowledge the Lamont Doherty Earth Observatory of Columbia University, New York for making available the global sea surface temperature data. The authors like to thank Australian Bureau of Meteorology for Southern Oscillation Index data. Funding for this research came from the Australian Research Council and the Sydney Catchment Authority. This research used the freely available R statistical computing platform (R Development Core Team, 2004).

2.6 Reference

Barry, B.A., 1978. Errors in practical measurement in science, engineering, and technology. Wiley, New York.

Berry, S.M., Carroll, R.J. and Ruppert, D., 2002. Bayesian smoothing and regression splines for measurement error problems. Journal of the American Statistical Association, 97(457): 160-169.

Beven, K. and Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. Journal of Hydrology, 249(1-4): 11-29.

Bottomley, M., Folland, C. K., Hsiung, J., Newell, R. E., Parker, D. E., 1990. Global Ocean Surface Temperature Atlas, Her Majesty's Station Office, Norwich, England.

Brown, B.W. and Mariano, R.S., 1993. Stochastic simulations for inference in nonlinear errors-in-variables models. Handbook of Statistics, 11. North Holland, New York, 611-627 pp.

Butts, M.B., Payne, J.T., Kristensen, M. and Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. Journal of Hydrology(298): 242-266.

Carpenter, T.M. and Georgakakos, K.P., 2004. Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model. Journal of Hydrology, 298(1-4): 202-221.

Carroll, R.J., Maca, J.D. and Ruppert, D., 1999. Nonparametric regression in the presence of measurement error. Biometrika, 86(3): 541-554.

Carroll, R.J., Ruppert, D., Stefanski, 1995. Measurement Error in Nonlinear Models. Monographs on Statistics and Applied Probability, 63. Chapman and Hall, 305 pp.

Chen, D., Cane, M.A., Kaplan, A., Zebiak, S.E. and Huang, D., 2004. Predictability of El Niño over the past 148 years, Nature, pp. 733 - 736.

Chowdhury, S. and Sharma, A., 2005. Errors in Hydrological Variables and Their Effect on Model Parameters, MODSIM Conference, Melbourne.

Cook, J.R. and Stefanski, L.A., 1994. Simulation-Extrapolation Estimation in Parametric Measurement Error Models. Journal of the American Statistical Association, 89(428): 1314-1328.

Devanarayan, V. and Stefanski, L.A., 2002. Empirical simulation extrapolation for measurement error models with replicate measurements. Statistics & Probability Letters, 59(3): 219-225.

Dieck, R.H., 2002. Measurement Uncertainty: Methods and Application. Instrumentation, Systems and Automation Society (ISA), United States.

Fuller, W.A., 1987. Measurement Error Models. John Wiley & Sons, New York.

Goddard, L., Mason, S.J., 2002. Sensitivity of seasonal climate forecasts to persisted SST anomalies. Climate Dynamics, 19: 619-631.

Holcomb, J., 1999. Regression with covariates and outcome calculated from a common set of variables measured with error: Estimation using the SIMEX method. Statistcs in Medicine, 18(21): 2847-2862.

Huard, D. and Mailhot, A., 2006. A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc". Water Resources Research, 42: WR004661.

Jaech, J.L., 1985. Statistical analysis of measurement errors. Wiley, New York.

Jonsdottir, H., Madsen, H. and Palsson, O.P., 2006. Parameter estimation in stochastic rainfall-runoff models. Journal of Hydrology, 326(1-4): 379-393.

Kaplan, A. et al., 1998. Analyses of global sea surface temperature 1856-1991. Journal of Geophysical Research (C. Oceans). 103(C9): 8,567-18,589.

Kaplan, A., Kushnir, Y., Cane, M.A. and Blumenthal, M.B., 1997. Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. Journal of Geophysical Research (C. Oceans). 102(C13): 27,835-27,860.

Kavetski, D., Franks, S. and Kuczera, G., 2002. Confronting Input Uncertainty in Environmental Modelling. In: H.V. Gupta, Sorooshian, S., Rousseau, A. N. and R. Turcotte (Editor), Calibration of Watershed Models. AGU Water Science and Applications Series, Duan, pp. 49-68.

Kavetski, D., Kuczera, G. and Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research, 42(W03407).

Kavetski, D., Kuczera, G. and Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resources Research, 42(W03408).

Kim, C., Hong, C. and Jeong, M., 2000. Simulation-extrapolation via the Bezier curve in measurement error models. Communications in Statistics-Simulation and Computation, 29(4): 1135-1147.

Kim, J. and Gleser, L.J., 2000. SIMEX approaches to measurement error in ROC studies. Communications in Statistics-Theory and Methods, 29(11): 2473-2491.

Latif, M., D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, E. Schneider, 1998. A review of the predictablility and prediction of ENSO. Journal of Geophysical Research, 103(C7): 14,375-14,393.

Lee, S.J., Georgakakos, K.P. and Schnoor, J.L., 1990. Effects of Parameter Uncertainty on Long-Term Simulations of Lake Alkalinity. Water Resources Research, 26(3): 459-467.

Liang, K.Y. and Liu, X.H., 1991. Estimating equations in generalized linear models with measurement error. In: V.P. Godambe (Editor), Estimating Functions. Clarendon Press, Oxford.

Lin, X.H. and Carroll, R.J., 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American Statistical Association, 95(450): 520-534.

Linder, E., Patil, G.P. and II, G.S., 1993. Errors-in-variables analysis of extrapolation procedures in environmental toxiology. In: G.P. Patil and C.R. Rao (Editors), Multivariate Environmental Statistics. Elsevier Science Publishers B.V., pp. 227-254.

Marcus, A.H. and Elias, R.W., 1998. Some useful statistical methods for model validation. Environmental Health Perspectives, 106(Suppl. 6): 1541-1550.

Marshall, L., Nott, D, Sharma, A, 2007. Towards Dynamic Catchment Modelling: A Bayesian Hierarchical Modelling Framework. Hydrological Processes, 21:847-861 .

R Development Core Team, 2004. R: A language and environment for statistical computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rabinovich, S.G., 1993. Measurement errors : theory and practice. American Institute of Physics.

Rajaram, H. and Georgakakos, K.P., 1989. Recursive parameter estimation of hydrologic models. Water Resources Research, 25(2): 281-294.

Ruiz, J.E., Cordery, I. and Sharma, A., 2005. Integrating Ocean Subsurface Temperatures in Statistical ENSO Forecasts. Journal of Climate, 18(17): 3571.

Staudenmayer, J. and Ruppert, D., 2004. Local polynomial regression and simulation-extrapolation. Journal of the Royal Statistical Society Series B-Statistical Methodology, 66(Part 1): 17-30.

Stefanski, L.A., 2000. Measurement error models. Journal of the American Statistical Association, 95(452): 1353-1358.

Stefanski, L.A. and Cook, J.R., 1995. Simulation-extrapolation: The measurement error jackknife. Journal of the American Statistical Association, 90(432): 1247-1256.

Yapo, P.O., Gupta, H.V. and Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology, 181(1-4): 23-48.

CHAPTER THREE

This chapter reprinted the following article:

Chowdhury, S. and Sharma, A. (2008), A simulation based approach for representation of rainfall uncertainty in conceptual rainfall runoff models. *Hydrological Research Letters*, 2: 5-8. DOI: 10.3178/HRL.2.2.

Reprinted with permission from *Hydrological Research Letters*.

Permitted by the chief editor Dr. Yosuke Yamashiki,. dated: 15 August 2008

3 REPRESENTATION OF RAINFALL UNCERTAINTY IN CONCEPTUAL RAINFALL RUNOFF MODELS

ABSTRACT

One of the common contributors to the uncertainty in any rainfall runoff model is the error distribution within the rainfall inputs. The uncertain rainfall introduces systematic bias in the estimated parameters. We present here the application of a method, known as simulation extrapolation (SIMEX), to ascertain the extent of parameter bias. SIMEX requires a knowledge of the standard error associated with the rainfall at any given time step. With this knowledge, it generates multiple sets of rainfall with artificially inflated error variance, and then assesses whether this leads to any trend in the resulting parameters. This trend is then extrapolated back to assess the most suitable parameter value when the input is error free. The applicability of the method is investigated using a synthetic example where rainfall uncertainty is multiplicative and temporally invariant. This paper ascertained the bias trend in three key storage parameters of the Sacramento Rainfall Runoff Model representing surface and subsurface flow mechanisms respectively. This initial investigation confirmed the stability of SIMEX for use in hydrological model specification studies; which hints the possibility of embedding this simple method to improve runoff estimation.

3.1 Introduction

Rainfall is the prime input variable of any rainfall runoff model. The availability and quality of rainfall data varies in time and across multiple locations of most catchments. Rainfall runoff models require prior transformation of these point rainfall time series to a sub-catchment scale. The transformation of point to aerial scale introduces input error to the rainfall runoff model. What is the effect of this input error on the resulting flows? Given known transformation error profiles, can we reduce the total uncertainty caused by this error on the simulated flows? This paper discusses a method to minimise the impact of rainfall error on flow simulations from any rainfall-runoff model.

Errors in input data introduce systematic bias in model parameterisation during calibration (Carroll et al., 1995; Chowdhury and Sharma, 2007). The bias in parameter estimation of rainfall runoff model due to uncertain rainfall has been studied by Kavetski et al. (2002). Later studies (Kavetski et al., 2006a; Kavetski et al., 2006b) used an orthogonal regression scheme termed as ‘total least square’ to mitigate the effect input error. An alternative functional estimator termed Simulation Extrapolation, or SIMEX in short, was introduced for use in hydroclimatic studies by Chowdhury and Sharma (2007), that can mitigate parameter bias with fewer underlying assumptions. The Chowdhury and Sharma (2007) study was limited to hydrological applications in linear settings whereas this research letter investigates practical non linear extensions of the approach that include rainfall runoff modelling. More details on the rationale used are presented next.

3.2 Methodology

SIMEX, as the name implies, simulates parameter sensitivity to artificially inflated input error which in turn allows the extrapolation of parameters to an error free state. The method was originally developed in mid nineties (Cook and Stefanski, 1994; Stefanski and Cook, 1995) and more recently applied in hydrological area by Chowdhury and Sharma (2007). We briefly repeat the method here for the benefit of the readership, detail of which can be referred to publications mentioned above.

Consider an ideal case of a modelling response variable y in presence of the error free input variable x:

$$y = G(\beta_x, x) + \varepsilon_x \qquad [1]$$

Here, β_x is the true parameter of the regression $y \sim x$, G() is an underlying model, and ε_x is the associated total error term. In practise we seldom know the true input variable (x) and need to set up our model using uncertain input variable, say w. The input variable w results the naive parameter estimate β_w of the regression $y \sim w$ instead.

The SIMEX methodology requires a known error distribution for w with respect to x. Say, w includes an additive Gaussian error with variance σ_w^2 and mean zero. Synthetic input data w^* is generated adding artificially inflated error $\lambda\sigma_w^2$ to w, where the inflation factor is $\lambda > 0$. Say the resulting biased parameter estimate of $y \sim w^*$ is β^*. The simulation is repeated for a number of inflation factors for example $\lambda=\{0.5,1.0,1.5..\}$. Hence a regression relationship of $\beta^* \sim \lambda$ can be derived. The regression is extrapolated to $\lambda = -1$ to compute β_{simex}. For a number of generalised linear models (Carroll et al. 1995), analytical solutions dictate that $\beta_{simex} \approx \beta_x$.

3.3 Rainfall Runoff Model Parameters

The Rainfall runoff relationship can be highly non linear and difficult to model. While an analytical proof of SIMEX in simple linear settings may be viable, a more practical option for implementing the method in a nonlinear rainfall runoff modelling study is through a detailed numerical simulation. This section evaluates the utility of SIMEX in a practical hydrological setting using the Sacramento conceptual rainfall-runoff model.

The Sacramento Model

The Sacramento Model is a water balance model widely used in Australia (Boughton, 2005). Sacramento Model is also alternatively known as National Weather Service River Forecasting System (NWS-RFS) in USA (Burnash, 1975; Burnash et al., 1973). A general specification of the model is:

$$Q_t = Sac(I_t, E_t; \theta) + \varepsilon_t \qquad [2]$$

where *Sac*() is the Sacramento model which uses rainfall and evaporation data (I_t, E_t) to generate flow (Q_t) at time t with residual error ε_t . The model parameter vector θ consists of 16 parameters. The parameters define the following five major characteristics of the conceptual modelling system:

a) soil moisture storages,

b) rate of outflow,

c) percolation from upper to lower storages,

d) direct runoff from impervious areas, and

e) evapo-transpiration and deep seepage loss.

The Sacramento model has five soil moisture storages. The model essentially simulates water movements between storages, loss and routing as represented in Figure 3.1.

We use daily rainfall and evaporation at Golspie, NSW in Australia from 1980 to 1992 as a notional true estimate of catchment rainfall and evaporation, $\{I_t, E_t; t=1,2.. (12x365)\}$, the time series being shown in Figure 3.2. The average rainfall per wet day in the catchment equals 8 mm/day with a standard deviation of 9 mm. The fraction of wet days in the recorded period is 26 %.

We generate flow based on a set of given values for all 16 model parameters. The resulting flow sequence, illustrated in Figure 3.2, is the synthetic true flow series $\{Q_t; t=1,2.. (12x365)\}$ used in the results presented below. This study monitored only three storage parameters with notional true values of {60, 150, 38} mm. We appreciate that limiting the parameter number reduces the complexity of the Sacramento model. Nevertheless this significantly expedites computation without compromising the objective of demonstrating the application of our proposed approach in a rainfall-runoff modelling setting.

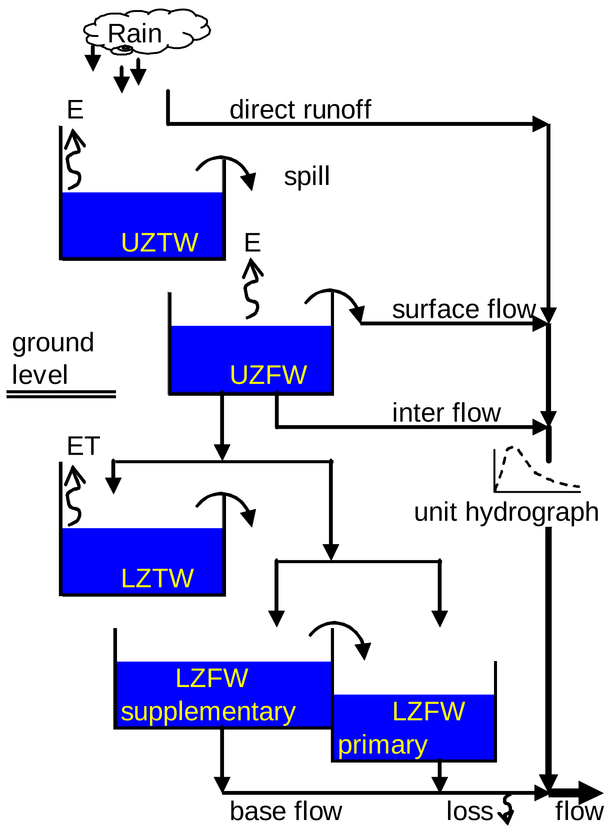


Figure 3.1 The Sacramento Model. The three storage parameters allowed to vary in this study are upper zone tension water (UZTW), upper zone free water (UZFW) and lower zone tension water (LZTW). The effect on lower zone free water (LZFW) and the other 12 parameters (not shown here) are not considered.

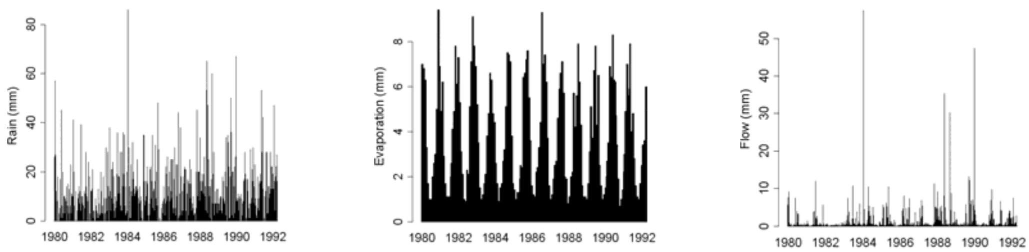


Figure 3.2 The recorded daily rainfall, evaporation and the synthetic flow used in this study.

The next step in this synthetic study is to generate a corrupted rainfall data set. The scope for errors in rain records during drier times is low (a dry day reading is error free). Nevertheless a storm may completely miss the rain gauge. Multiplicative error in rainfall (Kavetski et al., 2006b) or a combination of multiplicative and a small additive error (Carpenter and Georgakakos, 2004) have been assumed as appropriate in prior studies. We assumed the rain record error to be multiplicative in this study. Accordingly we artificially corrupt the rainfall series I_t by multiplying it by a lognormal series $U_t \equiv \log N(1, \log 0.1^2)$ as in Equation [3]. Figure 3.3 illustrates the effect of error density on rainfall estimates. The corrupted series, W_t becomes the notional recorded rainfall. For simplicity we assume the evaporation estimate to be error free as evaporation has much lesser spatial and temporal variability compared to rainfall.

$$W_t = I_t \quad U_t \qquad [3]$$

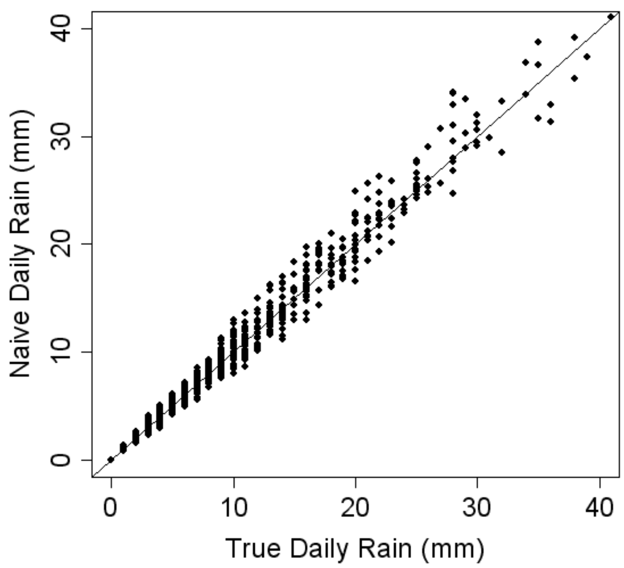


Figure 3.3. The relationship of true daily rainfall to naive rainfall due to log$N$(1, log 0.1^2) error.

Three parameters are allowed to vary keeping the remaining 13 parameters constant. They are UZTW (upper zone tension water storage capacity), UZFW (upper zone free water storage capacity) and LZTW (lower zone tension water storage capacity). The relative roles of these parameters are outlined in the Sacramento Model conceptual representation in Figure 3.1. We now need an objective function prior to calibrating a rainfall runoff model.

The design of objective function depends on the purpose behind the formulation of hydrologic models. Robust hydrologic models require multiple objective functions (Khadam and Kaluarachchi, 2004; Madsen, 2000). Our objective function to calibrate the Sacramento Model is based on flow hydrograph and flow duration curve. Equal weight has been assigned to both criteria as follows.

$$Obj(\theta) = Argmin \mid \sum_t (Q^{est}_t - Q_t)^2 + \sum_i (Q^{est}_i - Q_i)^2 \mid \qquad [4]$$

Where, Q_i, Q^{est}_i = observed and estimated flow at i^{th} percentile, the number of percentiles being set equal to the total number of time steps (t) in the sample. *Argmin*() minimises the objective function.

A suitable optimisation scheme is required to efficiently solve Equation [4] and thus estimate $\{\theta_p; p=1,2,3\}$. The choice of suitable optimisation schemes to solve non-linear hydrological models is intricate (Duan et al., 1992; Kuczera, 1997; Sorooshian et al., 1993; Yapo et al., 1998). We chose the L-BFGS-B optimisation algorithm (Byrd et al., 1995; Zhu et al., 1997). The L-BFGS-B is a limited-memory quasi-Newton algorithm for solving large non-linear optimization problems subject to simple bounds on the variables. It is intended for problems in which information on the Hessian matrix is difficult to obtain, or for large (or dense) problems.

The naive estimate $\{\theta^{naive}\}$ is attained by solving the objective function Equation [4] where the rainfall is W_t in the Sacramento model as shown:

$$Q^{est}_t = Sac(W_t, E_t; \theta_p^{naive} \mid \theta_q) \qquad [5]$$

Where, $p \equiv \{1,2,3\}$ and $q \equiv \{4,5.. 16\}$.

The rainfall data with error produces the following naive parameter estimates: $\theta^{naive} \equiv \{63, 151.5, 38.7\}$ mm. The naive estimate is a SIMEX step when $\lambda = 0$. Note the difference between these and the true parameter values {60, 150, 38} mm. Can SIMEX help in modifying these naive estimates so as to offer a better representation of the truth?

Application of SIMEX

We generate replicates of the corrupted rainfall W^*, with the addition of increasing levels of error $\{\lambda_i\sigma^2\}$, thereby resulting in an increasingly biased parameter vector θ_p^*. In this example, we use variance inflation factors $\lambda \equiv \{0.25, 0.5, 1.0, 1.5\}$ with 300 random trials each. The expected value of θ_p^* is estimated as the sample mean over the 300 trials. The distributions of the 300 trials of each parameter and interdependence among three parameters $\{\theta = \theta_1^*, \theta_2^*, \theta_3^*\}$ are analysed below followed by SIMEX extrapolation.

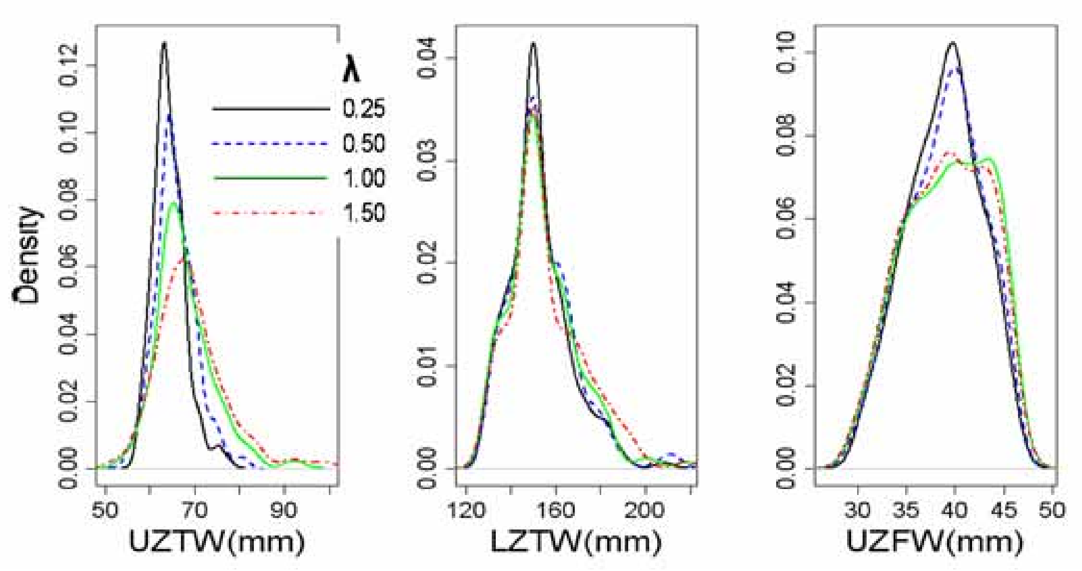


Figure 3.4 The empirical probability density function showing the affect of increase in the variance inflation factor (λ) on parameter estimates.

The probability density functions of the sampled parameters (Figure 3.4), ascertained using kernel density estimation approaches, show the increase in variance in the parameter sampling distribution with increase in λ. The effect of rainfall error is more significant for the upper storage parameter (UZTW), which is expected. The lower storage (LZTW) is less sensitive to input error due to the dampening effect of the prolonged accumulation involved. The three parameters exhibit near independence with respect to each other a feature that allows us to formulate separate relationships between λ and the estimated parameter value for each of the parameters considered.

Figure 3.5 illustrates the relationship between (λ~ θ_p*) for each p. The presence of a clear trend in the relationship allows the formulation of the following regression relationship:

$$\theta_p^* \sim G_p(\lambda) \qquad [6]$$

where G_p(.) denotes generic regression equations that may be linear or non-linear depending on the relationship that exists. In this example, a linear regression relationship is found to be satisfactory. The extrapolation to $G_p(\lambda = -1)$ gives us the SIMEX corrected estimate as shown in Figure 3.5 and listed (along with naive estimates and the true values) in Table 3.1. Note that the SIMEX corrected estimates are closer to the true values compared to the naive estimates.

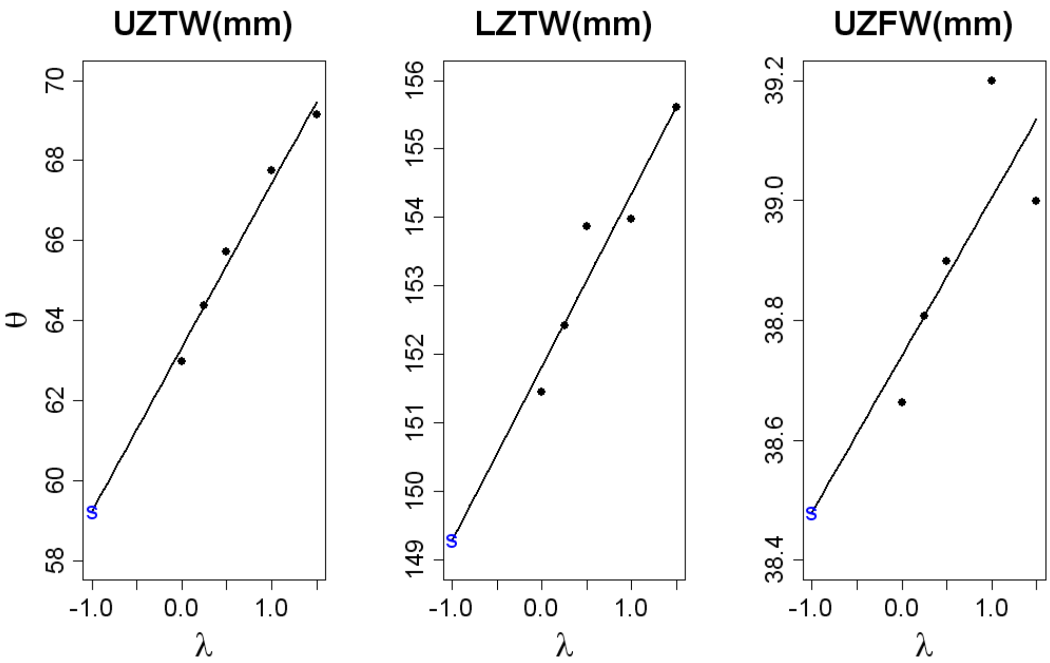


Figure 3.5. The biased parameter values are plotted against the multiples of error variance (λ). The point ($\theta|\lambda$ = 0) is the naive estimate and ($\theta|\lambda$ = -1) is the SIMEX corrected estimate (s).

Table 3.1. Improvement in parameter values due to SIMEX.

Parameters (mm)	True	Naive	SIMEX
UZTW	60	63	59.2
LZTW	150	151.5	149.3
UZFW	38	38.7	38.4

3.4 Discussion and Conclusion

Uncertainty in input variable results in systematic bias in parameter estimates. We ascertained that SIMEX helps to mitigate the bias of the Sacramento model parameters. The study demonstrated that parameters with higher influence on direct runoff are more sensitive to rainfall error. This research letter primarily aims to present the stability of SIMEX in a non linear rainfall runoff modelling setting. However the problem definition has been simplified in many respects which need further work. The three model parameters considered were independently related to the variance inflation factor λ. In a full 16 parameter setting, any likely interdependence of the parameter needs to be addressed using multivariate statistics.

SIMEX needs prior knowledge of the error distribution and any associated non stationarity in the rainfall time series. One of the future works will be the specification of error in rainfall data, which requires several considerations. The instrumentation error is often non stationary due to change in accuracy and method of measurements over time, which can be obtained from the manufacturers or experimentally quantified (Barry, 1978; Jaech, 1985). The translation of radar rainfall record to ground rainfall (Chumchean and Sharma, 2006) or downscaling of GCM simulations of rainfall and other atmospheric variables to rainfall at the local catchment scale (Mehrotra and Sharma, 2006) introduces uncertainty that may be possible to quantify and specify error distributions for use in a SIMEX or similar framework. The transformation of point rainfalls to aerial rainfall involves interpolation error which can be ascertained using available statistical methods (Kaplan et al., 1997). Alternatively non parametric SIMEX may be useful when multiple replicates (or possibilities) of rainfall scenario are available (Devanarayan and Stefanski, 2002).

Due to less onerous assumptions and computational burden, we envisage application of SIMEX to be a routine practice among practitioners to improve parameter estimation during any rainfall runoff model production.

3.5 Acknowledgement

We acknowledge the New South Wales Department of Water and Energy, Australia for rainfall and evaporation data. Funding for this research came from the Australian Research Council and the Sydney Catchment Authority. We recognize the contributors of the freely available statistical software R which has been used in this study (R Development Core Team, 2006).

3.6 Reference

Barry BA. 1978. *Errors in practical measurement in science, engineering, and technology*. Wiley: New York.

Boughton W. 2005. Catchment water balance modelling in Australia 1960–2004. *Agricultural Water Management* 77**:** 91-116. DOI:10.1016/j.agwat.2004.10.012

Burnash RJ. 1975. The NWS river forecast system–catchment modelling. *Computer Models of Watershed Hydrology*, V. P. Singh, Ed., Water Resources Publications: Colorado.

Burnash RJ, Ferral RL, McGuire RA. 1973. A Generalized Streamflow Simulation System–Conceptual Modelling for Digital Computers, Department of Water Resources: Californina.

Byrd RH, Lu P, Nocedal J, Zhu C. 1995. A limited memory algorithm for bound constrained optimization. *Scientific Computing* 16**:** 1190-1208.

Carpenter TM, Georgakakos KP. 2004. Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model. *Journal of Hydrology* 298**:** 202-221. DOI:10.1016/j.jhydrol.2004.03.036

Carroll RJ, Ruppert D, Stefanski LA. 1995. *Measurement Error in Nonlinear Models*. *Monographs on Statistics and Applied Probability 63*, Chapman and Hall; 305 pp.

Chowdhury S, Sharma A. 2007. Mitigating Parameter Bias in Hydrological Modelling due to Uncertainty in Covariates. *Journal of Hydrology* 340**:** 197-204. DOI:10.1016/j.jhydrol.2007.04.010.

Chumchean S, Sharma A. 2006. An integrated approach to error correction for real-time radar-rainfall estimation. *Journal of Atmospheric and Oceanic Technology* 23**:** 67-79. DOI: 10.1175/JTECH1832.1

Cook JR, Stefanski LA. 1994. Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association* 89**:** 1314-1328.

Devanarayan V, Stefanski LA. 2002. Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics & Probability Letters* 59**:** 219-225. DOI:10.1016/S0167-7152(02)00098-6

Duan Q, Sorooshian S, Gupta V. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28**:** 1015-1031.

Jaech JL. 1985. *Statistical analysis of measurement errors*. Wiley: New York.

Kaplan A, Kushnir Y, Cane MA, Blumenthal MB. 1997. Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *Journal of Geophysical Research (C. Oceans)* 102**:** 27,835-27,860.

Kavetski D, Franks S, Kuczera G. 2002. Confronting Input Uncertainty in Environmental Modelling. In *Calibration of Watershed Models*, Gupta HV, Sorooshian S, Rousseau AN, Turcotte R (eds); pp 49-68.

Kavetski D, Kuczera G, Franks S. 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research* 42:W03407. DOI:10.1029/2005WR004368

——. 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research* 42: W03408. DOI:10.1029/2005WR004376

Khadam IM, Kaluarachchi JJ. 2004. Use of soft information to describe the relative uncertainty of calibration data in hydrologic models. *Water Resources Research* 40**:** W11505. DOI:10.1029/2003WR002939

Kuczera G. 1997. Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research* 33**:** 177-185.

Madsen H. 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology* 235**:** 276-288. DOI:10.1016/S0022-1694(00)00279-1

Mehrotra R, Sharma A. 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. *Journal of Geophysical Research - Atmospheres* 111. DOI:10.1029/2005JD006637

R Development Core Team. 2006. R Foundation for Statistical Computing: Vienna. ISBN 3-900051-07-0. URL http://www.R-project.org.

Sorooshian S, Duan Q, Gupta VK. 1993. Calibration of rainfall runoff models: Application of global optimization to the Scramento Soil Moisture Accounting Model. *Water Resources Research* 29**:** 1185-1194. DOI:10.1029/92WR02617

Stefanski LA, Cook JR. 1995. Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association* 90**:** 1247-1256.

Yapo PO, Gupta HV, Sorooshian S. 1998. Multi-objective global optimization for hydrologic models. *Journal of Hydrology* 204**:** 83-97.

Zhu C, Byrd RH, Lu P, Nocedal J. 1997. Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. *ACM Transactions on Mathematical Softwares* 23**:** 550-556.

(blank page)

CHAPTER FOUR

This chapter reprinted the following article:

Chowdhury, S. and A. Sharma, 2008: Long Range NINO3.4 Predictions Using Pair Wise Dynamic Combinations of Multiple Models. *Journal of Climate*, **22(3)**, 793-805. DOI: 10.1175/2008JCLI2210.1

Reprinted with altered style and minor corrections.

4 LONG RANGE NINO3.4 PREDICTIONS USING PAIR WISE DYNAMIC COMBINATIONS

ABSTRACT

The interest in climate prediction has seen a rise in the number of modelling alternatives in recent years. One way to reduce the predictive uncertainty from any such modelling procedure is to combine or average the modelled outputs. Multiple model results can be combined such that the combination weights may either be static or vary over time. This research develops a methodology for combining forecasts from multiple models in a dynamic setting. We mix models on a pair wise basis using importance weights that vary in time reflecting the persistence of individual model skills. Such an approach is referred to here as a dynamic pair wise combination tree, and is presented as an improvement over the case where the importance weights are static or constant over time. The pair-wise importance weight is modelled as a product of a 'mixture ratio' and a 'bias direction', the former representing the fraction of the absolute residual error associated with each of the paired models, and the latter representing an indicator of the sign of the two residual errors. The mixture ratio is modelled using a generalised autoregressive model and the bias direction using ordered logistic regression.

The method is applied to combine three climate models, the variables of interest being the monthly sea surface temperature anomalies averaged over the NINO3.4 region from 1956 to 2001. We test the combined model skill using a 'leave ± 6 months out cross validation' approach along with validation in 10 year blocks. This study attained a small but consistent improvement of the predictive skill of the dynamically combined models compared to the existing practice of static weight combination.

4.1 Introduction

Climate models vary in complexity from simplistic conceptualisations of the underlying physics, to statistical or empirically based methods, to detailed physical representations of the processes involved. The relative strengths and weaknesses of climate prediction models vary depending on the assumed model structure, the data quality, the period of calibration and the method of validation. The increase in computing power and availability of more accurate input data has resulted in significant improvements in climate predictions. However the fact remains that each type of model is often able to capture some aspect of the underlying behaviour better than the other. The differential improvement in prediction, achieved by investigating a single modelling approach in isolation, diminishes asymptotically. Climate modellers are now combining various models in order to exploit the strength of individual approach and reduce the variance of predictive uncertainty (Barnston, A. G. *et al.*, 2003: Colman, A. W. and Davey, M. K., 2003: Greene, A. M. *et al.*, 2005: Peng, P. *et al.*, 2002: Raftery, A. E. *et al.*, 2005: Robertson, A. W. *et al.*, 2004: Sharma, A. and Lall, U., 2004). The combination parameters are usually estimated based on overall performance of the component models. The consideration of persistence of the component model skills while building a combination of predictive models is introduced in this paper. The method is referred to as a pair wise dynamic model combination, the term dynamic being used to denote the fact that the mode of combination varies with time, with combination weights (or importance weights as they are referred to later in the paper) being modelled on the basis of the persistence they exhibit in some local time window. The aim is to improve upon the existing method of combining model predictions that overlooks persistence in the individual model skills. We present a general review of the existing developments that set the background of our current research in the remainder of this section.

The rationale behind model combination is the statistical principle that the weighted mean of two zero centred symmetrical distributions has a lower variance. This principle raises the possibility that two or more inaccurate but independent predictions of the same future event can be combined to yield a prediction that is on average more accurate than either of them taken individually (Bates, J. M. and Granger, C. W. J., 1969: Fraedrich, K. and Smith, N., 1989: Granger, C. W. J. and Newbold, P., 1977: Sanders, F., 1963: Thompson, P. D., 1977). The advantage of combining predictions using various methods is researched in the field of biometrics, econometrics and decision sciences (Dawes, R. *et al.*, 1994: de Menezes, L. M. *et al.*, 2000: Larrick, R. P. and Soll, J. B., 2006: Phillips-Wren, G. E. *et al.*, 2004). Clemen (1989) drew conclusions based on an extensive literature search that combining forecasts leads to increased forecast accuracy. Accordingly, the ensembles of various models are being routinely used now to issue predictions in various disciplines (Armstrong, J. S., 2001: Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T., 1999).

The methods of model combination can be classed into two broad categories: *static combination* and *dynamic combination*. The static combination method, as the name implies, leads to a weighted average output in which the weights are time invariant and hence does not consider any temporal variations of the component model skill. Variations of such a weighted average combination (which includes Bayesian model averaging) have been used in a variety of applications such as rainfall runoff modelling (Granger, C. W. J. and Newbold, P., 1977: Kim, Y. O. *et al.*, 2006: Marshall, L., 2006: McLeod, A. I. *et al.*, 1987: Ragonda, S. K. *et al.*, 2006: Shamseldin, A. Y., O'Connor, K.M., Liang, G.C., 1997: Xiong, L. *et al.*, 2001) and climate modelling (Coelho, C. A. S. *et al.*, 2004: Fritsch, J. M. *et al.*, 2000: Peng, P. *et al.*, 2002: Raftery, A. E. *et al.*, 2005: Rajagopalan, B. *et al.*, 2002: Robertson, A. W. *et al.*, 2004). The weighted average combination forms the benchmark against which we compared the performance of our proposed pair wise dynamic combination approach. It should be noted that estimation of the weights can be performed using an equal weighting for all models, or a weighting that reflects the accuracy of individual models, or using more appropriate optimisation based approaches that maximise the performance of the weighted combination output (Coelho, C. A. S. *et al.*, 2004: Doblas-Reyes, F. J. *et al.*, 2005: Kondrashov, D. *et al.*, 2005: Pavan, V. and Doblas-Reyes, F. J., 2000: See, L. and Abrahart, R. J., 2001: Xiong, L. *et al.*, 2001). However, such an approach is unable to allow dynamic variations in weights to form outputs that resemble more the outputs from the better performing models at any given point in time.

The dynamic combination allows the combination weights to vary over time. This combined outputs to take into account local nonstationarities and inhomogeneities in individual model outputs, thereby resulting in a forecast that is less susceptible to sudden and unexplained variations. The early attempt of dynamic combination was in the form of a switching regression model (Deutsch, M. *et al.*, 1994) which later evolved into more complex non linear combinations (Lundberg, S. *et al.*, 2000: Terui, N. and van Dijk, H. K., 2002: Zou, H. and Yang, Y., 2004). These studies are related to forming econometric models with the relevant papers appearing in econometric and applied statistics literature, with no similar work being reported in climate science. We present next an approach that dynamically mixes climate prediction models in a pair wise hierarchical tree structure, offering significant advantages to model combinations which are of a more static form.

4.2 Model Combination

4.2.1 Static combination

We first introduce a static combination of m=1,2,… M models for a period of t = 1,2,…$t_{\max}$. Let us define component predictions as {$\hat{u}_{m,t}$; m=1,2,..M ; t=1,2..,$t_{\max}$ }, with residual error as {$e_{m,t}$; m=1,2,..M ; t=1,2..,$t_{\max}$ } so that:

$$y_t = \hat{u}_{m,t} + e_{m,t} \qquad [1]$$

Where y_t is the observed response at time t.

Then the combined prediction $\hat{y}_t^{(s)}$ is ascertained as:

$$\hat{y}_t^{(s)} = \sum_m \hat{u}_{m,t}\, w_m^{(s)} \qquad [2]$$

Where, $w_m^{(s)}$ are the static weights, superscript (s) is for static weight, conditional to $w_m^{(s)} \geq 0$ and $\sum_m w_m^{(s)} = 1$.

The parameter vector $\mathrm{w}^{(s)}$ = {$w_m^{(s)}$; m=1,2,..M } is estimated through a constrained minimisation of the error variance of $\hat{y}^{(s)}$ = {$\hat{y}_t^{(s)}$; t=1,2..,$t_{\max}$ }, under the constraint that the parameters lie in the range 0 to 1 and sum up to unity. This constraint reflects the assumption that each component model is unbiased. For a two model case the maximum likelihood estimate of $w^{(s)}$ from a bi-variate normal error distribution can be derived to be:

$$w^{(s)} = \frac{\sum_t (e_{2,t}^{\ 2} - e_{1,t} \cdot e_{2,t})}{\sum_t (e_{1,t}^{\ 2} + e_{2,t}^{\ 2} - 2 \cdot e_{1,t} \cdot e_{2,t})} \qquad [3]$$

where, $e_{1,t}$ and $e_{2,t}$ are residual errors of model 1 and model 2 respectively.

Note that this paper follows notations of lower case *italics* for scalar values (e.g.: $\hat{u}_{m,t}$, $w_m^{(s)}$) and straight roman (e.g.: u_m, $\mathrm{w}^{(s)}$) for vectors and **bold** fonts (eg: **u, w**) to represent higher dimension matrices if not mentioned otherwise (exception being *M* used for a scalar value). Vector series are enclosed by curly brackets. Functions are denoted by italic names followed by brackets, such as minimum being specified as *Min*(.).

This weighted average combination method is referred as static combination hereafter due to its time invariant weight $\mathrm{w}^{(s)}$. This paper proposes the dynamic weight instead, $\{\omega_t\}$, which incorporates the persistence of component model skills as described in the rest of this Section. The error variance of the static combination prediction $\hat{\mathrm{y}}^{(s)}$ is used as the benchmark of performance of this proposed method.

4.2.2 Paired dynamic combination

Consider a case of combining predictions of a pair comprised of i^{th} and j^{th} component models $\{m = i, j\}$ using dynamic weights. If the component predictions at time t are $\hat{u}_{i,t}$ and $\hat{u}_{j,t}$ then the two models can be combined as follows:

$$y_t = \hat{u}_{i,t}\, \omega_t + \hat{u}_{j,t}\, (1 - \omega_t) + \ddot{e}_t \qquad [4]$$

Here $\ddot{e}_t$ is the residual of the combination where the true weight ω_t is available. We continue the assumption of the static weight formulation that the component predictions are unbiased and hence restrict the weights within 0 to

1. This constraint reduces the serial correlations in the combined forecast error (Aksu, C., Gunter, S., 1992).

Early research on model combination by Bates and Granger (1969) acknowledged the possible nonstationarity of Equation [3] and hence the need for estimating weights dynamically. The approach adopted in this and earlier studies are to investigate the possibility of using a dynamic structure in formulating the weights. Two possible ways of specifying the autoregressive structure in the dynamic weights are (Granger, C. W. J. and Newbold, P., 1977: McLeod, A. I. *et al.*, 1987):

$$\hat{\omega}_{t+1} = \lambda \hat{\omega}_{t-1} + (1-\lambda) \frac{\sum_{\tau=t-h}^{t} e_{2,\tau}{}^{2}}{\sum_{\tau=t-h}^{t} (e_{1,\tau}{}^{2} + e_{2,\tau}{}^{2})} \quad [5]$$

$$\hat{\omega}_{t+1} = \frac{\sum_{\tau=1}^{t} \gamma^{\tau} e_{2,\tau}{}^{2}}{\sum_{\tau=1}^{t} \gamma^{\tau} (e_{1,\tau}{}^{2} + e_{2,\tau}{}^{2})} \quad [6]$$

Where, t is the current time, h is a time bandwidth representing a local window centred around time t, $1 \geq \lambda > 0$ and $\gamma \geq 1$ are parameters that control the degree of autocorrelation. Note that these methods are primarily based on precision (inverse of prediction error variance) of component predictions. The case of ($\lambda = 0$, $h = t-1$) in Equation [5] and $\gamma = 1$ in Equation [6] collapses them to the static weight estimate of Equation [3] when combining two independent predictions.

Our proposed method starts by first computing a time series of target weights $\{\omega_t \in [0,1]\}$ that would produce perfect combined hindcast out of the component hindcast pairs. The target weights are predicted using generalised linear models (Chandler, R. E., 2005: Dunteman, G. H. and Ho, M. R., 2006: Helsel, D. R. and Hirsch, R. M., 2002: McCullagh, P. and Nelder, J. A., 1989: Yang, C. *et al.*, 2005). Generalised linear regression requires that the response variable belongs to exponential family of distributions, in contrast the target weights which follow a beta distribution (Bates, J. M. and Granger, C. W. J., 1969). The requirement is met by formulating separate linear models in two steps.

The first step uses a generalised linear autoregressive (GLAR) model (Shephard, N., 1995: Yu, L. *et al.*, 2005) as the basis of predicting the *mixture ratio* r_t :

$$r_t = |e_{j,t}| / (|e_{j,t}| + |e_{i,t}|) \qquad [7]$$

The GLAR is a special case of a generalised linear model that includes both autoregressive and exogenous covariates. Exogenous covariates imply the predictors external to the ones used in the component models and thus potentially adjoin additional predictive information.

GLAR estimates the predicted mixture $\hat{r}_t$ ratio as follows:

$$g(\hat{r}_{t+1}) = \theta_t + \varphi\, r_{t-.} + \psi\, z_{t-.} \qquad [8]$$

Where,

$r_{t-.}$: {1, $r_{t-h(1)}$, $r_{t-h(2)}$, …}, stepwise autoregressive covariates at lags of $h(1)$, $h(2)$, …, representing the persistence that is exhibited in r_t ;

$z_{t-.}$: {$z_{1,t-.}$, $z_{2,t-.}$, …}, exogenous covariates at earlier times (subscript $t\text{-.}<t$),

θ_t : seasonally variant intercept, varying from one season to the other, but not varying across years; and

φ, ψ : $\{\varphi_0, \varphi_1, \varphi_2, \ldots, \psi_1, \psi_2, \ldots\}^T$ the time invariant model parameters and the intercept (φ_0).

$g(.)$: a function transforming the response variable known as link function of the generalised linear model.

The link function $g(.)$ is chosen in a way that it transforms the bounded *mixture ratio* ≡ {0 → 1} to unbounded values ≡ {-∞ → +∞}. This research applies the following *Logit*(.) link function, which was used in the studies of forecast probabilities (Carrasco, J. A. and Ortuzar, J. D., 2002: Kamstra, M. and Kennedy, P., 1998).

$$Logit(r_t) = \log\,[r_t / (1 - r_t)] \qquad [9]$$

GLAR parameters are estimated using Maximum Likelihood of the beta binomial distribution (Gelman, A. and Hill, J., 2006: Yang, C. *et al.*, 2005) for the response r_t, which was found to be suitable in the context of the over dispersion of **r** described in later sections.

The mixture ratio, Equation [7] is not sufficient to keep $\ddot{e}_t^2 \le Min(e_{i,t}^2, e_{j,t}^2)$. The second step introduces additional criteria that aim to identify the direction of the bias of each model. The models are combined based on r_t only when $e_{i,t}$ and $e_{j,t}$ have opposing sign i.e. two predictions are bracketing the true value. On the other hand while both predictions exhibit bias in the same direction the better prediction is chosen ignoring r_t. This shields the combination from erratic high outlier with same sign. The *bias direction* $\{b_t;\ t = 1,2,..t_{max}\}$ is mapped into three categories (see Figure 4.1 (top)) where b_t is a categorical variable as follows:

$$b_t = \begin{cases} mix & 0 > e_{j,t} / e_{i,t} \\ zero & 0 < e_{j,t} / e_{i,t} < 1 \\ one & 1 < e_{j,t} / e_{i,t} \end{cases} \qquad [10]$$

The optimum measure of ω_{t+1} is defined as follows:

$$\omega_{t+1} = \begin{cases} r_{t+1} & \text{when } b_{t+1} = mix \\ 0 & b_{t+1} = zero \\ 1 & b_{t+1} = one \end{cases} \qquad [11]$$

The prediction of ω_{t+1} is done in a two step process. The first step involves predicting r_{t+1}, the mixture ratio model using GLAR. The second step is predicting the bias direction b={*mix*, *zero*, *one*} using an ordered logistic regression (OLR) model (Agresti, A., 1996). In OLR, the cumulative probability of b, $\hat{P}$ (b) is estimated as:

$$Logit\,[\hat{P}\,(b_{t+1} = mix)] = \alpha_1 + \mathrm{x}_t\,\beta \qquad [12]$$

$$Logit\,[\hat{P}\,(b_{t+1} = mix \text{ or } b_{t+1} = zero)] = \alpha_2 + \mathrm{x}_t\,\beta$$

where α={α_1, α_2} are intercepts and x_t ={$x_{0,t}$, $x_{1,t}$, $x_{2,t}$, ..} are predictor vectors inclusive of a periodic intercept ($x_{0,t}$), and autoregressive and any exogenous covariates, and β={1, β_1, β_2, ..}^{T} are model parameters. No third equation is necessary since P(b = *one*) = 1 - P(b = *mix* or b = *zero*). The logic for generation of this three category ordered regression variable is presented in Figure 4.1(bottom).

If the predicted mixture ratio is $\hat{r}_t$ and the bias direction is $\hat{b}_t$ then the fitted dynamic weight ($\hat{\omega}_t$) of the pair wise model combination can be estimated from Equation [11]. The dynamically combined prediction of the hydrologic response variable, $\hat{y}_t^{(d)}$, is as follows:

$$\hat{y}_t^{(d)} = \hat{u}_{i,t}\,\hat{\omega}_t + \hat{u}_{j,t}\,(1 - \hat{\omega}_t) \qquad [13]$$

4.2.3 Multiple model combination

The last section presented the basis for a pair wise combination of models. The exercise is now extended to M component models, where M > 2. We propose a paired combination hierarchical tree as shown for a four model case in Figure 4.2.

Denoting the component prediction errors of the component models as **e** = {$e_{m,t}$; m=1,2,..M; t=1,2..,t_{max}}, one can estimate the variance-covariance matrix of the residuals Cov(**e**) as {c_{ij}; i=1,2,..M; j=1,2,..M}. The model pair with smaller covariance has a higher potential of improvement after combination.

Hence the model pairing is performed by first sorting the models in order of their individual residual variance, and then starting from the lowest variance model and finding its pair as the model with which it has the lowest covariance. This process is repeated for the models that remain until all models are exhausted. If the number of component models is even, one would expect all models to be paired, if not, one would expect one component model to remain on its own. In the notation used in Figure 4.2, the indices of the models have been altered to reflect the pairs as (1,2) and (3,4). This notation will be followed in the remainder of this paper.

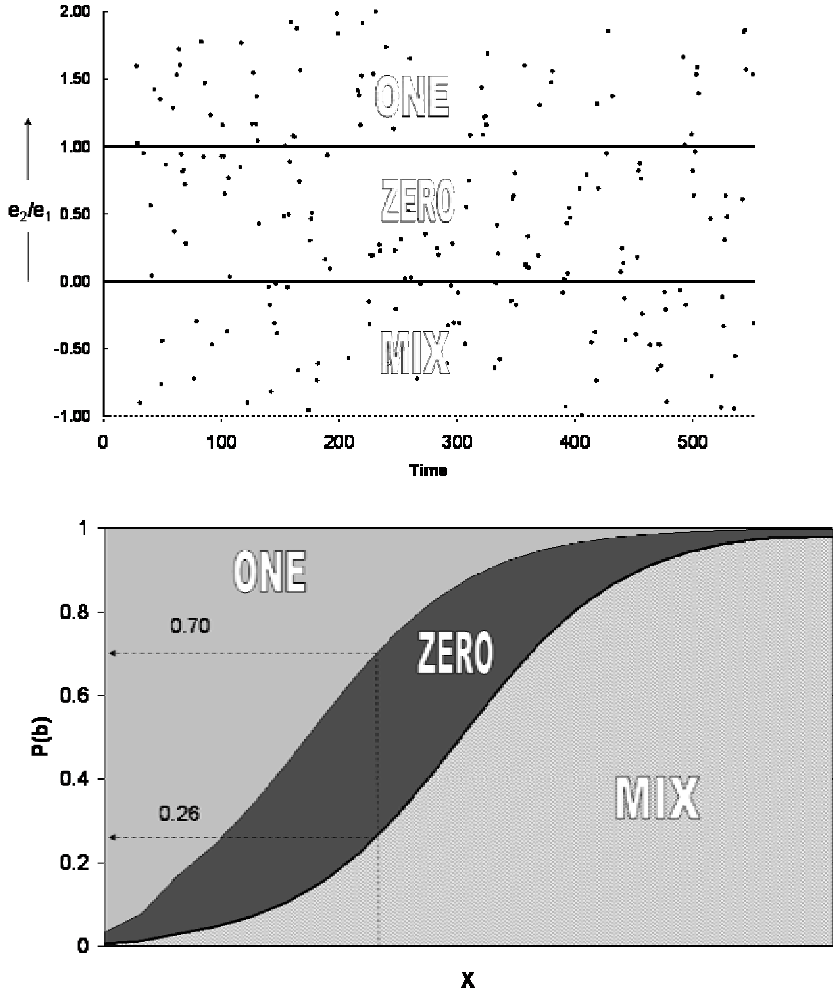


Figure 4.1 (top) The classification method of the bias direction variable. The time series of the ratio of two model residuals (e_2/e_1) are grouped into three zones. The residuals are classified into a three category response variable {*mix*, *zero*, *one*} as shown. (bottom) The simple ordered logistic model for a

three category response variable. The regression lines are dividing the probability space. For a given value of the predictor x, the dashed line is showing P(b=*mix*) equal to 0.26, P(b=*zero*) equal to 0.70-0.26 and P(b=*one*) equal to 1-0.70.

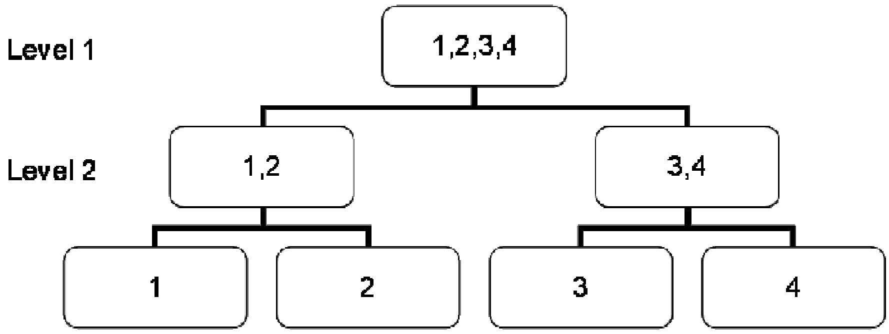


Figure 4.2 The tree showing pair wise hierarchical mixing of four component models.

The hierarchical combination tree will have multiple levels depending on the number of component models present. The hierarchical tree contains l levels which satisfy the following constraints: $2^l \geq M$ and $2^{l-1} < M$ where the exact value depends on the binary divisibility of M. This hierarchical tree uses the same (M -1) number of weight parameters as the static combination method described in Section 4.2.1. If $w_{t,i}^{(k)}$ includes $\{\omega_t \in w_{t,i}^{(k)}\}$ and represents the i^{th} weight time series vector at k^{th} level of the mixing tree then the weight matrix **W** can be shown as following:

$$\mathbf{W} = \{w_{t,i}^{(k)};\ k=1,2,..l;\ i=1,2,..n_k,\ t=1,2,..t_{\max}\}, \qquad [14]$$

where, $n_k \leq 2^{k-1}$.

The predicted value of **W** consists of (M-1 x $t_{\max}$) elements of $\hat{\omega}_t$ where the full set of component predictions are $\hat{\mathbf{u}} = \{\hat{u}_{m,t};\ m=1,2,..M;\ t=1,2..,t_{\max}\}$. The hierarchical extension of Equation [13] for the tree shown in Figure 4.2 where M =4 is as follows,

$$\hat{y}_t^{(d)} = \hat{u}_{1,t}\,(w_{t,1}^{(1)} \bullet w_{t,1}^{(2)}) + \hat{u}_{2,t}\,(w_{t,1}^{(1)} \bullet (1- w_{t,1}^{(2)}))$$
$$+ \hat{u}_{3,t}\,((1- w_{t,1}^{(1)}) \bullet w_{t,2}^{(2)}) + \hat{u}_{4,t}\,((1- w_{t,1}^{(1)}) \bullet (1 - w_{t,2}^{(2)})) \quad [15]$$

4.2.4 Model combination algorithm

The algorithm for combining M models using the pair wise dynamic procedure described above is as follows.

i) Index the models in such a way that the pairs satisfy the logic in Section 4.2.3.

ii) Choose the pair $\{\hat{u}_{1,t}$, $\hat{u}_{2,t}$;t=1,2..,$t_{\max}$ } and compute the target mixture ratio and bias direction $\{r_t$, b_t; t=1, 2…$t_c\}$ for the period of calibration t_c using Equations [7] and [12].

iii) Identify any autoregressive structure in {r, b} and plausible exogenous predictors using a model selection algorithm such as the Akaike Information Criterion (AIC) (Chambers, J. M., 1992: Hastie, T. and Pregibon, D., 1992: Hastie, T. *et al.*, 2000). Ascertain parameters φ, ψ, α and β of the selected model.

iv) Apply the developed model to obtain estimates of combination weights for a forecast period t_{c+} , $\{\hat{\omega}_t$; t= t_{c+} }.

v) Repeat step ii) to iv) for all M/2 pairs of combination, if M is even. Otherwise repeat for (M-1)/2 pairs and add the remaining component model at lower tree level.

vi) Repeat step 1 to 5 for all l levels of the hierarchical tree and thus estimate the weight matrix $\hat{\mathbf{W}}$.

Compute the final estimate by Equation [15]. The error variance of $\hat{y}^{(d)}$ should be narrower than that $\hat{y}^{(s)}$.

4.3 Application

The pair wise dynamic combination approach is applied to three component models selected from the pool of models available to predict globally gridded sea surface temperature anomalies (SSTA). The base of the anomalies was the Global Ocean Surface Temperature Atlas (GOSTA) Climatology of 1951 to 1980 (Bottomley, M. *et al.*, 1990). The extended SST data set, reconstructed at the US National Climate Data Centre (Smith, T. M. and

Reynolds, R. W., 2002), was used as observed SST. The component models predicted monthly SSTA at NINO3.4 region at 3 months in advance. For example, the SSTA value in April, May or June 1980 corresponds to the forecast of those months as made in January, February or March 1980 respectively. The first set of the three models was developed at University of California, Los Angles, USA, hereafter referred as UCLA model (Kondrashov, D. *et al.*, 2005). This is a multilevel quadratic inverse stochastic model formulated using global sea surface temperature data from 1950 to 2003 with an emphasis on ENSO variability. The second set of the three models was developed at the Climate Prediction Centre of the National Oceanic and Atmospheric Administration, USA. It uses a statistical method known at constructed analogue and referred to as the CACPC model (van den Dool, H., 2000). The third model was prepared by the Demeter project of European Centre for Meteorological Forecast and referred to as the ECMF model (Palmer, T. N. *et al.*, 2004). The concurrent hind-casts during the period of January 1956 to December 2001 of these component models are used as the basis of evaluating the accuracy of the model combination procedure. This study used an available set of hind cast realisation only and it may not be the most up to date version of the component model. All SSTA time series, except UCLA, were downloaded from the data library of International Research Institute for Climate and Society, New York (accessed on February 2006). The UCLA data was collected as *per comm.* D Kondrashov. The relative performances of our hind cast data set (represented as residuals from the observed) are illustrated in Figure 4.3. Note that the component model pair, which shows a low residual covariance, offers a potential of improvement based on the pair wise dynamic combination approach.

This research briefly trialled the two existing dynamic weight estimation methods presented in Equations [5] and [6]. The proposed GLAR plus OLR method is tested in greater detail. The first step in the proposed approach is the identification of relevant predictors, followed by an evaluation of the resulting model in a predictive sense. Details on each of these are presented next.

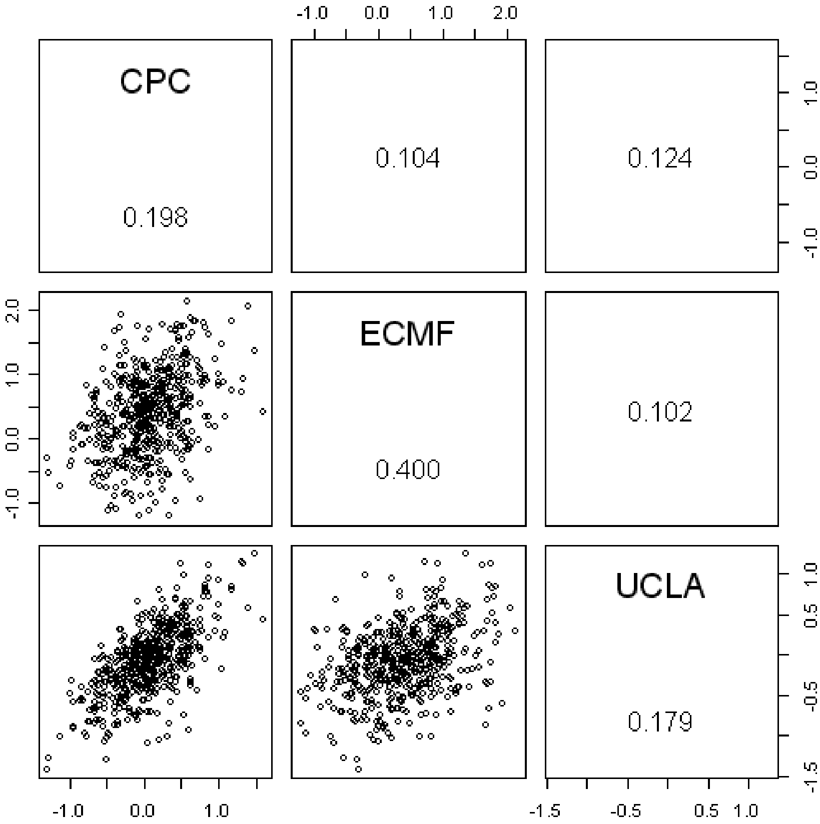


Figure 4.3 The paired plot of the three model residuals. The value in the diagonals shows the variance of the individual model error, the numbers in upper boxes are the covariance. The lower covariance has higher potential of improvement via combination.

4.3.1 Predictor selection

The predictors for the mixture ratio ($r_{t-.}$) model in Equation [8] are ascertained from lagged values of the response (r_t) over the past 12 time steps (months). Predictors for the categorical bias direction (b_t) are selected from lagged values of the ratio $e_{j,t}/e_{i,t}$. This ratio (termed residual ratio) is constrained to lie within [-1, 2] to avoid numerical instability when $e_{i,t} \approx 0$. The inclusion of hydrologic variables to the pool of candidate exogenous predictors involves detail knowledge of the component model pairs. We do not include any exogenous variables at this stage of the research.

The final predictor vector is chosen using standard statistical model selection procedures involving partial autocorrelation to the response, backward stepwise model selection using the partial F test (Chambers, J. M., 1992: Hastie, T. and Pregibon, D., 1992: Hastie, T. *et al.*, 2000) and partial mutual information (Sharma, A., 2000). The statistical analysis selected various autoregressive lags and a periodic intercept with 12 values of 3 monthly means, eg. February intercept contains the mean of January to March values of the mixture ratio (or residual ratio) of the entire calibration period. The lag period of autoregressive covariates are listed in Table 4.1. The preference was towards smaller number of predictors to preclude over parameterisation, which is why a limited number of (autoregressive) covariates were considered in the regression formulations.

Table 4.1 Predictor variables for pair wise model combinations. Note that U: UCLA, C:CPC, E:ECMF in the table. Note also that t-3 implies a lag of 3 months, see Equation [8] & [12] for further description of the notations.

(a) Precision ratio (GLAR) model

Pairs	Autoregressive predictor ($r_{t-.}$)	Periodic Intercept (θ_t)
U + E	r_{t-3} , r_{t-6}	$\{r_s ; s = 1,2..12\}$
(U+E)+C	r_{t-3}	$\{r_s ; s = 1,2..12\}$

(b) Bias direction (OLR) model

Pairs	Autoregressive predictor (x_t)	Periodic Intercept ($x_{0,t}$)
U + E	$e_{j,t-3}/e_{i,t-3}$, $e_{i,t-3}$	$\{e_{j,s}/e_{i,s} ; s = 1,2..12\}$
(U+E)+C	$e_{j,t-11}/e_{i,t-11}$, $e_{i,t-6}$	$\{e_{j,s}/e_{i,s} ; s = 1,2..12\}$

4.3.2 Results

The results presented next evaluate the performance of the modelling framework presented in the previous section in the context of forecasting weights as the basis of reducing predictive uncertainty of the NINO3.4 forecasts.

(1) Bias direction forecasts

We consider next the predictive accuracy of the ordered logistic regression of the categorical bias direction as stated in Equation [12]. The expected bias direction at any time step is the category with the highest predicted probability. The validation results for the classification obtained using the model structure illustrated in Figure 4.4 are presented in Table 4.2. As can be inferred from the diagonal values in the table, the correct classification rate is 52% and 44% at level 2 and 1 combinations respectively. The potential of combination error arising from misclassification of *one* to *zero* or vice versa is higher than the misclassification of the category *mix*. Such misclassification rate (*zero* to *one*, *one* to *zero*) was small in this study. The result table shows that there are only 5 out of possible 30 instances in level 2 combination and another 3 out of 20 instances in level 1 misclassified *zero* to *one* or vice a versa.

(2) Mixture ratio forecasts

Dynamic weights are predicted based on precision forecast conditional on bias direction forecast (Equation [11]). In the results that follow, a leave ± 6 months out cross validation where data blocks of 6 months from either side of a validation month are excluded in the formulation of the model, is performed to ascertain the predictive accuracy associated with the categorical forecast. For example, the July 1970 validation is based on a calibration period of January 1956 to December 1969 and February 1971 to December 2001. In addition to this, we also validate the model in four ten year blocks, for example validation from 1992 to 2001 that has been calibrated for the period of 1956 to 1991. The mean of squared error (MSE) of the predictions is used as a measure of forecast accuracy.

Table 4.2 Contingency table (represented as percentages) showing predictions from ordered logistic model. The diagonal indicates the correct classification rate; (a) the level two tree where UCLA and ECMF models are combined; (b) the level one tree where (UCLA+ECMF) is combined with CPC.

	Obs		
Predicted	*one*	*zero*	*mix*
one	25	5	17
zero	0	0	0
mix	17	9	27
Total	42	14	44

(a)

	Obs		
Predicted	*one*	*zero*	*mix*
one	17	3	7
zero	0	0	0
mix	34	12	27
Total	51	15	34

(b)

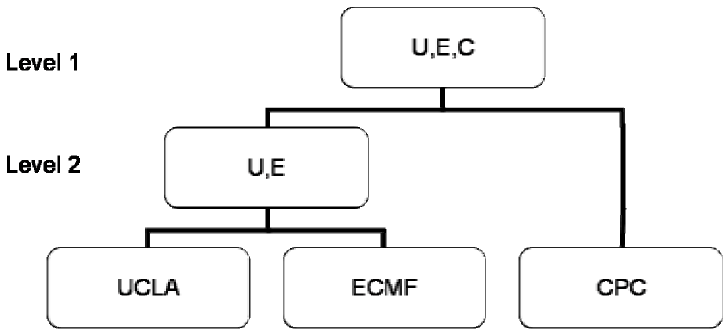


Figure 4.4 Pair wise hierarchical mixing tree of UCLA(U), CPC(C) and ECMF(E). The first pair UCLA + ECMF is chosen due to its lowest covariance.

The static combination forecasts (Equation [2]) is used as the benchmark to evaluate the performance of this dynamic combination method. Let us index {UCLA, ECMF, CPC} as models {1, 2, 3} based on lower to higher paired covariance. The static weight of UCLA and ECMF models are $w_1^{(s)}$ and $w_2^{(s)}$.

These weights can be compared to the overall weight assigned to individual models at Level 1 of the hierarchical tree in Figure 4.4. Using the notation as of Equation [14] for the dynamic weight $\{w_{t,m}{}^{(k)}\}$, where *m* refers to the pair number for the *l*'th tree level, the weight representing the weight for the first model included in the pair (the weight for the other model being (1- $w_l^{(l)}$), the overall weights associated with individual models can be derived as shown in Table 4.3.

The dynamic and static weights, as explained above, from the ± 6 months out cross validation are presented in Figure 4.5 and Figure 4.6 The static weights do not change over time (the wiggly appearance being a result of presenting results in cross validation). As expected the static weights represent the centroid of the observed values of the dynamic weights. A close scrutiny of the predicted weights revealed that the UCLA model dominated during El Nino period, reflecting prime calibration intent of UCLA (Kondrashov et al., 2005). The weights during La Nina phase slightly favoured the CPC model. While overall ECMF contribution was minor, this minor role is more evident during La Nina and the neutral years.

Table 4.3 The notation and comparative association of the dynamic weight (46 years of monthly values) to static weight.

Model	Index	Dynamic Weight	Static Weight
UCLA	1	$\{w_{t,1}{}^{(1)} \bullet w_{t,1}{}^{(2)}\}$, t=1,2,…46x12	$w_1{}^{(s)}$
ECMF	2	$\{w_{t,1}{}^{(1)} \bullet (1- w_{t,1}{}^{(2)})\}$, t=1,2,…46x12	$w_2{}^{(s)}$
CPC	3	$\{(1- w_{t,1}{}^{(1)})\}$, t=1,2,…46x12	$1- w_1{}^{(s)} - w_2{}^{(s)}$

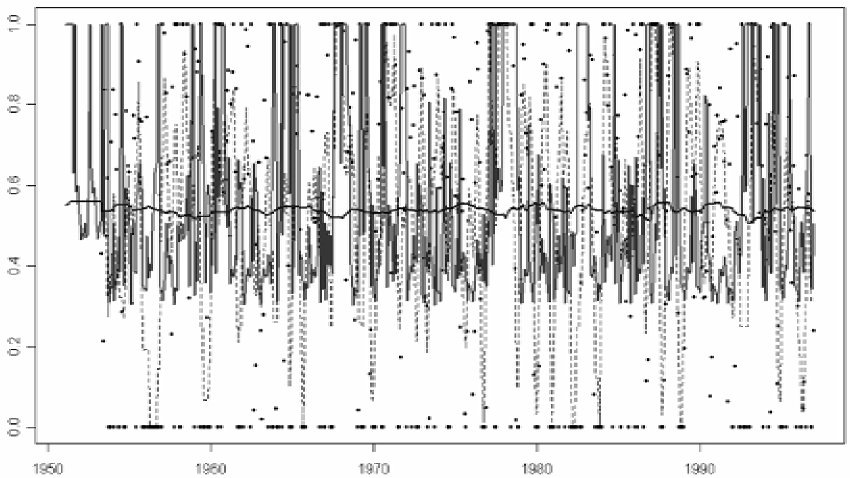


Figure 4.5 The overall dynamic weights for the UCLA model. The predicted weights in validation are drawn in solid lines and the optimum weights as black dots. The broken line is a 4 monthly moving average of the optimum weights included for clarity. The near horizontal line along 0.55 is the static weight.

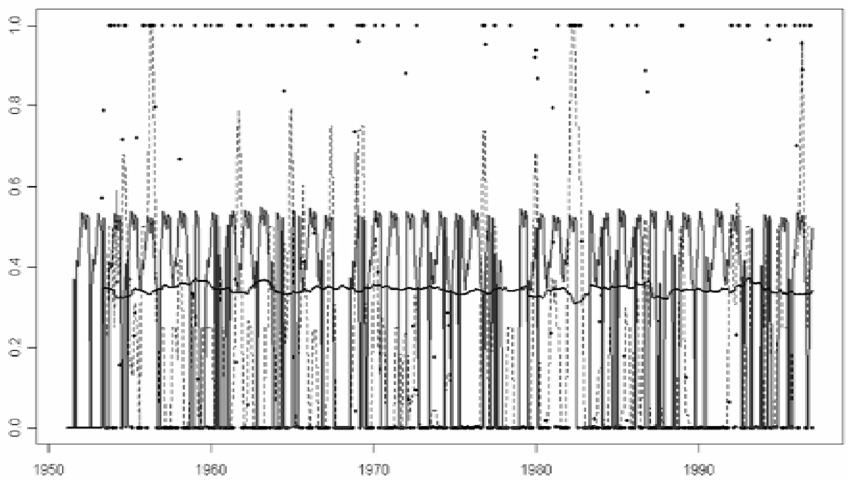


Figure 4.6 The overall dynamic weights for the CPC model. . The predicted weights in validation are drawn in solid lines and the optimum weights as black dots. The broken line is a 4 monthly moving average of the optimum weights included for clarity. The near horizontal line along 0.35 is the static weight.

(3) NINO3.4 prediction from the combined models

This subsection presents improvements achieved in predicting NINO3.4 due to model combination by dynamic weight. We start by offering an assessment of the performance of existing precision ratio based combination methods (Equation [6] and [7]). The performances of these precision ratio based estimates are scrutinized based on last ten years (1991 to 2001) of hindcasts. Various combinations of parameters *γ*, *h* and *λ* of Equation [5] and [6] did not improve the predictive error variance compared to that of (0.192) static weight based combination (Figure 4.7), confirming the need of a more flexible dynamic weight formulation.

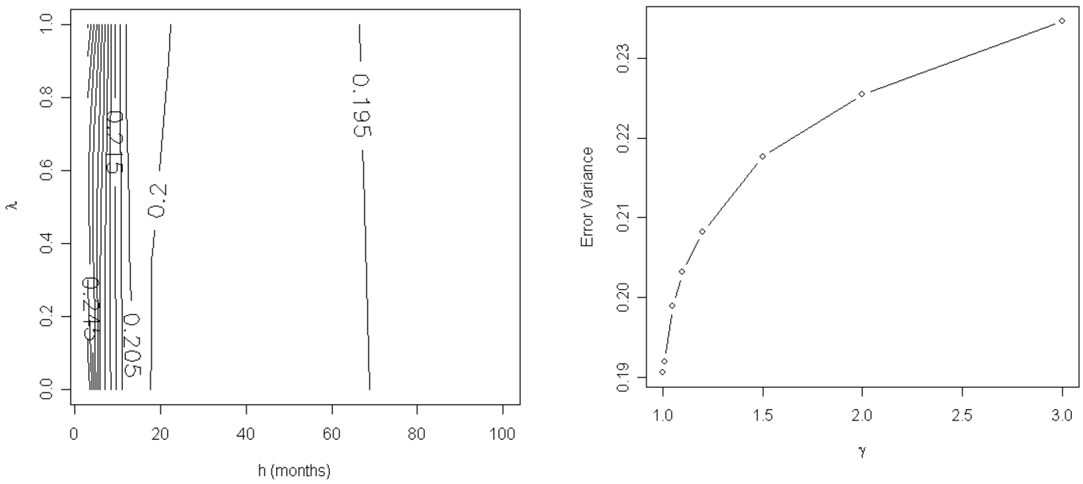


Figure 4.7 The error variance of the combined prediction is drawn against the parameters of Equation [5] and [6]. The predictive variance of static weight combination of period 1992 to 2001 is 0.192, higher *h* value or *γ*= 1 collapses the model to the static weight case. None of the two methods exhibit any improvement (i.e. error variance < 0.192) by localised estimation of the weights (by smaller *h* or *γ*> 1).

Readers should recall that our proposed model consists of two stages, the first stage aiming to ascertain the magnitude of the mixture ratio between the two models, and the second step the direction of the respective errors the two models have. These stages are referred to as the mixture ratio and the bias direction models. Table 4.4 presents the mean squared error (MSE) of calibration and various validation results. Improvements can be noted in all validation cases presented.

Table 4.4 The mean of squared errors (MSE) of NINO3.4 predictions prior to and after the inclusion of the bias direction model. The first row represents the calibration using the full data set (1951 to 2001). The second row represents leave ±6 months out cross validation skills. The other rows represent validation in ten year blocks. Note that all cases exhibit superior performance after consideration of the bias model.

	Precision ratio only	Including bias direction
Calibration	0.1472	0.1401
±6 mth CV	0.1484	0.1444
2001 to 1992	0.1875	0.1853
1991 to 1982	0.1917	0.1845
1981 to 1972	0.1199	0.1071
1971 to 1962	0.1167	0.1159

The MSE of dynamic combination (same as the second column in Table 4.4), the static combination and the component models for the same period for the NINO3.4 response variable are listed in Table 4.5. The overall reduction of MSE (Table 4.5) and improved prediction skill in either combination method reconfirms the past finding that the model combination improves the prediction accuracy (Armstrong, J. S., 1989: Clemen, R. T., 1989). We note that the reduction of MSE of dynamic combination to that of static combination is minor. However the better results are consistence for all the cases. The reduction of MSE, if analysed by an one tail t test, is found to be significant at confidence level of p= 0.0117. A notable point here is that these improvements are based on the use of persistence of various orders only, with no exogenous predictors being considered for simplicity in our presentation. In addition to the presented MSE, the results are analysed using alternative measures like mean absolute error, mean error in probability space and Nash-Sutcliffe Efficiency or R^2. These measures drew similar conclusion and are not presented here.

Table 4.5 The mean of squared errors (MSE) for various model results for a concurrent period for NINO3.4. The first row represents the calibration using the full data set (1951 to 2001). The second row represents leave ±6 months out cross validation skills. The other rows represent validation in ten year blocks.

	DW	Static W	UCLA	CPC	ECMF
Calibration	0.1401	0.1477	0.1821	0.1998	0.5576
±6 mth CV	0.1444	0.1503	0.1821	0.1998	0.5576
2001 to 1992	0.1853	0.1946	0.2499	0.2281	0.8476
1991 to 1982	0.1845	0.1908	0.2259	0.2536	0.6946
1981 to 1972	0.1071	0.1121	0.1483	0.1631	0.4400
1971 to 1962	0.1159	0.1205	0.1322	0.1909	0.325

4.4 Discussion

While the results presented in the previous section do point to the utility of the proposed dynamic combination approach, there are a number of issues that need to be discussed. Foremost amongst these are predictive models used to ascertain the dynamic combination weights.

The inclusion of bias direction model reduces the chance of combined prediction being inferior to individual model. No strong and clear basis of setting the order of the categories of the bias direction variables exists. There are alternatives to the ordered logistic regression used for the bias direction model. They are linear discriminant analysis, quadratic discriminant analysis and multinomial models. However an evaluation of these alternatives did not yield any better prediction in the case study presented here. In addition, we preferred OLR as it needs lesser parameter than the multinomial model. It is worthwhile here to flag the basic assumption of each component model being unbiased while formulating the combination method. However this study did not remove the minor bias from component predictions prior to combination. We formed the view that any apparent fine bias is local to the time window of analysis only; any pre whitening of the predictions in absence of detail calibration knowledge may be precarious.

The class of linear predictive models, in our case the GLAR, does not represent well the probability distribution of the weights because of its limited ability to assume their extreme limits of 0 and 1, and thus underestimates the confidence interval of the combined prediction. The localised regression models (Cleveland, W. S., 1979: Lall, U. *et al.*, 2006) such as Loess (Cleveland, W. S. *et al.*, 1988) and Generalised Additive Models (Hastie, T. and Tibshirani, R., 1986) yielded better predictions in some trials. Our findings concur with Bates and Granger (1969) that the weights (mixture ratio) follow a beta distribution; improvements can be expected by choosing a predictive model that represents this to a better extent such as beta regression model (Ferrari, S. L. P. and Cribari-Neto, F., 2004) ; something that was not attempted to keep our presentation simple and concise.

One way to avoid the need of a design distribution altogether is to use non parametric predictive models (Lall, U. and Sharma, A., 1996: Mehrotra, R. and Sharma, A., 2004: Sharma, A. and Lall, U., 1999). The details of alternative regression options and nonparametric methods are not included here in order to maintain the focus of the paper on the rationale of combining models in a dynamic manner as presented here.

One can infer the weights ω_t as the probability of one component model exhibiting a lower error as compared to its pair in each pair wise combination. The estimation of these weights, or gating function as referred in (Hastie, T. *et al.*, 2000: Marshall, L., Nott, D, Sharma, A, 2007), proceeds through the assumption of a linear form in the logistic transformed space. This is an assumption often used for simplicity and can be improved upon by having other gating functions (instead of the *Logit* transform used here) or nonlinear/nonparametric models instead of the linear one used. A notable departure in this research from existing combination weight formulation is to use absolute residual ratio (Equation [7]) instead of precision ratio (inverse of variance). Precision ratio based dynamic weights (variants of Equation [3], [5] or [6]) requires a minimum band width (time window) to obtain a stable estimate of the variance. Whereas the absolute ratio (which is the analytical solution of the weights when the predictions are bracketing the true value) can be deduced at every time step.

It should also be noted that the notion of formulating dynamic combination weights has been explored in earlier studies, although not in the context of formulating climate model forecasts. Marshall *et al.* (2007) dynamically applied the Bayesian Model Averaging (Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T., 1999) where the method assumes knowledge of exogenous predictors (with imbedded persistence) and full structural details of all component models. In contrast this study regards each component model as a black box, i.e. no structural knowledge is needed, giving way to mix wide types of (statistical or dynamical) predictive models available off the shelves, and considers models on a pair wise basis. Another example of a similar application is found in Robertson, A. W. *et al.*, (2004) where all the component models were paired against the climatological forecasts as a way to stabilise the multivariate weight computation. This resulted in a multivariate extension to the static weight combination presented in Section 4.2.1. Kim *et al.* (2006) applied the artificial neural network method to dynamically combine hydrological models. However the predictor identification (of the weights) was performed in a full multivariate setting and could represent added predictive variance due to the complexity of the neural network models used.

While this paper advocates model combination using the dynamic weight rationale, the robustness of the simple static combination can not be underestimated. The simple method often gives satisfactory result compared to more computationally intensive approaches (Gooijer, J. G. D. and

Hyndman, R. J., 2006, pp 17), which is more true if the component predictions exhibits good precision. The complexity of combination increases with higher number of component models. Increased number of hyper parameters (parameters external to the component models) reduces the degrees of freedom which may eventually compromise the strength of combination. However, the use of cross validation as the basis for evaluating model performance removes this concern to a significant extent, as the cross validation mean square error represent the predictive error the models may have. Any reduction of the number of component models can be carried out during the design of the combination tree by pre combining (using static weight) the model partners showing high residual covariance. Although empirical studies in the economic forecasting literature recommend a maximum number of components as 6 to 8 (Armstrong, J. S., 2001, pp 420: Hibon, M. and Evgeniou, T., 2005), it is unclear if that finding holds for the type of models considered here, and especially so if our aim were to formulate such combinations over a multivariate response representing grided sea surface temperature anomalies spread over the full world surface. Nevertheless the bias direction model here, to some extent, shields the loss of parsimony of the combined prediction when the number of models M is high. We are of the opinion that the optimum size of M largely depends on the extent of the uncertainty and level of independence of the component predictions. This empirical study is based on ensemble mean only; nevertheless the predicted weights can be applied to combine the full set of component realisations in order to attain full probability range. Our future work aims to expand this method, now applied to a univariate response, to a multivariate spatially distributed response vector. In a multivariate extension, we envisage the challenge will be to maintain spatial dependence in the predicted responses with minimal loss of degrees of freedom while consistently exhibiting improvement.

4.5 Conclusion

This paper presented a methodology for combining forecasts from multiple models in a dynamic manner. Multiple models were mixed in pairs based on importance weights that were allowed to vary in time reflecting the persistence of individual model skills and of any relevant exogenous variable. The model pairs were first matched based on the sample error covariance. Then the pairs were combined by ascertaining a weight for each time step. The weights were structured in a hierarchical pair wise combination tree. This process provided a low dimension setting for investigating any predictive structure of the relative model strengths. A two step regression model was used to predict the weights; the steps being the formulation of the mixture ratio model and the bias direction model. The mixture ratio was predicted by a generalised linear autoregressive model and the bias direction by an ordered logistic regression.

The method was applied to combine two statistically based and one dynamic climate models. The variables of interest was the monthly sea surface temperature anomalies averaged over the NINO3.4 region from 1956 to 2001 predicted three months in advance. The combined model skill was tested using a 'leave ± 6 months out cross validation' along with validation in individual 10 year blocks. This empirical study first reconfirmed the concept that the predictions from static weight combination (or a weighted model average) of multiple models improves the skills compared to any single model prediction. Secondly, we found that the predictions using existing precision ratio based dynamic weight did not offer any improvement over predictions using static weight combination. Thirdly, the proposed dynamic weight computation method is an improvement over the existing precision ratio based dynamic weights. The proposed method exhibited a very small but consistent increase in prediction skill over that of static weight method for the entire six validation scenario with no case of worsening results. These consistent results suggest that the potential of improvement is real if multiple predictions are combined using our proposed dynamic weights.

4.6 Acknowledgement

We acknowledge the assistance of Upmanu Lall, Andrew Robertson and Lisa Goddard of the Lamont Doherty Earth Observatory Centre of Columbia University, New York and for access to their component model prediction database. This research was funded by the Australian Research Council and the Sydney Catchment Authority. The computation was performed using the freely available R statistical computing platform (R Development Core Team, 2006). The helpful comments of anonymous Journal of Climate reviewers are gratefully acknowledged.

4.7 Reference

Agresti, A., 1996: *An Introduction to Categorical Data Analysis*. *Wiley Series in Probability and Statistics*.

Aksu, C., Gunter, S., 1992: An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. *International Journal of Forecasting*, **8,** 27-43.

Armstrong, J. S., 1989: Combining forecasts: The end of the beginning or the beginning of the end? *International Journal of Forecasting*, **5,** 585-588.

——, 2001: Combining forecasts. *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong, Ed., Kluwer Academic Publishers, 417-439.

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,** 1783-1796.

Bates, J. M. and C. W. J. Granger, 1969: The combination of forecasts. *Operational Research Quarterly*, **20,** 451-468.

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: Global Ocean Surface Temperature Atlas.

Carrasco, J. A. and J. D. Ortuzar, 2002: Review and assessment of the nested logit model. *Transport Reviews*, **22,** 197-218.

Chambers, J. M., 1992: Linear Models. *Statistical Models in S*, J. M. Chambers and T. Hastie, Eds., Wadsworth & Brooks.

Chandler, R. E., 2005: On the use of generalized linear models for interpreting climate variability. *Environmetrics*, **16,** 699-715.

Clemen, R. T., 1989: Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, **5,** 559-583.

Cleveland, W. S., 1979: Robust locally weighted regression and smoothing scatterplots. *Journal of American Statistical Association*, **74,** 829-836.

Cleveland, W. S., S. J. Devlin, and E. Grosse, 1988: Regression by Local Fiting Methods, Properties, and Computational Algorithms. *Journal of Econometrics*, **37,** 87-114.

Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson, 2004: Forecast Calibration and Combination: A Simple Bayesian Approach for ENSO. *Journal of Climate*, **17,** 1504-1516.

Colman, A. W. and M. K. Davey, 2003: Statistical prediction of global sea-surface temperature anomalies. *International Journal of Climatology*, **23,** 1677-1697.

Dawes, R., R. Fildes, M. Lawrence, and K. Ord, 1994: The past and the future of forecasting research. *International Journal of Forecasting*, **10,** 151-159.

de Menezes, L. M., D. W. Bunn, and J. W. Taylor, 2000: Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, **120,** 190-204.

Deutsch, M., C. W. J. Granger, and T. Terasvirta, 1994: The combination of forecasts using changing weights. *International Journal of Forecasting*, **10,** 47-57.

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

Dunteman, G. H. and M. R. Ho, 2006: *An Introduction to Generalized Linear Models*. Vol. 145, *Quantitative Applications in the Social Sciences*, Sage Publications.

Ferrari, S. L. P. and F. Cribari-Neto, 2004: Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, **31,** 799-815.

Fraedrich, K. and N. Smith, 1989: Combining Predictive Scheme in Long-Range Forecasting. *Journal of Climate*, **2,** 291-294.

Fritsch, J. M., J. Hilliker, and J. Ross, 2000: Model Consensus. *Wether Forecasting*, **15,** 571-582.

Gelman, A. and J. Hill, 2006: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 1st edition ed. Cambridge University Press.

Gooijer, J. G. D. and R. J. Hyndman, 2006: 25 years of time series forecasting. *International Journal of Forecasting*, **22(3)**, 443-473.

Granger, C. W. J. and P. Newbold, 1977: *Forecasting economic time series*. Academic Press.

Greene, A. M., L. Goddard, and U. Lall, 2005: Probabilistic multimodel climate change scenarios for the evolution of regional temperatures.

Hastie, T. and R. Tibshirani, 1986: Generalized Additive Models. *Statistical Science*, **1,** 297-310.

Hastie, T. and D. Pregibon, 1992: Generalised Linear Models. *Statistical Models in S*, J. M. Chambers and T. Hastie, Eds., Wadsworth & Brooks.

Hastie, T., R. Tibshirani, and J. Friedman, 2000: *The Elements of Statistical Learning, Data Mining, Inference and Prediction*. Springer Series in Statistics.

Helsel, D. R. and R. M. Hirsch, 2002: Statistical Methods in Water Resources. *Book 4, Hydrologic Analysis and Interpretation*, United States Geological Survey.

Hibon, M. and T. Evgeniou, 2005: To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, **21,** 15-24.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T., 1999: Bayesian Model Averaging: A Tutorial (with discussion). *Statistical Science*, **14,** 382-417.

Kamstra, M. and P. Kennedy, 1998: Combining qualitative forecasts using logit. *International Journal of Forecasting*, **14,** 83-93.

Kim, Y. O., D. Jeong, and I. H. Ko, 2006: Combining rainfall-runoff model outputs for improving ensemble streamflow prediction. *Journal of Hydrologic Engineering*, **11,** 578-588.

Kondrashov, D., S. Kravtsov, A. W. Robertson, and M. Ghil, 2005: A Hierarchy of Data-Based ENSO Models. *Journal of Climate*, **18,** 4425-4444.

Lall, U. and A. Sharma, 1996: A nearest neighbor bootstrap for time series resampling. *Water Resources Research*, **32,** 679-693.

Lall, U., Y.-I. Moon, H.-H. Kwon, and K. Bosworth, 2006: Locally weighted polynomial regression: Parameter choice and application to forecasts of the Great Salt Lake. *Water Resources Research*, **42**.

Larrick, R. P. and J. B. Soll, 2006: Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, **52,** 111-127.

Lundberg, S., T. Tera¨svirta, and D. van Dijk, 2000: Time-varying smooth transition autoregressive models.376.

Marshall, L., 2006: Bayesian analysis of rainfall-runoff models: Insights to parameter estimation, model comparison and hierarchical model development, School of Civil and Environmental Engineering, University of New South Wales, 222.

Marshall, L., Nott, D, Sharma, A, 2007: Towards Dynamic Catchment Modelling: A Bayesian Hierarchical Modelling Framework. *Hydrological Processes*, **21,** 847-861.

McCullagh, P. and J. A. Nelder, 1989: *Generalized Linear Models*. 2nd ed. ed. Chapman & Hall.

McLeod, A. I., D. J. Noakes, K. W. Hipel, and R. M. Thompstone, 1987: Combining hydrologic forecast. *Journal of Water Resources Planning Management*, **113,** 29-41.

Mehrotra, R. and A. Sharma, 2004: Conditional resampling of hydrologic time series using multiple predictor variables: A k-nearest neighbour approach. *Water Resources Research*.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Delecluse, Deque, M. , 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.

Pavan, V. and F. J. Doblas-Reyes, 2000: Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. *Climate Dynamics*, **16,** 611-625.

Peng, P., A. Kumar, and H. v. d. Dool, 2002: An analysis of multimodel ensemble predictions for seasonal climate anomalies. *Journal of Geophysical Research*, **107,** 4710.

Phillips-Wren, G. E., E. D. Hahn, and G. A. Forgionne, 2004: A multiple-criteria framework for evaluation of decision support systems. *Omega (Oxford)*, **32,** 323.

R Development Core Team, 2006. R Foundation for Statistical Computing.

Raftery, A. E., T. Geneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review*, **133,** 1155-1174.

Ragonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona, 2006: A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, **42,** W09404.

Rajagopalan, B., U. Lall, and S. E. Zebiak, 2002: Categorical Climate Forecasts through Regularization and Optimal Combination of Multiple GCM Ensembles{ast}. *Monthly Weather Review*, **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. *Monthly Weather Review*, **132,** 2732-2744.

Sanders, F., 1963: On subjective probability forecasting. *Journal of Applied Meteorology*, **2,** 191-201.

See, L. and R. J. Abrahart, 2001: Multi-model data fusion for hydrological forecasting. *Computers & Geosciences*, **27,** 987-994.

Shamseldin, A. Y., O'Connor, K.M., Liang, G.C., 1997: Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology*, **197,** 203-229.

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

Sharma, A. and U. Lall, 1999: A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation*, **48,** 361-371.

——, 2004: Model averaging and its use in probabilistic forecasting of hydrologic variables. *Hydrology, Science and practice for the 21st century*, Imperial College, London, British Hydrological Society, 372-378.

Shephard, N., 1995: Generalized Linear Autoregressions. *Economics Working Paper 8*, February 6 ed., Nuffield College.

Smith, T. M. and R. W. Reynolds, 2002: Extended reconstruction of global seas surface temperatures based on COADS data 1854 to 1997. *Journal of Climate*, **16,** 1495-1510.

Terui, N. and H. K. van Dijk, 2002: Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, **18,** 421-438.

Thompson, P. D., 1977: How to Improve Accuracy by Combining Independent Forecast. *Monthly Weather Review*, **105,** 228-229.

van den Dool, H., 2000: Constructed analogue prediction of the east central tropical Pacific SST and the entire world ocean for 2001. *Experimental Long-Lead Forecast Bulletin*, **9,** 38-41.

Xiong, L., A. Y. Shamseldin, and K. M. O'Connor, 2001: A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. *Journal of Hydrology*, **245,** 196-217.

Yang, C., R. E. Chandler, and V. S. Isham, 2005: Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research*, **41,** W11415.

Yu, L., S. Wang, and K. K. Lai, 2005: A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Computers & Operations Research*, **32,** 2523-2541.

Zou, H. and Y. Yang, 2004: Combining time series models for forecasting. *International Journal of Forecasting*, **20,** 69-84.

CHAPTER FIVE

This chapter reprinted the following article:

> Chowdhury, S. and A. Sharma, 2009: Improving prediction skill by a dynamic combination of multiple global sea surface temperature forecasts. *Journal of Climate*, under review.

5 GLOBAL SEA SURFACE TEMPERATURE FORECAST

ABSTRACT

Current approaches for combining hydroclimatic forecasts to reduce predictive uncertainty are mainly limited to using combination weights that remain static over time. Recent research has shown the advantage of time variant mixing weights (dynamic weights) over static weights in long range prediction of El Nino Southern Oscillation Indices. This paper extends the dynamic combination algorithm to predict a multivariate sea surface temperature field. Two approaches for forecasting multivariate dynamic weights are presented in this paper. The first of the two approaches uses a mixture of two basis distributions which are three category ordered logistic regression model and a generalised linear autoregressive model. The second method uses a modified nearest neighbour approach to forecast the future weights.

The case study presented combines predictions from three climate models for the period 1958 to 2001. The variables of interest here are the monthly global sea surface temperature anomalies at a 5°x5° latitu de-longitude grid, predicted three months in advance. The prediction from static weight combination is used as the base case for comparison. The predicted sea surface temperature using the dynamic combination algorithm consistently exhibits better accuracy to that of static combination in each season. Improved skill is achieved at 93% of the global grid cells compared to the static weights, in four, 10 year independent validation segments. Our results also consistently outperform the best performing single model amongst the three models considered.

5.1 Introduction

There are various dynamic and stochastic models for hydroclimatic prediction, each model being subject to their own strengths and weaknesses. The selection of a single *best* model introduces selection uncertainty as well as discards any superior strengths of the models not selected over the full historical record but a shorter period of time (Hibon, M. and Evgeniou, T., 2005). This paper investigates combining multivariate responses from multiple models as an alternative to selecting a single model. Linear combinations of multiple model responses are used to reduce the predictive uncertainty of climate variables (Barnston, A. G. *et al.*, 2003: Pena, M. and van den Dool, H., 2008: Peng, P. *et al.*, 2002: Raftery, A. E. *et al.*, 2005) . While climate scientists and hydrologists have been aware of the advantage of model combination (Fraedrich, K. and Smith, N., 1989: Sanders, F., 1963: Thompson, P. D., 1977), research on improving the linear combination methods has been limited. One scope of improvement is to allow the linear model combination weights to be non stationary, by allowing weights to change as a function of time (Bates, J. M. and Granger, C. W. J., 1969: Miller, C. M. *et al.*, 1992). Such an approach is similar to the rationale behind dynamic linear models (Huerta, G. and Sansó, B., 2007: Lundberg, S. *et al.*, 2000: West, M. and Harrison, J., 1997), where the parameters of the model (analogous to the model combination weights here) are allowed to vary with time.

The rationale for model selection weights that vary over time has been investigated in hydroclimatology by McLeod *et al.* (1987) and more recently by Chowdhury and Sharma (2009) (hereafter referred to as CS2009). Such weights were termed dynamic model combination weights (or dynamic weights in short), and the improvement resulting from their use for 3-month-ahead forecasts of the NINO3.4 index in contrast to temporally invariant weights (or static weights) were documented in CS2009. While the CS2009 study was limited to univariate application, this paper extends the method for multivariate prediction of global sea surface temperature anomalies. The existing developments that set the background of our current research are discussed in next paragraph.

Extensive review and a bibliography on time series combination research compiled by Clemen (1989) corroborate the earlier studies on this topic. Combinations of multiple models have been widely adopted in practice in econometrics (Armstrong, J. S., 2001: Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T., 1999). Research on non stationary model combination weights (mostly sourced from disciplines outside of hydroclimatology) has investigated various non continuous and non linear methods for combining constituent or component models (Deutsch, M. *et al.*, 1994: Terui, N. and van Dijk, H. K., 2002: Yu, L. *et al.*, 2005: Zou, H. and Yang, Y., 2004). In contrast, hydroclimatic research has commonly assumed a stationary skill of the component models (Butts, M. B. *et al.*, 2004: Georgakakos, K. P. *et al.*, 2004: Ragonda, S. K. *et al.*, 2006: Shamseldin, A. Y., O'Connor, K.M., Liang, G.C., 1997: Sharma, A. and Lall, U., 2004), with few exceptions that explore a dynamic combination (Devineni, N. *et al.*, 2008: Marshall, L. *et al.*, 2007: See, L. and Abrahart, R. J., 2001: Xiong, L. *et al.*, 2001). Most of the above mentioned studies focus on models that produce single or univariate outputs. A dynamic model combination where multiple or multivariate outputs are of interest requires consideration of the multivariate or spatial dependence that needs to be simulated in the resulting response.

This paper is organised as follows. We first describe the multivariate dynamic weight formulation and two techniques for predicting the dynamic weights forward in time. We next present the application of the proposed methodology to dynamically combine global sea surface predictions from three predictive models. Finally we securitized the results with relevant discussion and conclusion.

5.2 Methodology

The alternative forecasts issued by various models are combined using parameters referred to as weights in this paper. The persistence of individual model skill is reflected by using weights that vary over time (are dynamic instead of static). The time variant property of the weight raises additional requirement of forecasting the dynamic weight time series. Before presenting the rationale for forecasting dynamic model combination weights, we first assess the rationale for estimating static weights or combination weights that do not vary with time. Static model combination weight, in its various forms, embodies a major portion of the current state of practice; hence it is used as a benchmark of our proposed method. Later we present the dynamic weight formulation and two methods for forecasting the dynamic weights forward in time.

5.2.1 Static weight

Consider multivariate predictions of a component model m for a time step t at any location l as $\{\hat{u}_{m,l,t} : m=1,2,.. ; l=1,2,.. ; t=1,2.. \}$. Then the combined prediction of all component models $\hat{y}_{l,t}^{(s)}$ can be ascertained as:

$$\hat{y}_{l,t}^{(s)} = \sum_m \hat{u}_{m,l,t}\, w_{m,l} \qquad [1]$$

Where, $w_{m,l}$; m=1,2,.. are the static weights at location l, conditional to $w_m \geq 0$ and $\sum_m w_m = 1$.

The Equation [1] above denotes weighted average of multiple alternative forecasts. Note that the weights are equal in case of simple multi-model mean. The estimation of $\{w_{m,l}\}$ here largely follows a methodology similar to Robertson *et al.* (2004). The methodology can be divided into three steps. First, the weights at all grid points are derived by combining only two predictions rather than all of the models combined together. For example, each component model prediction at each grid point is paired against a common reference prediction. The reference prediction may be the climatology values (i.e. mean prediction over a long period). The static weight of each pair is estimated by minimising a loss function, chosen as the sum of squared errors in this study. Next, the resulting weights (paired against climatology) of all component models are normalised to add to unity. The above two steps are repeated for all grid points forming the raw weights. At third step, the spatial variations of these raw weights are reduced by a smoothing function centred at each grid point and spreading across its neighbours.

This method of estimating static combination weight ensures stability in the estimated values; an important consideration for combining multivariate forecast variables from a number of alternative models. Such an approach is attractive, but the fact remains that it is incapable of dealing with cases where individual model skills may be non stationary, in which case the combination weights may need to change with time as proposed in combination approach outlined in next section.

Note that this paper follows notations of lower case *italics* for scalar values (e.g. $\hat{u}_{m,l,t}$) and straight Romans for vectors (e.g. $\hat{\mathrm{u}}_{m,t} \equiv \{\hat{u}_{m,l,t}\ ;\ l=1,2,..\}$) and **bold** fonts (e.g. $\hat{\mathbf{u}}_t \equiv \{\hat{u}_{m,l,t}\ ;\ m=1,2,..;\ l=1,2,\ldots\}$ or $\hat{\mathbf{U}} \equiv \{\hat{u}_{m,l,t}\ ;\ m=1,2,..;\ l=1,2,\ldots;\ t=1,2,..\}$) to represent higher dimension matrices if not mentioned otherwise. Vector series is shown within curly brackets {.}. Functions are denoted by capital italic names followed by brackets, such as minimum being specified as *Min*(.). Any departure from this notation (eg *K* in nearest neighbour method) is mentioned exclusively.

5.2.2 Dynamic weight

Consider the case of combining predictions of two component models, $\hat{u}_{1,l,t}$ and $\hat{u}_{2,l,t}$, at a location l. For ease of notation let us conceal the location subscript l now within the straight font notation as $\hat{\mathrm{u}}_{1,t} \equiv \{\hat{u}_{1,l,t}\ ;\ l=1,2,3..\}$:

$$\mathrm{y}_t = \hat{\mathrm{u}}_{1,t}\,\omega_t + \hat{\mathrm{u}}_{2,t}(1 - \omega_t) + \ddot{\mathrm{e}}_t \qquad [2]$$

Where, y_t: observed predictand at time t

ω_t: weight dynamically assigned to model 1, and

$\ddot{\mathrm{e}}_t$: the residual of the combined prediction.

The algebraic solution of the weights in Equation [2] that makes $\ddot{\mathrm{e}}_t = 0$, ρ_t equals

$$\rho_t = \mathrm{e}_{2,t} / (\mathrm{e}_{2,t} - \mathrm{e}_{1,t}) \qquad [3]$$

where $\mathrm{e}_{1,t}$ and $\mathrm{e}_{2,t}$ are residuals of the model 1 and 2.

Analytically, $\boldsymbol{\rho}$ may span from $-\infty$ to $+\infty$. We assume that the component models are unbiased and hence weights are constraint to be $\{\omega_t \in 0 \rightarrow 1\}$. So the observed dynamic weight is written as:

$$\omega_t = \begin{cases} \rho_t & 0 < \rho_t < 1 \\ 1 & 1 < \rho_t \\ 0 & 0 > \rho_t \end{cases} \qquad [4]$$

Consequently, given the above derivation of ω_t , Equation [2] can be seen as dynamic extension of Equation [1] that is weighted average of two alternative forecasts at every time step.

The combination problem can now be considered as a time-series forecasting problem where the response is the dynamic weight ω_t, conditional to predictors which could be chosen from past lags of $\boldsymbol{\omega}$ and selected exogenous variables. Additionally, the prediction problem needs to account for the significant mass of observations that fall precisely at 0 or 1, resulting in the prediction comprising of two prediction problems – (1) whether ω_t is 0, 1 or in-between, and (2) if in-between, what is the numeric value predicted.

This paper presents two methods for forecasting the observed weight time series $\{\omega_t;\ t=1,2..,t_{\max}\}$ forward in time. The first method uses a mixture of two generalised linear regression approaches (GLM) and the second method uses a modified nearest neighbour sampling process (KNNW) for predicting the dynamic combination weights.

(1) Mixture regression method

The first method for predicting the combination weight $_t$ uses a generalised linear regression rationale. Generalised linear regression requires the response variable to belong to a family of exponential distributions, in contrary the weights $\{\omega_t\}$ follow a beta distribution (Bates, J. M. and Granger, C. W. J., 1969) with high inflation at zero and one. Hence the linear regression is formulated by first defining $\{\omega_t\}$ as an aggregation of two intermediate variables, each of which is predicted separately. This approach of simplifying the response distribution inflated at a point or several points as a mixture of two basis distributions is synonymous to modelling rainfall as a mixture of amount and occurrences (Bruhn, J. A. *et al.*, 1980: Harrold, T. I. *et al.*, 2003: Srikanthan, R. and McMahon, T. A., 2001: Yang, C. *et al.*, 2005) or other applications involving a mixture of a discrete and continuous distributions e.g. the Zero Inflated Poisson Model (Hall, D. B., 2000: Martin, T. *et al.*, 2005). These two basis distributions are: (1) mixture ratio $\{r_t\}$ and the (2) categorical variable bias direction $\{b_t\}$, as specified in Equations [5] and [6] below.

$$r_t = |e_{2,t}| / (|e_{2,t}| + |e_{1,t}|) \qquad [5]$$

$$b_t = \begin{cases} mix & \text{when } 0< \rho_t <1 \\ zero & 0> \rho_t \\ one & 1< \rho_t. \end{cases} \qquad [6]$$

It should be noted that the dynamic weight ω_t equals the mixing ratio r_t when the bias direction b_t equals "*mix*", and equals 0 or 1 when that is not the case In the above formulation, the bias direction indicates whether both the forecasts are under or over predicting, or alternately the case where the bias is in opposing directions. The approach of using bias direction and mixture ratio as the basis for predicting the dynamic weight is similar to the forecasting method presented for the case of a single predictand in CS2009.

The first stage of the forecasting weights models the mixture ratio $\{r_t\}$ using a generalised linear autoregressive (GLAR) structure (Shephard, N., 1995). The second stage models the bias direction using a three category ordered logistic regression (OLR) (Agresti, A., 1996). For sake of space the details of GLAR and OLR are not included here, the fundamentals of these regression methods are available at the reference quoted above, and their application is illustrated with sufficient detail in CS2009. The brief algorithm for this method is repeated from CS2009 below:

i) Compute the time series of mixture ratio and bias direction, as defined in Equations [5] and [6], from the residuals of the hindcast series.

ii) Formulate an OLR model to forecast the three bias categories {*mix*, *zero*, *one*}. Account for any seasonality in the bias category time series by formulating the prediction model on a seasonal basis, or introducing seasonally varying coefficients in the model. More specific examples on predictor selection methods are provided in the case study presented at later sections.

iii) Formulate a GLAR model to forecast the mixture ratio $\{r_t\}$ using an autoregressive structure and may be additional exogenous predictors. A logistic transformation of $\{r_t\}$ is needed, after which the GLAR parameters are ascertained. If necessary, account for seasonality using a periodic intercept term.

iv) If the GLAR forecast of mixture ratio at a future time t is $\check{r}_t$, the dynamic weight at t is $\check{r}_t$, 0 or 1 if OLR forecast result is *mix*, *zero* or *one*.

As evident in this section, the regression based forecasting methods are constraint by the inflexible assumption of design distributions. The onerous requirement of conformity to a design distribution (here exponential distribution) can be avoided using a non parametric approach. This paper introduces such a non parametric method in the next section.

(2) Nearest neighbour method

The second method of forecasting model combination weights uses a non parametric weighted nearest neighbour approach known as KNNW (Mehrotra, R. and Sharma, A., 2006). The KNNW approach aims to ascertain the conditional dependence of predictands (ω_t) on a specified set of predictors (here mainly $r_{t-.}$) by identifying nearest neighbours of the predictors in the historical record. A forecast is then expressed as an expected value of the conditional probability distribution formed based on the nearest neighbours.

Identification of the nearest neighbours proceeds by ranking historical responses using a modified squared Euclidean distance (ξ) metric:

$$\xi_{\tau}^{(t)} = \sum_{p} \beta_{p} (x_{p,\tau} - x_{p,t})^{2} \qquad [7]$$

where, $x_{p,\tau}$ = the scaled p^{th} predictor vector at a past time τ,

p = 1, 2, … index of multiple predictor vectors,

$\tau = t$-1, t-2, … index of past time,

β_p=the influence load to p^{th} predictor vector.

The time series of historical weights $\{\omega_\tau;\ \tau = t\text{-}1, t\text{-}2, t\text{-}3 \ldots\}$ is ranked based on the order of the current $\xi_{\tau}^{(t)}$. If k_τ is the sorted rank of ω_τ then $k_\tau \in \{1, 2, 3, \ldots K, \infty\}$, where K is the farthest neighbour considered for ascertaining the prediction in the KNNW method. The probability of re-sampling a past observation at a future time t is then specified as follows (Lall, U. and Sharma, A., 1996):

$$Pr\ (\omega_t = \omega_\tau \mid \mathbf{X}) = k_\tau^{-1} / (1 + 2^{-1} + 3^{-1} \ldots + K^{-1}) \qquad [8]$$

Where **X** is the multiple predictor vector $\{x_{p,\tau}\}$.

The nearest neighbour method is implemented as follows.

i) Compute the time series of observed dynamic weights $\{\omega_t\}$, using Equations [3] and [4].

ii) Identify the predictor set **X** to forecast $\{\omega_t\}$ from a pool of candidate predictors. If the dynamic weights exhibit a marked seasonal structure, it may be necessary to include a seasonal indicator in the pool of candidate predictors. Examples of possible candidate predictors and method of selection are shown in the case study.

iii) Formulate simple linear regression of $\omega_t \sim \mathbf{X}$. Use the normalised absolute value of the multiple regression parameters to determine the influence load $\{\beta_p,\ p=1, 2,..\}$. This technique of assigning influence loads (also known as predictor weights for the KNN method) is as described in Souza Filho and Lall (2003), and represents a simplification of the optimisation based alternative presented in Mehrotra and Sharma (2006).

iv) Specify an appropriate K, recommended to be the nearest integer to the square root of the training time series length (Lall, U. and Sharma, A., 1996). Rank past observed weights $\{\omega_T\}$ based on the modified Euclidian distance (ξ) of current predictor to all predictors in the past as shown in Equation [7].

v) The expected value at a future time t, as computed by the conditional probability distribution defined in Equation [8], is the weight at future time t.

(3) Multivariate response

We discuss here the steps adopted to represent spatial dependence in the forecasts of the multivariate sea surface temperature field. Examples of multivariate responses include global grided sea surface temperatures (Robertson, A. W. *et al.*, 2004), wind speed (Yan, Z. *et al.*, 2002) or multi site rainfall (Yang, C. *et al.*, 2005). Combination of multivariate responses requires a matrix of spatially distributed weight time series. It is reasonable to expect that each component model of multivariate predictions would preserve spatial statistics to a varying degree. Our combination weights are restricted to $\omega_t \in \{0 \rightarrow 1\}$ and thus help to conserve the characteristics of the component model's spatial dependence, allowing simple multivariate extension of the dynamic combination method. The inter dependence of the spatially distributed weights are simulated through the dependence characteristics of the predictor matrix. This is done by smoothing the raw predictors' across the neighbourhood as shown in Equation [9]. The readers are reminded that the location subscript l in Equation [2] to [8] is concealed in the straight font, since those equations are applicable to all grid points $\mathrm{x}_{p,t} = \{x_{p,l,t}\ ;\ l=1,2,..\}$. If the smooth p^{th} predictor at a location l at any time t is $x_{p,l,t}$ then,

$$x_{p,l,t} = \sum_n c_{p,n} \bullet x^*_{p,n,t} \qquad [9]$$

Where,

$n = l \pm 1, l \pm 2, \ldots$ location index of the neighbours relative to this location l,

$x^*_{p,n,t}$ = the raw scaled p^{th} predictor at a neighbouring location n at any time t, and

$c_{p,n}$ = a measure of influence (eg. correlation coefficient) of the predictor at any neighbour n to this location l (note that $c_{p,n=l} = 1$ if the measure is correlation coefficient).

In summary, the dynamic weight combination method begins with computing the observed weights, see Equation [3] and [4], from the hind cast series of the two component models. Then the raw predictors of the observed weights at each grid point $\{x^*_{p,n,t}\}$ are selected from a pool of candidate predictors. Spatial variability of the selected raw predictors is condensed later, see Equation [9]. Once the predictors are ascertained the weights are forecasted applying either the mixture GLM or the non parametric method KNNW.

5.3 Application

The method is applied to improve the 3 months ahead prediction of global sea surface temperature anomalies (SSTA) at 5° by 5° gr ids of the global sea surface between 60°N to 40°S. The base of the anoma lies was the climatology of Global Ocean Surface Temperature Atlas (GOSTA) from 1951 to 1980 (Bottomley, M. *et al.*, 1990: Reynolds, R. W. and Smith, T. M., 1995). The observed data set came from the extended SSTA set, reconstructed at the US National Climate Data Centre (Smith, T. M. and Reynolds, R. W., 2002). Three model predictions were combined. The first and second models were prepared by the DEMETER project (ECMWF, 2004) of European Centre for Medium Range Weather Forecast (ECMWF). One of these two models was developed at ECMWF is referred as ECM here (Wolff, J. E. *et al.*, 1997), the other model comes from Météo-France (Madec, G. *et al.*, 1997) which is referred as MetF in short. The DEMETER models (ECM, MetF) are global coupled ocean-atmosphere models. The third model was developed at the

Climate Prediction Centre of the National Oceanic and Atmospheric Administration, USA and referred to as the CPC model (van den Dool, H., 2000: van den Dool, H. *et al.*, 2003). The CPC model used a statistical technique known as constructed analogue to forecast SSTA as linear combination of all past observation at the same month. All SSTA time series were downloaded from the data library of International Research Institute for Climate and Society, New York. The common period of hind-casts among these three models extends for a period of March 1958 to December 2001. Note that this study accepts the component model as black box and ignores any minor biases. The removal of apparent bias by looking only through a certain time window might affect any unique strength of the model at an alternative time window.

As mentioned at the beginning of the Section 5.2, we decided to compare the performance of dynamically combined prediction to the predictions obtained by static combination. Each of the three models (ECM, MetF, CPC) were first combined against the GOSTA climatology in isolation (Robertson, A. W. *et al.*, 2004). The three sets of weights were normalized at the next step. The spatial noises of the normalized raw weights were condensed by taking an average value of weights of all surrounding grid cell within ±20° distance. These smooth static weights were used to combine the three component models and form the static weight alternative $\hat{y}_{l,t}^{(s)}$ of our proposed method.

The readers may note that the dynamic weight formulation and forecasting methods developed in Section 5.2 are based on two component models only. Application of this method to higher number of component models requires a hierarchical pair-wise combination tree. The design of the tree may be influenced by a number of considerations, such as residual covariance and existence of static combination nodes. The formation of pair-wise combination tree of the three component model is illustrated in Figure 5.1. The two global circulation models, ECM and MetF were first combined using static weight as described in Section 5.2.1. The joined ECM+MetF model were then combined with CPC using the dynamic weight logic. As a first step, the dynamic combination method computed the observed weights of (ECM+MetF) and

CPC pair using Equation [3] and [4]. Contours of the mean of the observed weight time series are shown in Figure 5.2. The background shades in Figure 5.2 llustrates the cells with high concentration of 0 or 1 in the observed weight time series. The next step requires selection of predictors to forecast these observed weight time series.

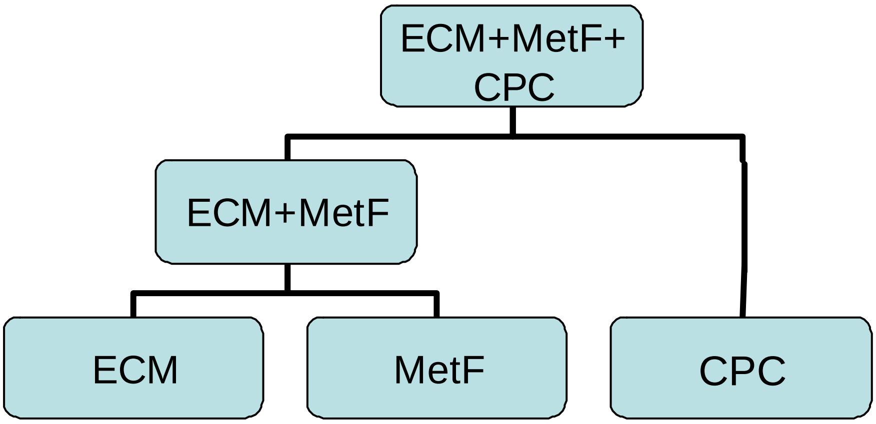


Figure 5.1 The combination tree of the three component models. The component models are, MetF: Météo-France, ECM: European Centre for Medium Range Weather Forecast, CPC: Climate Prediction Centre of the National Oceanic and Atmospheric Administration, USA.

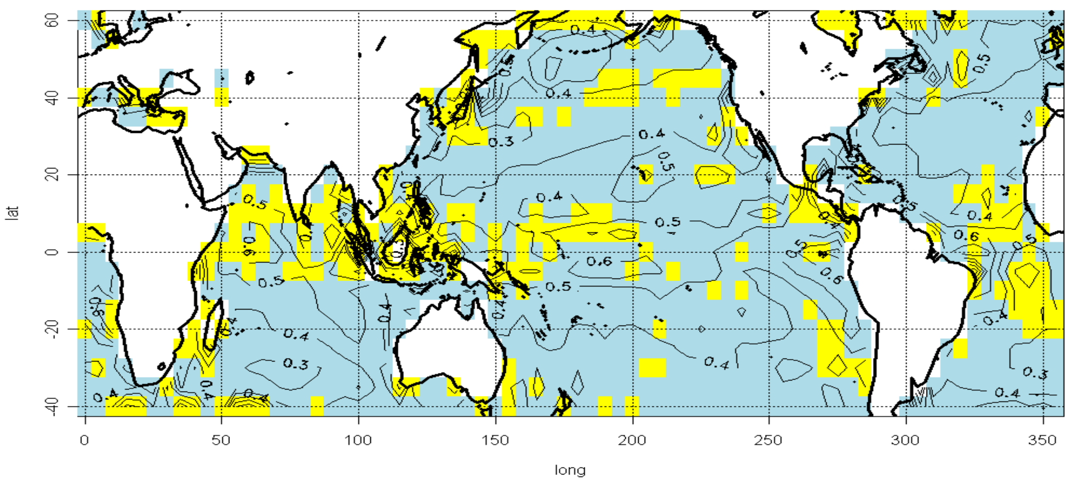


Figure 5.2 Mean observed weights from 1958 to 2001 for the dynamic combination of the (ECM+MetF) and CPC model combination. The yellow zones are locations where the majority of the three category observed weights {0, 0<$\boldsymbol{\omega}$<1, 1} are either 0 or 1.

5.3.1 Predictor selection

The predictors for both the mixture GLM and the KNNW forecasting methods are ascertained from lagged values of the mixture ratio (r_t) and an indicator of the relative bias associated with both models, referred to as the residual ratio ($ŕ_t$) as shown in Equation [10].

$$ŕ_t = e_{2,t}/e_{1,t} \qquad [10]$$

The residual ratio $ŕ_t$ is constrained to fall within {-1 → 2} avoiding numerical instability when $e_{1,t} \rightarrow 0$. A common set of predictors is used for the entire global sea surface for simplicity. This is achieved by first identifying the geographical spread of the concentration of high loadings of the first few principle components of observed **ω** or $\{\omega_{l,t};\ l=1,2..,\ t=1,2..\}$. We narrowed our predictor search by primarily aiming to forecast **ω** at these identified locations. At each identified location, we select a set of predictors based on the partial autocorrelation to the response, backward stepwise selection using an F test (Chambers, J. M., 1992: Hastie, T. and Pregibon, D., 1992: Hastie, T. *et al.*, 2000) and partial mutual information (Sharma, A., 2000). This research uses the same predictor variables for all the seasons. Seasonality is represented by using a seasonal intercept or a variable that remains constant for each season (month). Based on the predictor selection procedure outlined above, the optimal predictors were identified as autoregressive lags of the mixture ratio of order 3, 6 and 12 months and residual ratio of 3 and 12 months as shown in Table 5.1. The approach of using minimal predictor set ensures strong parsimony, an important feature of any forecasting model. Note that the predictors are limited to persistence only, to maintain simplicity in our presentation we do not attempt to search for any predictors exogenous to the observed weights.

Spatial dependence in the dynamic weight is represented by using predictors that represent a weighted average over a local spatial domain, as explained through Equation [9]. In this study the correlation coefficient depicts the parameter $\{c_n\}$ of Equation [9]. An exhaustive cross correlation analysis of raw predictors at each grid points against all other grid locations concludes that the influence beyond ±20° distances reduces signifi cantly. Figure 5.3 displays one such analysis at a grid point in equatorial Pacific Ocean. Accordingly each raw predictor vector is smoothed by a weighted linear combination of neighbouring predictors within ±20° where c_n >0.4.

Table 5.1 Predictor variables for pair wise model combinations.

MODELS	Predictor vector
GLAR	r_τ ; $\tau = t$– 3, 6, 12 months
OLR	$\acute{r}_\tau$; $\tau = t$– 3, 12 months
KNNW	r_τ ; $\tau = t$– 3, 6, 12 months

r: mixing ratio, ŕ: residual ratio, see Equation [5]

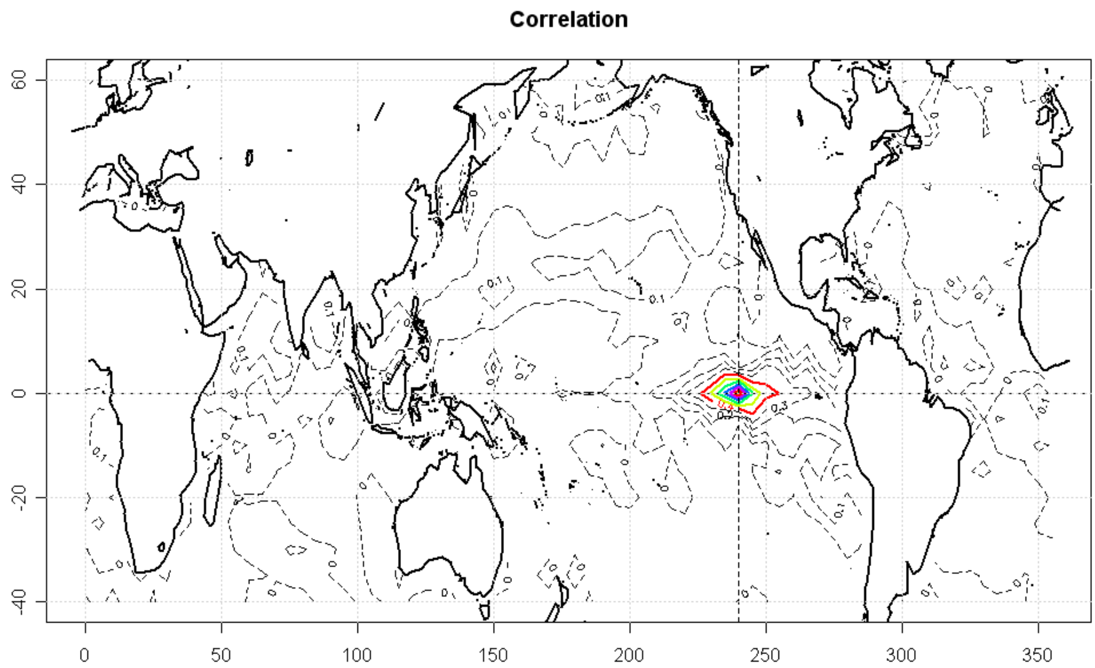


Figure 5.3 Spatial correlation of weights in all cells to a reference cell at 0°N x 240°E. Correlations ≥ 0.4 are drawn in thicker line in colour, and lower values (<0.4) are shown in black broken line. Note the little or no correlation beyond 20° from the reference cell (0°N x 240°E). Similar decay pattern is observed when the reference cell is moved across the ocean surface.

5.3.2 Forecasting dynamic weights

As mentioned earlier in the text, the two methods used to forecast the weights are the mixture GLM and the KNNW. The calibration of the mixture GLM is based on maximizing the likelihood of mixture ratio in GLAR and bias categories in OLR model estimates separately. The value of *K* in KNNW method is estimated based on square root of calibration data length (Lall, U. and Sharma, A., 1996). The influence load β_p of Equation [7] is estimated using scaled absolute value of the linear regression coefficients (Mehrotra, R. and Sharma, A., 2006) of predictand (observed weight) versus the predictor set. Only one set of influence loads are used for the entire world for simplicity. This spatially uniform influence load is ascertained by first estimating linear regression coefficients at all grid points. The marginal distribution for each regression coefficient across the globe is ascertained next and the value corresponding to the highest probability density (mode) is selected as representative for the entire world. Results from these two dynamic weight forecasting methods are presented in the next section.

5.4 Results and Discussion

It should be noted that forecast combination adds parameters supplementary to each component model and hence any combination exercise is prone to over fitting (Pena, M. and van den Dool, H., 2008). Moreover dynamic combination introduces additional parameters and complexity to the overall prediction scheme in comparison to a static combination approach. It is imperative that the performance of any such “complex” model be evaluated in a carefully designed cross-validation setting. We attempt to do so by estimating model prediction error, reflective of the performance of the model in a pure forecast setting. This study validates the results obtained by all procedures by applying the models in four ‘ten year’ blocks. For example, results from 1992 to 2001 are based on parameters that have been calibrated for the period of 1958 to 1991 only. We believe that the ten year gap would minimise any likely artificial inflation of skill because of boundary influences. It is pertinent to note the caution (DelSole, T. and Shukla, J., 2009) against inadvertently instilling artificial skill in validation results by biased predictor selection. The possibility of overestimating validation performance arises if the validation period is not removed prior to predictor selection. In this research the predictors are limited to only three auto regressive terms, the choice of which does not change by expunging the validation data. Besides, as recommended by DelSole and Shukla (2009), the predictors are chosen based on a variety of rigorous statistical analyses (see Section 5.3.1) rather than simple correlation which may exhibit spurious signal. Overall we believe that our validation result reflects realistic predictive skill of the models.

The prediction obtained by the dynamic weights is compared against the static weight prediction which in turn is compared against the selection of the best performing single component model (MetF in this study).

First we compare the extent of the overall reproduction of global SSTA distribution by MetF to that of combined models. The probability density of the observed and predicted SSTA at all grid locations pooled together is presented in Figure 5.4. The figure illustrates unconditional distribution and biases. It shows that the density of the SSTA predictions using dynamic weight combination is closer to the observed SSTA when compared to static weight; whereas static weight combination outperformed the results compared to the best performing single model MetF.

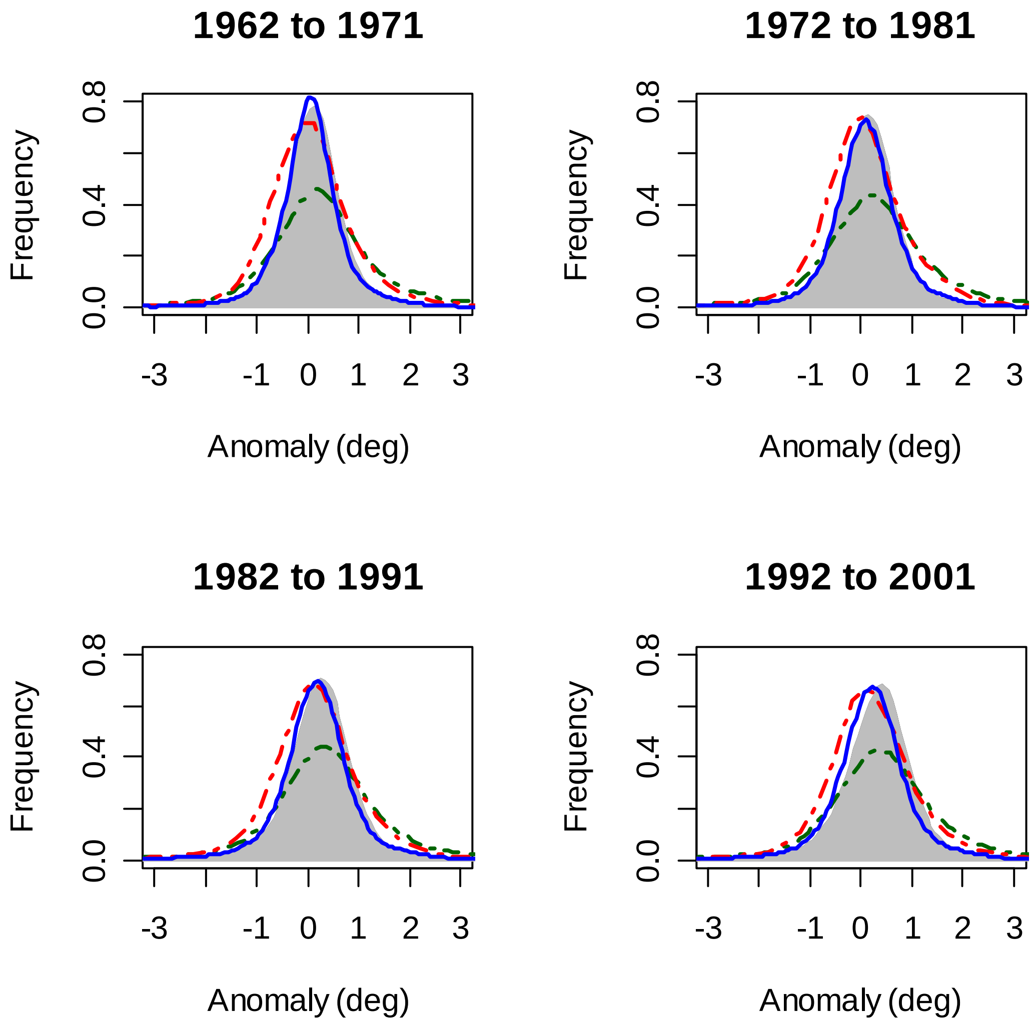


Figure 5.4 The pooled probability densities of the global sea surface temperature anomalies are drawn over four validation periods. The grey shade represents the observed anomaly, the green dash line the MetF prediction, the red dash line the static combination results, and the blue solid line the dynamic combination results. Note the improvement in the representation of the overall variation in the anomalies through the use of the dynamic combination.

This study used mean squared error (MSE) as a relative measure of skill between different models. Table 5.2 demonstrates a consistent reduction of MSE due to model combination across all the four validation blocks. Static combination offered reductions of 5 to 13% of MSE of component models whereas the dynamic combination yields higher reduction of 13 to 21%. There was no case of increased MSE after predictions are combined. The improvements achieved by the either method of forecasting are similar with minor but consistently better performance of KNNW method over the mixture GLM method. We further examined the performance of the combination methods on a month by month basis by analysing (as shown in Table 5.3) the percentage reduction of MSE in each month with respect to MSE of the best single model MetF. The dynamic combination of predictions reduced MSE compared to that of static combination which in turn is better than any single model predictions. We next assess the performance of dynamically combined prediction in representing spatial dependence across the forecasted SSTA field.

Table 5.2 MSE of SSTA predictions across various validation periods.

	DWR	DWK	SW	CPC	ECM	MetF
1962 to 1971	5.51	5.49	6.09	6.97	6.63	6.56
1972 to 1981	5.52	5.50	6.10	6.95	6.68	6.52
1982 to 1991	5.56	5.52	6.11	6.90	6.67	6.49
1992 to 2001	5.58	5.55	6.13	6.97	6.77	6.45

(DWR: dynamic weight forecasted using regression, DWK: dynamic weight forecasted using KNNW, SW: static weight, CPC ECM MetF: component models)

The contours of MSE of statically combined predictions minus that of dynamically combined prediction using the KNNW approach across the global sea surface grid are presented in Figure 5.5. The cells with positive difference denote improvement and are shown using blue shades and the cells with no improvement are shown in yellow, the cell counts being summarised in Table 5.4. Figure 5.5 and Table 5.4 demonstrates that 96 to 98% of the locations exhibited improved predictions due to dynamic weight method compared to static weight. The decrease of sum of squared error is found statistically significant when analysed by a paired one tailed t test ($p= 2.94 \times 10^{-5}$). Similarly paired one tailed t test of MSE at all the grid points accepted the null hypothesis that the MSE of KNNW is less than that of mixture GLM at p=0.0034.

Table 5.3 Percentage reduction of MSE using the dynamic combination approach compared to the MSE of the MetF model. The errors of all four 10 year blocks (1962-2001) validations are aggregated here.

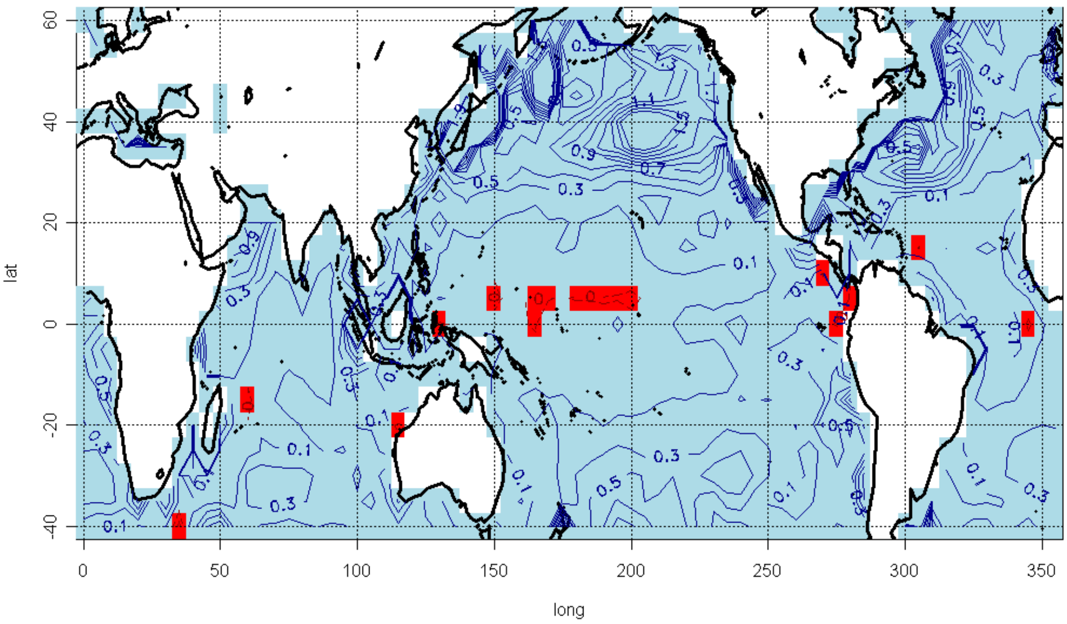
	DWR	DWK	SW		DWR	DWK	SW
Jan	15.7	16.2	9.4	Jul	28.3	29.3	19.6
Feb	10.4	10.8	3.9	Aug	12.1	13.1	7.6
Mar	8.3	8.7	1.8	Sep	7.6	8.1	0.1
Apr	18.8	19.1	9.8	Oct	15.9	16.2	3.7
May	14.7	15	2.8	Nov	12.9	13.3	0.2
Jun	16.9	16.8	7.2	Dec	12.1	11.9	2.5

(DWR: dynamic weight forecasted using regression, DWK: dynamic weight forecasted using KNNW, SW: static weight.)

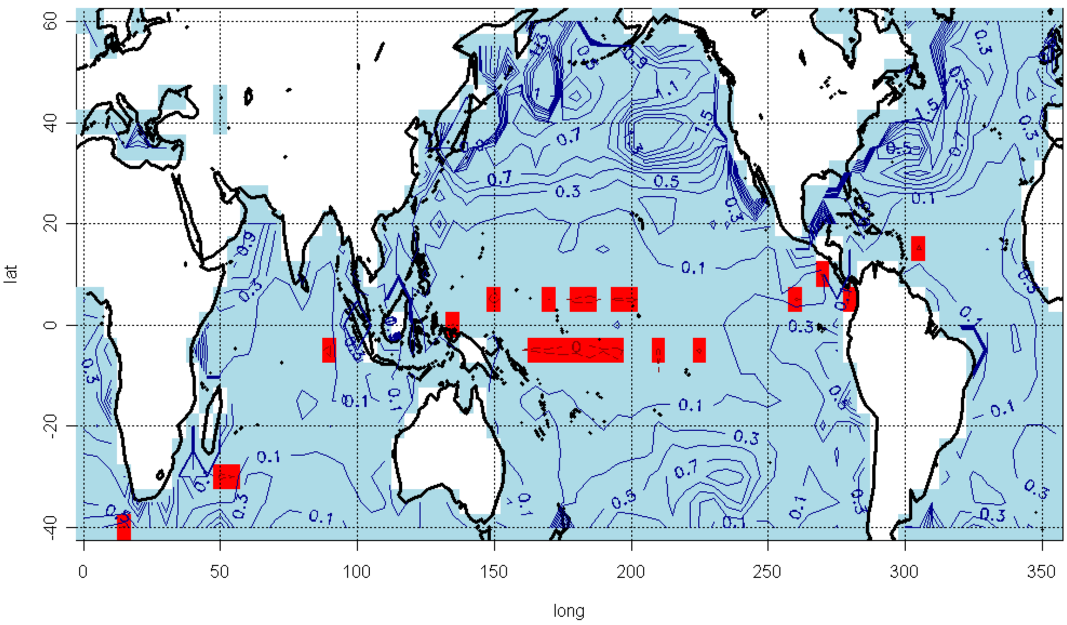
Table 5.4 Percentage of cells where the KNNW dynamically combined model MSE is smaller compared to that of static weight combination.

Years	1962 to 1971	1972 to 1981	1982 to 1991	1992 to 2001
Percent	98.2	97.7	97.7	96.1

1962 to 1971



1972 to 1981



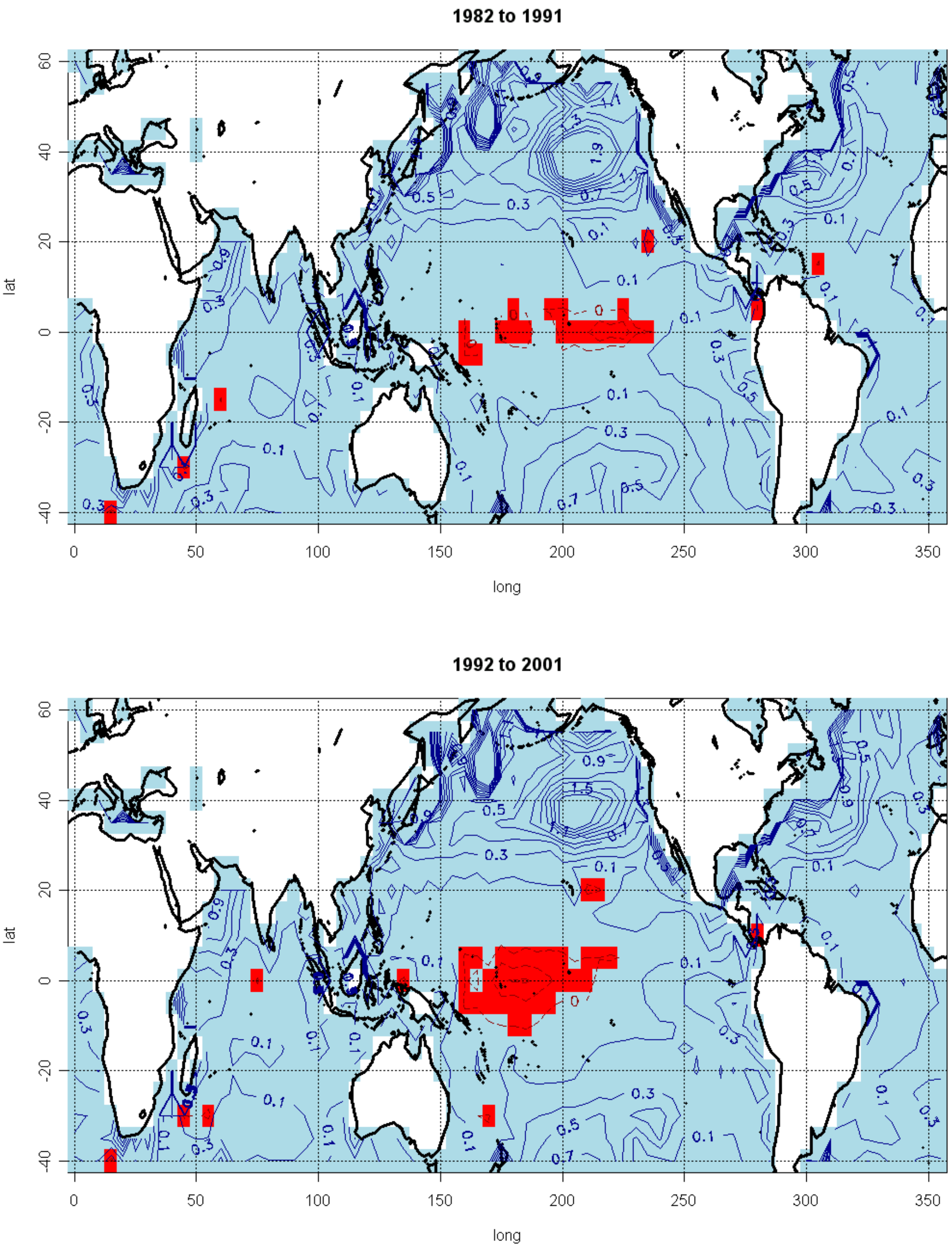


Figure 5.5 The reduction in prediction MSE due to the KNNW dynamic weight combination compared to that of static weight combination. The lighter shades (blue) and solid contours are the zones with improved prediction (or lower MSE). The darker boxes (red) and broken contours are the zones exhibiting indifference or worse prediction.

The representation of spatial dependence is analysed by measuring linear inter-dependence across the global grid. The correlation of the SSTA time series at a reference grid point to all other cells in the grid is measured first. The reference grid point is then moved across all the available grid cells $\{l=1, 2, \ldots l_{max}\}$, resulting in $\frac{1}{2}\, l_{max}\, (l_{max} -1)$ estimates of the correlation. A subset of these correlations is compared against their historical values as illustrated in Figure 5.6. Few systematic losses (or gain) of spatial correlation were evident except where the observed correlation is low. The artificial inflation of linear dependence around low correlation zone can be attributed to the underlying component model predictions. We have drawn a least square fitted line to correlations for both the dynamically combined prediction and the best component prediction MetF. The fitted line correspond to the combined prediction is closer to the 1:1 line indicating a better match to the observed statistics.

Our case study found that the improvement due to dynamic combination is spatially and temporally consistent. We observed that due to the lower accuracy of the component models at extra tropics (hence higher scope of improvement), the reduction of MSE around extra tropics is higher than that of equatorial zones. The weaker improvements around equatorial zones concur the earlier study of univariate dynamic weight application of CS2009 study that had achieved small improvement in NINO3.4 prediction. The spatial consistency of improvement is scrutinized as follows. First, a spatial spread of minimum MSE out of three models (ECM, MetF, CPC) separately at each grid point is selected. The percent improvement over this *potential minimum MSE* spread is drawn in Figure 5.7. We found a spatially consistent reduction of MSE (by approximately 50%) due to dynamic combination. On the contrary, the static combination could not outperform the best component model in significantly large pockets in the South Pacific Ocean and around the east and west coast of the North American continent. This observation is analogous to past studies that had experienced lesser skill of combined models in the extra tropics (Colman, A. W. and Davey, M. K., 2003).

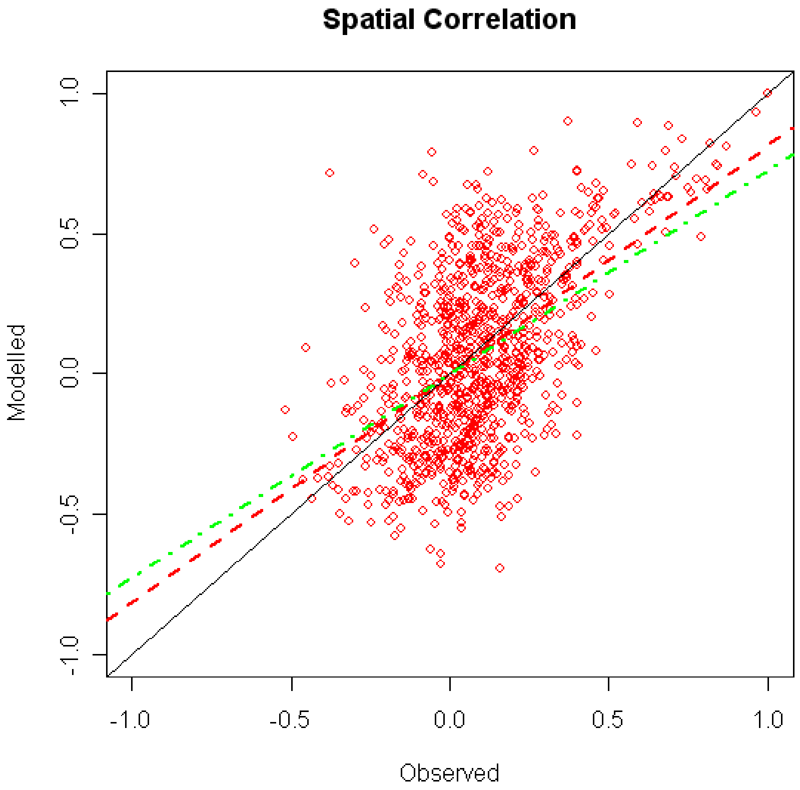


Figure 5.6. Correlation of SSTA at a reference grid point to the rest of the sea surface grid, the reference grid point being rotated across the grid. The correlation ascertained using the predicted time series is plotted against that using the observed time series during a validation period. Only 1000 randomly selected estimates of the correlation are drawn for clarity. The diagonal is the 1:1 line and the red dash line is the least square fitted line to the points. The green dash dot line is the least square fitted line to MetF correlation vectors (not shown here).

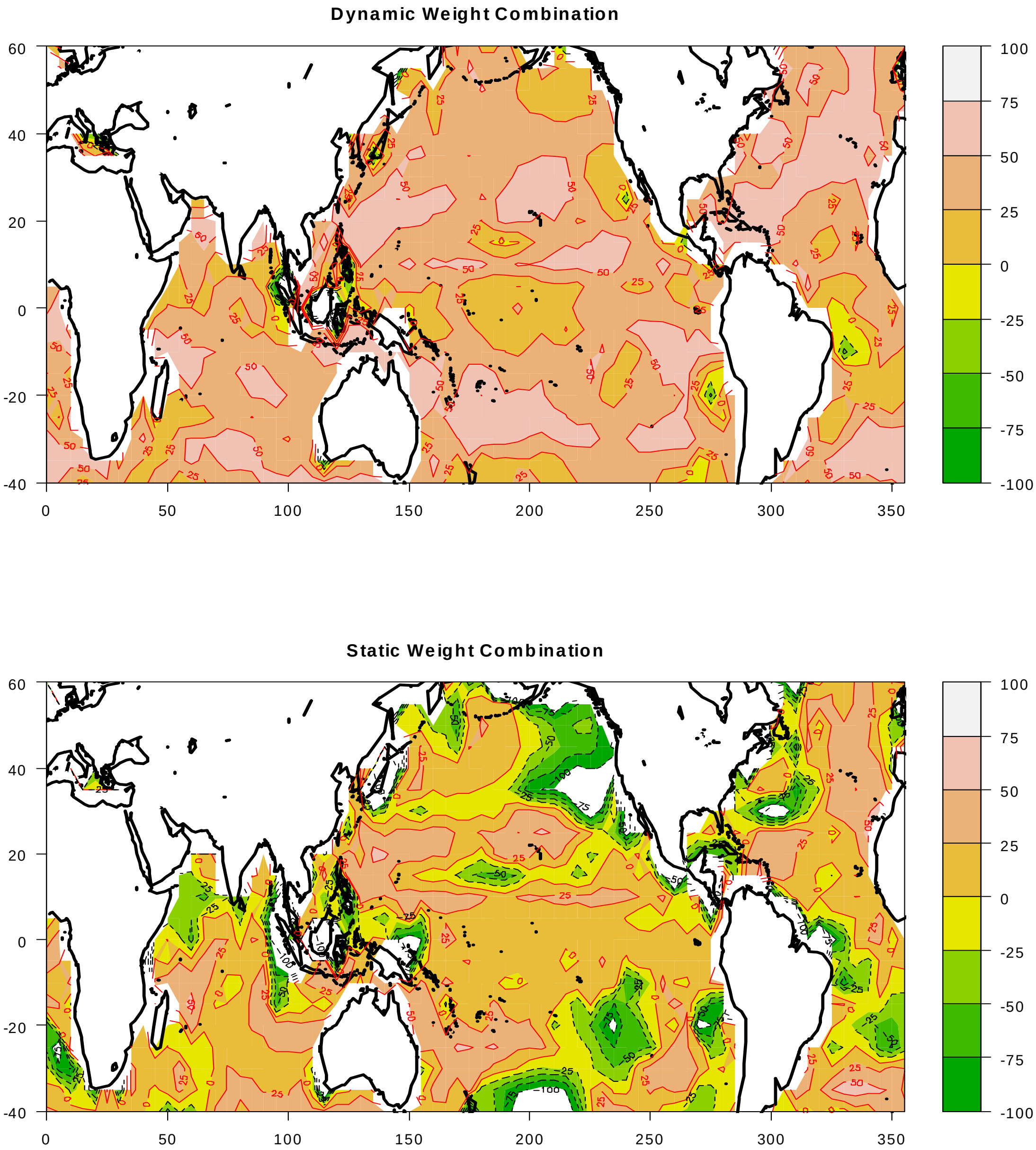


Figure 5.7 The percent reduction of MSE at each grid location compared to the best component model at that grid point. The positive reduction denotes improvement. The positive contours are red solid line, the negative contours are drawn in black broken line.

In this study, the dynamic combination is only applied at the highest pair of the combination tree (Figure 5.1) unlike the pair wise dynamic weight combination method of CS2009. This simplification in the combination method architecture did not worsen the MSE results for our case study significantly. Ideally one can expect predictions to be improved if the pair wise combination tree contains only one dynamic weight node with the rest remaining as static weight nodes. While multiple nodes of dynamic weights provide multiple possibilities of improvement, we recommend a single dynamic node only at the highest tree level in case of multivariate prediction, a sensible simplification when the number of component models is high. Dynamic combination is appropriate for cases when the number of alternative forecasts is small and the length of hindcasts is long. The issues associated to combining a large number of forecasts with small period of hindcast are discussed in detail by Pena and van den Dool (2008). This paper does not investigate the effect of varying the number of component models or combination tree architecture.

5.5 Conclusion

This study illustrated two methods of dynamically combining multiple multivariate predictions of globally grided sea surface temperature anomalies available at 5°x5° grids. These anomalies were pred icted three months in advance for a period of 1958 to 2001 using three separate models (referred to as component models in this study). As a first step, the observed vectors of combination weights (i.e. dynamic weights) were computed from the hindcast time series of the SSTAs at each sea surface temperature grid cell. Next, two approaches for forecasting multivariate dynamic weights 3 months ahead were applied, and performance tested using independent validation in 10-year blocks. The first of the two approaches uses a mixture of a three category ordered logistic regression model to forecast the state of bias of component prediction pairs and a generalised linear autoregressive model to ascertain the magnitude of the mixing weights in case of opposing bias is forecasted. The second approach uses a nearest neighbour formulation where Euclidean distances are ascertained based on the relevance of each predictor to the observed dynamic weight vector. The prediction skill from static weight combination was used as the base case to compare the merits of using dynamic weights instead. The predicted sea surface temperature using the dynamic combination algorithm consistently exhibited better accuracy, both in space and time, to that of the static combination. Improved skill, in four 10 year validation blocks tested, is achieved in 96% or more global grids with the rest showing indifference to static weight skills.

5.6 Acknowledgement

This research was funded by the Australian Research Council and the Sydney Catchment Authority. We acknowledge the assistance of Upmanu Lall, Andrew Robertson and Lisa Goddard of the Lamont Doherty Earth Observatory Centre of Columbia University, New York. The computation was performed using the freely available R statistical computing platform (R Development Core Team, 2008).

5.7 Reference

Agresti, A., 1996: *An Introduction to Categorical Data Analysis*. *Wiley Series in Probability and Statistics*.

Armstrong, J. S., 2001: Combining forecasts. *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong, Ed., Kluwer Academic Publishers, 417-439.

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,** 1783-1796.

Bates, J. M. and C. W. J. Granger, 1969: The combination of forecasts. *Operational Research Quarterly*, **20,** 451-468.

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: Global Ocean Surface Temperature Atlas.

Bruhn, J. A., W. E. Fry, and G. W. Fick, 1980: Simulation of Daily Weather Data Using Theoretical Probability Distributions. *Journal of Applied Meteorology***,** 1029-1036.

Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen, 2004: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, **298,** 242-266.

Chambers, J. M., 1992: Linear Models. *Statistical Models in S*, J. M. Chambers and T. Hastie, Eds., Wadsworth & Brooks, 95-144.

Chowdhury, S. and A. Sharma, 2009a: Long Range NINO3.4 Predictions Using Pair Wise Dynamic Combinations of Multiple Models. *Journal of Climate*, **22,** 793-805.

Clemen, R. T., 1989: Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, **5,** 559-583.

Colman, A. W. and M. K. Davey, 2003: Statistical prediction of global sea-surface temperature anomalies. *International Journal of Climatology*, **23,** 1677-1697.

DelSole, T. and J. Shukla, 2009: Artificial Skill Due to Predictor Screening. *Journal of Climate*, **22,** 331-345.

Deutsch, M., C. W. J. Granger, and T. Terasvirta, 1994: The combination of forecasts using changing weights. *International Journal of Forecasting*, **10,** 47-57.

Devineni, N., A.Sankarasubramanian, and S. Ghosh, 2008: Multimodel Ensembles of Streamflow Forecasts: Role of Predictor State in Developing Optimal Combinations. *Water Resources Research*, **44,** W09404.

ECMWF, 2004: Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER)434.

Fraedrich, K. and N. Smith, 1989: Combining Predictive Scheme in Long-Range Forecasting. *Journal of Climate*, **2,** 291-294.

Georgakakos, K. P., D.-J. Seo, H. Gupta, J. Schaake, and M. B. Butts, 2004: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, **298,** 222-241.

Hall, D. B., 2000: Zero-inflated Poisson binomial regression models with random effects: a case study. *Biometrics*, **56,** 1030-1039.

Harrold, T. I., A. Sharma, and S. J. Sheather, 2003: A nonparametric model for stochastic generation of daily rainfall amounts. *Water Resources Research*, **39**.

Hastie, T. and D. Pregibon, 1992: Generalised Linear Models. *Statistical Models in S*, J. M. Chambers and T. Hastie, Eds., Wadsworth & Brooks, 195-248.

Hastie, T., R. Tibshirani, and J. Friedman, 2000: *The Elements of Statistical Learning, Data Mining, Inference and Prediction*. Springer Series in Statistics.

Hibon, M. and T. Evgeniou, 2005: To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, **21,** 15-24.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T., 1999: Bayesian Model Averaging: A Tutorial (with discussion). *Statistical Science*, **14,** 382-417.

Huerta, G. and B. Sansó, 2007: Time-varying models for extreme values. *Environ Ecol Stat*, **14,** 285-299.

Lall, U. and A. Sharma, 1996: A nearest neighbor bootstrap for time series resampling. *Water Resources Research*, **32,** 679-693.

Lundberg, S., T. Tera¨svirta, and D. van Dijk, 2000: Time-varying smooth transition autoregressive models.376.

Madec, G., P. Delecluse, M. Imbard, and C. Levy, 1997: OPA Release 8, Ocean General Circulation Model Reference Manual, 200 pp.

Marshall, L., D. Nott, and A. Sharma, 2007: Towards Dynamic Catchment Modelling: A Bayesian Hierarchical Modelling Framework. *Hydrological Processes*, **21,** 847-861.

Martin, T., B. Wintle, J. Rhodes, P. Kuhnert, S. Field, S. Low-Choy, A. yre, and H. Possingham, 2005: Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8,** 1235-1246.

McLeod, A. I., D. J. Noakes, K. W. Hipel, and R. M. Thompstone, 1987: Combining hydrologic forecast. *Journal of Water Resources Planning Management*, **113,** 29-41.

Mehrotra, R. and A. Sharma, 2006: Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach. *Advances in Water Resources*, **29,** 987-999.

Miller, C. M., R. T. Clemen, and R. L. Winkler, 1992: The effect of nonstationarity on combined forecasts. *International Journal of Forecasting*, **7,** 515-529.

Pena, M. and H. van den Dool, 2008: Consolidation of Multi Model Forecasts by Ridge Regression: Application to Pacific Sea Surface Temperature. *Journal of Climate*, **21,** 6521-6538.

Peng, P., A. Kumar, and H. van den Dool, 2002: An analysis of multimodel ensemble predictions for seasonal climate anomalies. *Journal of Geophysical Research*, **107,** 4710.

R Development Core Team, 2008. R Foundation for Statistical Computing.

Raftery, A. E., T. Geneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review*, **133,** 1155-1174.

Ragonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona, 2006: A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, **42,** W09404.

Reynolds, R. W. and T. M. Smith, 1995: A High-Resolution Global Sea Surface Temperature Climatology. *Journal of Climate*, **8,** 1571-1583.

Robertson, A. W., U. Lall, S. E. Zebiak, and L. Goddard, 2004: Improved Combination of Multiple Atmospheric GCM Ensembles for Seasonal Prediction. *Monthly Weather Review*, **132,** 2732-2744.

Sanders, F., 1963: On subjective probability forecasting. *Journal of Applied Meteorology*, **2,** 191-201.

See, L. and R. J. Abrahart, 2001: Multi-model data fusion for hydrological forecasting. *Computers & Geosciences*, **27,** 987-994.

Shamseldin, A. Y., O'Connor, K.M., Liang, G.C., 1997: Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology*, **197,** 203-229.

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

Sharma, A. and U. Lall, 2004: Model averaging and its use in probabilistic forecasting of hydrologic variables. *Hydrology, Science and practice for the 21st century*, Imperial College, London, British Hydrological Society, 372-378.

Shephard, N., 1995: *Generalized Linear Autoregressions*. February 6 ed. *Economics Working Paper 8*, Nuffield College.

Smith, T. M. and R. W. Reynolds, 2002: Extended reconstruction of global seas surface temperatures based on COADS data 1854 to 1997. *Journal of Climate*, **16,** 1495-1510.

Souza Filho, F. A. and U. Lall, 2003: Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. *Water Resources Research*, **39,** 1307.

Srikanthan, R. and T. A. McMahon, 2001: Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, **5,** 653-670.

Terui, N. and H. K. van Dijk, 2002: Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, **18,** 421-438.

Thompson, P. D., 1977: How to Improve Accuracy by Combining Independent Forecast. *Monthly Weather Review*, **105,** 228-229.

van den Dool, H., 2000: Constructed analogue prediction of the east central tropical Pacific SST and the entire world ocean for 2001. *Experimental Long-Lead Forecast Bulletin*, **9,** 38-41.

van den Dool, H., J. Huang, and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. *Journal of Geophysical Research*, **108,** 8617.

West, M. and J. Harrison, 1997: *Bayesian Forecasting and Dynamic Models*. 2nd ed. Springer-Verlag.

Wolff, J. E., E. Maier-Reimer, and S. Legutke, 1997: The Hamburg Ocean Primitive Equation Model. Technical report no. 13.

Xiong, L., A. Y. Shamseldin, and K. M. O'Connor, 2001: A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. *Journal of Hydrology*, **245,** 196-217.

Yan, Z., S. Bate, R. E. Chandler, V. Isham, and H. Wheater, 2002: An Analysis of Daily Maximum Wind Speed in Northwestern Europe Using Generalized Linear Models. *Journal of Climate*, **15,** 2073-2088.

Yang, C., R. E. Chandler, and V. S. Isham, 2005: Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research*, **41,** W11415.

Yu, L., S. Wang, and K. K. Lai, 2005: A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Computers & Operations Research*, **32,** 2523-2541.

Zou, H. and Y. Yang, 2004: Combining time series models for forecasting. *International Journal of Forecasting*, **20,** 69-84.

CHAPTER SIX

This chapter reprinted the following article:

Chowdhury, S. and Sharma, A. (2009), Multi-site Seasonal Forecast of Arid River Flows Using a Dynamic Model Combination Approach. *Water Resources Research*, under review.

This work has been submitted for publication. Copyright in this work may be transferred to *American Geophysical Union* without further notice, and this version may no longer be the latest version.

6 MULTI-SITE SEASONAL FORECAST OF ARID RIVER FLOW

ABSTRACT

This paper dynamically combines three independent forecasts of multiple river flow volumes a season in advance for arid catchments with high variability. The case study considers five inflow locations in the upper Namoi Catchment of Eastern Australia. The seasonal flows are predicted based on concurrent Sea Surface Temperature Anomalies (SSTA), which are predicted a season forward using a dynamic combination of three SSTA forecasts. The river flows are predicted using three statistical forecasting models: 1) a mixture of generalised log normal and multinomial logit models; 2) the local regression of independent components of five inflows; 3) the weighted nearest neighbour method; where each of these models use the predicted SSTA along with prior lags of the flow as the main driving variables. The study demonstrates that improved SSTA forecast (due to dynamic combination) in turn improves all three flow forecasts, while the dynamic combination of the three flow forecasts results in further, although smaller, improvements.

6.1 Introduction

Seasonal forecast of river flow is vital for efficient water resource management, aiding irrigation, hydropower generation, flood mitigation, drinking water supply and managing water dependant eco systems. The water availability in the coming irrigation season is a prime consideration to decide plantation areas of water sensitive annual crops at the start of the season. The potential benefit of seasonal flow forecast is high for agricultural economies of arid regions (Podbury et al. 1998, Letcher et al. 2004). Hydrologists have a number of models at their disposal to forecast river flow, with a varying degree of success. The structural uncertainty of a single model can be reduced by combining alternative modelling platforms. This paper investigates the scope of improving forecast skill by dynamically combining alternative forecasting approaches. Various hydrological studies have reported improvement of flow forecast after combination of multiple methods (Ajami et al. 2006, Devineni et al. 2008, Georgakakos et al. 2004, Goswami and O'Connor 2007, Ragonda et al. 2006, Shamseldin 1997). This research uses a pair wise dynamic combination approach first introduced by Chowdhury and Sharma (2009a) in the context of combining forecasts of the NINO3.4 sea surface temperature anomaly (SSTA) based index representative of the strength of an El Nino Southern Oscillation anomaly. The improvement of global SSTA forecast was recently reported by Chowdhury and Sharma (2009a, 2009b) after dynamically combining three SSTA forecast models. Can similar improvements be achieved for flow forecast models? When can we expect forecast combination to exhibit improvement? What improvement of flow forecast is delivered by better global SSTA forecast? This paper seeks to explore these questions.

This study forecasts stream flow directly circumventing the uncertainty of needing to specify rainfall runoff models. This approach has certain advantages especially for regions with few rainfall events and sparse rain gauge networks. For example, flow is mostly a continuous variable with considerable memory which assimilates information spatially. Uncertainty of flow measurement is less demanding than the extrapolation of point rain gauge into catchment scale. Consequently forecast of flow rather than rainfall is often recommended (Chiew et al. 1998, Dutta et al. 2006), which is strengthen by the fact that stream flow is the ultimate variable of interest for water resource management. The main challenge in such a predictive framework is to identify the climate variables that constitute predictors of flow.

Various indices and transformations of SSTA have been found to be useful as predictors of hydrological variables such as rain and flow (Sharma 2000c, Drosdowsky and Chambers 2001a, Verdon et al. 2004). The relationship of El Nino Southern Oscillation (ENSO) and flow has been well established in many parts of the world (Hamlet and Lettenmaier 1999, Chiew and McMahon 2002, Muluye and Coulibaly 2007). Accordingly, the concurrent SSTA field at Pacific and Indian Oceans are used as the predictor source in our case study.

The case study comprise of forecasting seasonal flow at five locations of an arid catchment. The details of the catchment and flow characteristics are documented in Section 6.3. Since, forecast error of any model is generally a culmination of imprecise predictors and structural uncertainty (Butts et al. 2004, Huard and Mailhot 2006); we first attempt to reduce the flow forecast error by using a more precise predictor approximation selected from a SSTA forecast field that arises through a dynamic model combination approach (Chowdhury and Sharma, 2009b). Secondly the structural uncertainty is reduced by considering three flow forecast models and using a dynamic combination approach to arrive at the final forecast. Lastly, we analyse the performances of the various forecast methods considered and scrutinise the advantages of forecast combination in light of our case study.

This paper is organised as follows. First we describe the methodology of dynamic combination algorithm preceded by background literature on time series combination. Then we illustrate the benefit of dynamically combining three SSTA forecasts as reported in Chowdhury and Sharma (2009b). Section 6.3 describes the study catchment along with the three flow forecast models considered. This is followed by an analysis of results and related discussions on the merits and demerits associated with the various approaches presented.

6.2 Model Combination

6.2.1 Background

Combinations of multiple forecasts have been widely adopted in practice in the time series forecasting discipline (Clemen 1989, Hoeting 1999, Armstrong 2001). There has been various studies supporting and analysing a range of forecast combination methods (Menezes et al. 2000, Ajami et al. 2006, Kim et al. 2006, Devineni et al. 2008) that provide a good background to this area of research. In general the model combination involves linear combination of multiple response time series where the combination weights remains time invariant. While McLeod *et al.* (1987) were the first to propose a non stationary model combination weight, only few studies using such dynamic weights in hydrology (See and Abrahart 2001, Xiong et al. 2001, Marshall et al. 2007) have been reported since. One approach is to combine the models in pairs using combination weights that vary in time reflecting the persistence of individual model skills. A summary of the pair-wise dynamic weight combination method is presented next, with readers being referred to Chowdhury and Sharma (2009a, 2009b) for additional details on this topic.

6.2.2 Dynamic weight combination

Consider the case of two component hydrologic forecasts, $\hat{u}_{1,l,t}$ and $\hat{u}_{2,l,t}$, at a location *l*. For ease of notation let us conceal the location subscript *l* within the straight font notation as $\hat{\mathrm{u}}_{1,t} \equiv \{\hat{u}_{1,l,t}\, ; l=1,2,3..\}$. These can be combined as:

$$\hat{\mathbf{y}}_t = \hat{\mathrm{u}}_{1,t}\, \omega_t + \hat{\mathrm{u}}_{2,t}\, (1 - \omega_t) \qquad [1]$$

Where $\hat{\mathbf{y}}_t$: combined forecast at time *t* and ω_t: weight dynamically assigned to model 1.

The observed dynamic weight is defined as:

$$\omega_t = \mathrm{e}_{2,t} / (\mathrm{e}_{2,t} - \mathrm{e}_{1,t}) \qquad [2]$$

Here $\mathrm{e}_{1,t}$ and $\mathrm{e}_{2,t}$ are residuals of the model 1 and 2. Ignore any forecast bias and constraint the weights to positive fractions only, $\{\omega_t \in 0 \rightarrow 1\}$. The combination procedure is as follows.

First, prepare a time series of weights {ω_t; t=1,2,3,..} using component model hind cast residuals as shown in Equation (2). Next predict the weight time series forward by formulating a model with appropriate predictors and auto regressive components as well as multi site characteristic of ω_t that is {$\omega_{l,t}$; l=1,2,3..}. Two such models were presented in Chowdhury and Sharma (2009a), one being a mixed regression model and the other being a weighted nearest neighbour method. Once such a predictive model has been formulated, we use the predicted future weight (say ω_{t+1}) to combine two alternative forecasts $\hat{u}_{1,t+1}$ and $\hat{u}_{2,t+1}$ into $\hat{\mathbf{y}}_{t+1}$. Note that the model combination operates on a pair-wise basis, with multiple pairs being formulated if more than two model forecasts are to be combined. Details on the rationale behind the model combination and the logic for imparting spatial dependence in multivariate forecast fields are presented in Chowdhury and Sharma (2009b).

This paper demonstrates weighted combination of mean forecast only. As a result, the information about the uncertainty in the process contained in the forecast ensemble is not retained in the dynamic combination forecasts. However, estimates of the standard error associated with the forecasts can be obtained by building a conditional variance model (similar to the dynamic combination that is analogous to a conditional mean), or alternately by using well formulated conditional bootstrap alternatives to develop nonparametric error estimates. A practical application of improvements due to dynamically combining three separate sea surface temperature forecast is described next.

6.2.3 Combining sea surface temperature forecasts

Concurrent reconstructed sea surface temperature anomalies (SSTA) at 5° by 5° grids of the global sea surface between 60°N to 40°S are prime source of flow predictors used in this research. Reconstructed, monthly sea surface temperature anomalies, known as the Kaplan Optimal Smoother SSTA (Kaplan et al. 1997, Kaplan et al. 1998) are source of observed SSTA time series. This reconstructed dataset extends from 1856 to 2003.

One, two and three months SSTA forecast are necessary to predict flow volumes over the next three months. Two sets of SSTA forecasts have been used in the study reported here. The first of these originates from Météo-

France (Madec et al. 1997), which is referred as MetF here, it is developed by the DEMETER project (ECMWF 2004) of European Centre for Medium Range Weather Forecast. The second forecast set is developed by dynamically combining the following three models. They are (a) MetF, (b) another DEMETER model that is referred to as ECM (Wolff et al. 1997), and (c) a statistical model from Climate Prediction Centre (CPC) of the National Oceanic and Atmospheric Administration, USA (van den Dool 2000, van den Dool et al. 2003). The common hindcast period of these models used in our study is 1958 to 2001.

The combined SSTA forecast is named the DW model here, named after the dynamic combination method that is used. Figure 6.1 shows the percent reduction of mean of squared error of the DW forecast compared to that of MetF forecast for each SSTA grid cell considered. The MetF forecast is chosen for comparison due to its smallest error variance to that of CPC and ECM. The figure shows a considerable improvement of the SSTA forecast post combination; most of the region returns a 25 to 75% reduction in mean squared error. Can flow forecast be improved to the similar extent by using this improved SSTA forecast as the basis of concurrent prediction? Additionally, could further improvements in flow forecasts be possible by considering a dynamic combination of multiple flow forecast models? These are two of the questions we seek to address in the remainder of this paper.

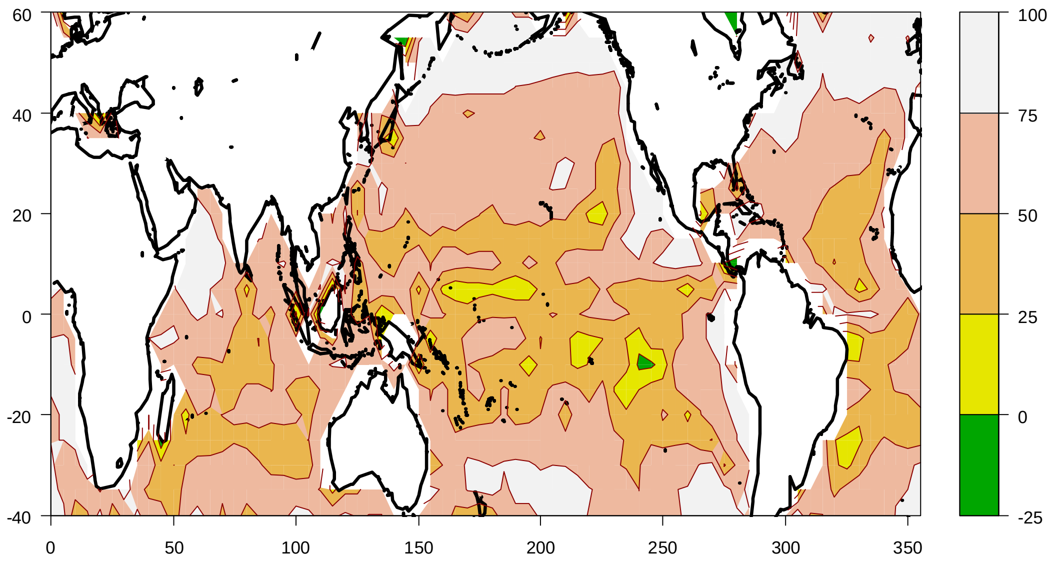


Figure 6.1 Percent reduction of combined SSTA forecast compared to that of single best SSTA forecast (MetF model).

6.3 Catchment Description and Flow Forecasting Methods

6.3.1 Catchment and river description

The Namoi River Catchment, with an area of 42,000 km^2, is a major contributor of flows within the Murray Darling Basin in Eastern Australia. The centroid of the catchment is about 450 km north-west of Sydney. There are three major reservoirs with a total storage capacity of 872x10^6 m^3 and numerous other small dams, weirs and on farm ponds. The river supports 96,000 ha of irrigated agriculture in addition to stock and domestic water use of the local population. Cotton is the major crop grown along with wheat and grazing pasture for the live stocks. Long term mean surface water use is estimated to be 320x10^6 m^3 which is close to half of the available runoff in the catchment (CSIRO 2007). Water extraction is regulated by the Government of the State of New South Wales (NSW). Available water resources are allocated at the start of a sowing season to the irrigators proportional to individual entitlement of annual extraction volume. The allocation is continually revised throughout the irrigation season. Australian climatic variability results in an unreliable pattern such as a low allocation (less than 50%) era of 5 years followed by a resource abundant decade. Hence, it is typical of farming in an arid environment to vary sowing area of irrigated annual crops from year to year. The land developed at the start of the sowing period is dependant on present allocation of irrigation water and a forecast of likely increase in allocation due to higher inflow in next few seasons. Overdevelopment of the land than the available water increases capital expenditure, while underdevelopment amounts to a lost opportunity of the rare water abundant years. Hence improved forecast of next seasons flow in the major rivers in the Namoi Catchment has good economic potential.

The eastern half of the catchment is relatively wet (Preece and Jones 2002) and includes all the major streams notably meeting irrigation demand. This study chose five major river flow locations in the eastern half of the Namoi Catchment as shown in Figure 6.2. The overall natural flow has been artificially altered to a varying degree by extractions, weirs and river regulations since the European settlement in the catchment in the nineteenth century. The flow of Namoi River at Keepit and the Peel River at Carroll Gap are regulated by the Split Rock Dam and the Chaffey Dam constructed in 1987 and 1979. We removed any human induced change in volumes (response time series) prior to seting up the seasonal forecasting models. This is done by modelling the river system by the Integrated Quantity Quality Model (IQQM). The hydrologic model IQQM was progressively developed in the 1990s by the NSW Government in Australia (Simons 1996). This is a conceptual deterministic model that mainly simulates (using a node link structure) daily rainfall runoff, river routing, reservoir operation, irrigation demand and associated extractions subject to legal compliance (Hameed and Podger 2001). The calibrated Namoi IQQM has been applied in various water management studies (CSIRO 2007) including the development of a legal framework to share water resources of the catchment (NSW Government 2003). All the dams, weirs, towns, irrigation and other extraction points are removed from the Namoi IQQM to simulate natural daily flow free of human interference. The natural daily flows of 1898 to 2007 are aggregated to estimate the seasonal flows used in this study.

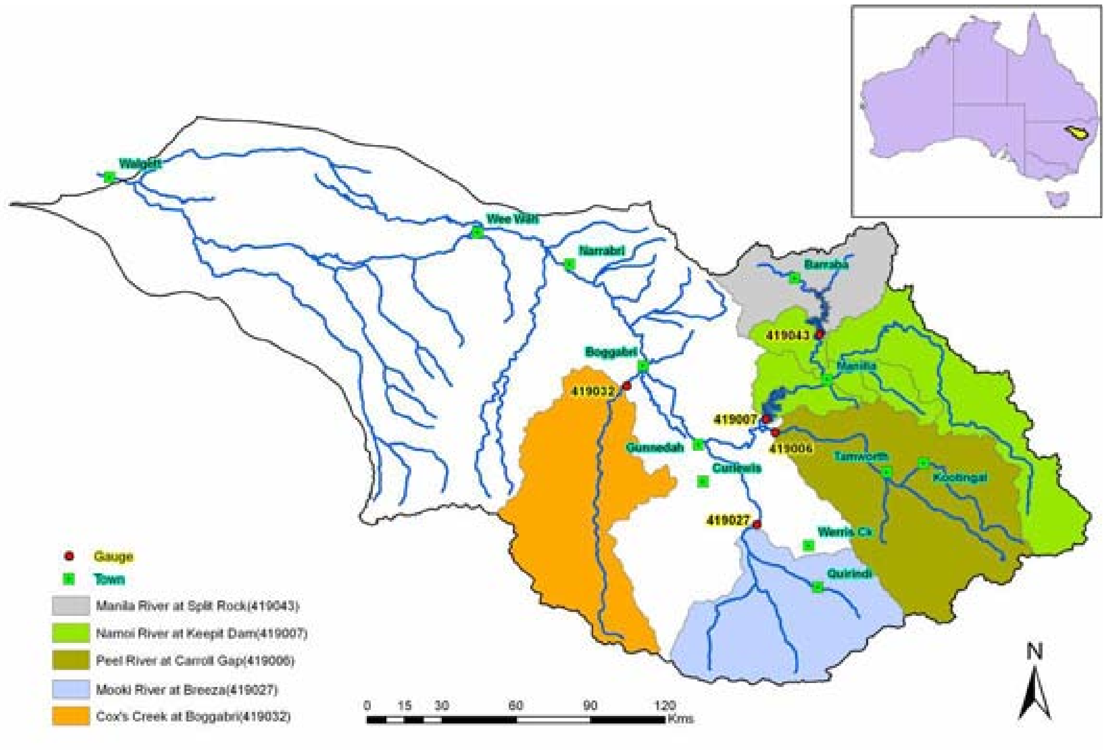


Figure 6.2 Namoi River Catchment flowing in a westerly direction. The five flow locations modelled are shown highlighted.

6.3.2 Flow variability

Strong flow variability is customary to arid rivers in Australia and southern Africa. A number of studies in past and present acknowledged the annual variability (McMahon 1979, Ward 1984, Chiew et al. 2003, McMahon et al. 2007) which tends to get more erratic at seasonal time scales. Mean volumes of Namoi inflows are two to ten times higher than the median flow volumes, see Table 6.1. Considerable spatial difference of flow height within Namoi Catchment were reported in past studies as well (Crapper et al. 1999).

Readers may refer to Kachroo (1992) for an overview of the formulation process behind a flow forecast model. Forecasting flows of arid rivers pose additional challenges related to long hydrograph tails (Anderson and Meerschaert 1998) as seasonal pattern and persistence forms a minor portion of the total variance. A number of prior researches identified various SSTA derived indicators (including NINO3.4) as prime predictors of flow and rainfall in Eastern Australia. However simulation of a variable that exhibits a high coefficient of skewness through a simple (linear) model using predictors that are not as highly skewed is difficult.

Table 6.1. Flow locations, catchment area and associated statistics in units of 1000 m^3. Note the spatial and temporal variability across the five rivers.

Catchments	(km^2)	Min	Med	Mean	Max
Manila River at Split Rock	1650	70	5790	18780	429920
Namoi River at Keepit Dam	5700	830	45410	94560	1050040
Peel River at Carroll Gap	4670	1440	29300	65830	640570
Mooki River at Breeza	3630	0	3150	29090	635860
Cox's Creek at Boggabri	4040	0	1770	16430	417860

This section presents three methods that are primarily aimed at our case study of forecasting flows at five locations using predictors from a concurrent SSTA field. Three flow forecasting models are developed with an aim of demonstrating the dynamic combination of three forecasts. One of the prime considerations in forming these three models is to maximise the structural independence among the three. Prior to presenting detail formula of the flow forecasting models in the following sub sections, let us introduce the notation style for the ease of readership. This paper denotes variables in *italics* for scalar values (e.g. $y_{l,t}$), in straight Romans for vectors (e.g. $\mathrm{\hat{y}}_t = \{\hat{y}_{l,t}\,;\, l=1,2,..\}$) and in **bold** fonts for higher dimension matrices (e.g. $\mathbf{V}=\{v_{i,t};\ i=1,2..I;\ t=1,2,..T\}$). Functions are defined by italic names followed by brackets, such as logarithm being specified as *log*(.). The response vector (river flow) at a location l at time t is denoted by $y_{l,t}$ and random predictors are usually given a notation of X or Z, Greek characters denotes parameters.

6.3.3 Mixture of linear and multinomial regression (GLM)

The Generalized Linear Model (GLM) (McCullagh and Nelder 1989, Chandler 2005, Yang et al. 2005) is an extension of linear regression to include response variables that follow a family of exponential distributions. The exponential family relevant to hydrology includes log Normal, multinomial and Gamma distributions. We have used a mixture of two GLMs' in our forecast model. The first GLM forecasts total flow (sum of flows across the five sites) assuming a log Normal distribution. The second GLM forecast the proportion of the total flow at each site assuming a multinomial distribution. This approach is based on the rationale that overall water availability in a catchment is dictated by global climate indices whilst the spatial variation is more localised related to factors such as the wind direction or vegetation profile (evapo-transpiration). The two stages are outlined below.

The first stage involves a forecast of the sum of five river flows using a generalised linear autoregressive (GLAR) model (Shephard 1995, Yu et al. 2005). The GLAR is a special case of GLM that includes both autoregressive terms and random covariates. The autoregressive variable exploits the persistence structure, a common feature in flow time series, and the random covariate allows inclusion of long range climate indicators known to influence the regional precipitation (Sveinsson et al. 2008). While the choice of sum of flows, instead of single site, simplify the regression into univariate GLAR, the summation also reduces the high variance of an arid river flow system. Besides, a skewness stabilising logarithmic transformation is used to allow proper specification of the forecasting model:

$$log(\hat{Y}_t)= \tilde{Y}_s + \alpha\, log(Y_{t-.})+ \beta\, X_t \qquad [3]$$

$\hat{Y}_t$: sum of multi-site seasonal flow forecasted at time t,

$Y_{t-.}$: sum of multi-site flow matrix at selected earlier time steps,

$\tilde{Y}_s$: seasonally varying intercept where s=1,2,3 and 4.

X_t : multiple exogenous indicators derived using sea surface temperature anomalies,

α, β : GLAR parameters.

The first two elements of Equation [3] model seasonality and persistence in the flow time series. The third element models climatic influences using variables derived from concurrent SSTA fields. Our case study explored various climatic indices and the first five principle components (PC) of the Indian and Pacific Ocean SSTA. Redundant predictors were screened out by backward stepwise model selection using the partial F test (Chambers 1992, Hastie and Pregibon 1992, Hastie et al. 2000). In addition to the seasonal intercept, the auto-regressors in this example were identified as $Y_{t-.} = \{Y_{t-1}; Y_{t-3}\}$. The following exogenous predictors were retained: $X_t = \{PC1^+, PC1^-, PC2\}$, where $PC1^+$ denotes the positive part in the first principal component with negative values being replaced by zeroes, $PC1^-$, the negative part, and PC2 the second principal component. While separating PC1 into the positive and negative parts may not adhere to distributional assumption (e.g. Gaussian) of linear model, it allows non proportional influence of two opposing climate state for example El Nino and La Nina. The loadings of the first two principle components are shown in Figure 6.3. The positive or negative PCs are designed to separate the strong positive or negative anomalies in the transformed SSTA series. This association of SSTA PCs to flows has been reported in numerous past studies (Araghinejad et al. 2006, Cardoso and Dias 2006, Cardoso et al. 2005, Guetter and Georgakakos 1996, Hamlet and Lettenmaier 1999, Hsieh et al. 2003, Piechota et al. 1998).

Next, the total catchment flow $\hat{Y}_t$ is apportioned to various locations using a multinomial logit model (Agresti 1996, pp 206, Augustin et al. 2008). The multinomial distribution is an extension to the widely used binomial distribution (eg. rainfall occurrence model). The multinomial logit model predicts the log odds ratio of proportions of total flow at a site as shown below:

$$log(r_{l,t} / r_{5,t}) = \theta_l + \lambda_l Z_t \qquad [4]$$

$r_{l,t}$: proportion of total flow at a location l at time t, l =1, 2, 3 and 4.

$r_{5,t}$: proportion of total flow at 5^{th} location at time t,

Z_t : multiple predictor set at time t,

θ_l, λ_l : location specific intercept and the coefficient vectors.

It should be noted that subscript *l* in Equation [4] can assume any 4 values, as $r_{5,t} = 1 - \sum r_{l,t}$. A common set of the multiple predictor vector Z_t is used for all 4 locations. In simple terms this can be seen as four linear models predicting the four log odds ratio based on a same predictor set.

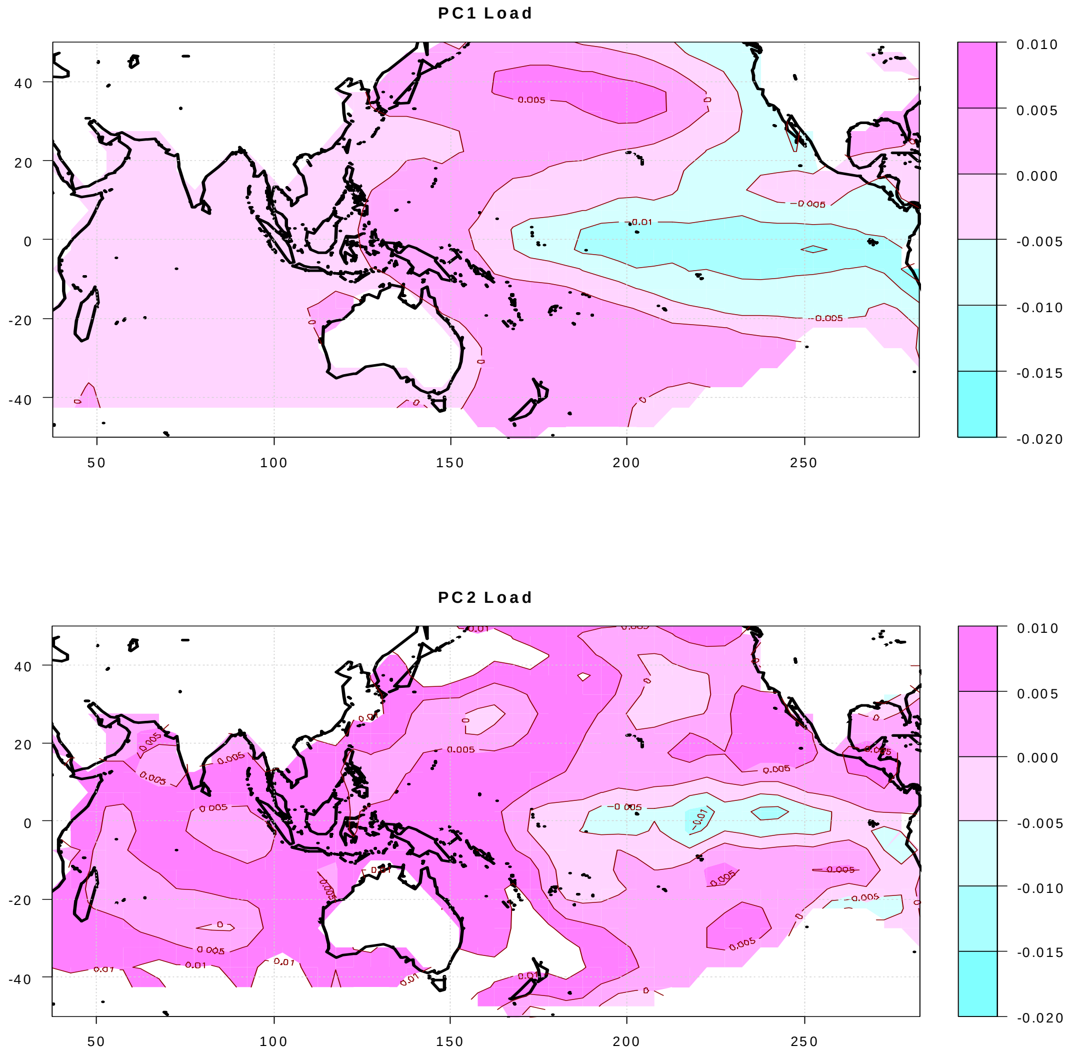


Figure 6.3. Loadings of the first two principle components used as predictors of the forecast model GLM.

Candidate predictors of the multinomial model may include inflow ratios in past time steps of all the rivers along with suitable climatic indices. The stepwise backward removal (Hastie *et al.* 2000) from the pool of candidate predictors of this case study retained two predictor vectors $Z_t = \{Y_{t-1}, r_{2,t-1}\}$.

The first predictor is the summation of the flows at the previous time step, Y_{t-1}. The second predictor is the proportion of the Namoi River flow of the previous season ($r_{2,t-1}= y_{2,t-1}/Y_{t-1}$, where $y_{2,t-1}$ is flow from the 2nd river). These predictor variables are useful indicators of the state of the overall wetness of the system and the spatial spread of the total catchment inflow at the previous time step. The flow at individual site is then estimated as $\hat{y}_{l,t}^{(G)} = r_{l,t} \bullet \hat{Y}_t$.

The above mixture of log normal and multinomial generalised regression models is necessary to replicate spatial dependence across the predicted flows over the five rivers. However there is an alternative way to maintain spatial dependence while utilising the flexibility of univariate regression as shown next.

6.3.4 Independent component and local polynomial (ICM)

The need of a multivariate regression arises from the inter dependence of the flow variables constituting the multiple response vector. The need to have a multivariate model can be removed if the multivariate response matrix can be transformed in a manner that removes the inter dependence. Independent Component Analysis (ICA) yields such transformation (Westra et al. 2007). As per the Central Limit Theorem, linearly mixing a number of independent signals leads to response that approaches a Gaussian probability distribution. ICA reverses the above logic by arguing that there must exist a set of independent signals that can be identified through transformations that result in these variables being maximally non-Gaussian, or, characterised by probability distributions that are as different as possible from the Gaussian distribution. Independent Components (ICs) are the result of un-mixing the multivariate response matrix using the above rationale. The ICA is performed on the log-transformed flow series in which the Markov order 1 persistence structure has been removed (Westra et al. 2008):

$$log(y_{l,t}) = \alpha_s \, log(Y_{t-.}) + q_{l,t} \qquad [5]$$

$y_{l,t}$:flow at location *l* and time *t*,

$Y_{t-.}$:sum of multi-site flow at various earlier time steps,

α_s :seasonal parameter, function of time t, where $s \equiv \{1, 2, 3 \text{ or } 4\}$,

$q_{l,t}$:regression residual at location l and time t,

Due to the common predictors for all 5 flow sites the dependence structure is maintained within the regression residual $\mathbf{Q}=\{q_{l,t}\ ;\ l=1,2..L;\ t=1,2,..T\}$. The derivation of ICs of **Q** simplifies the multivariate regression into multiple univariate regressions. The generic procedure is as follows (Hyvarinen and Oja 2000).

First, the data is centred by subtracting the mean of each column of the data matrix **Q**. The large data matrix may then be condensed by projecting the data onto it's principle component (PC) directions **QE** where **E** is eigenvector matrix. The number of PCs to retain can be specified by the user. Note that the intermediate step of estimating PC is not a pre requisite of ICA. The ICA algorithm deduces the un-mixing matrix **W** subject to **QE.W**=**V**. The un-mixing matrix **W** is chosen to maximize the neg-entropy approximation (Common 1991, Girolami and Fyfe 1996) i.e. non-Gaussianity of the components under the constraints that **W** is an ortho-normal matrix. As a result, the columns of **V** are independent of each other and can be modelled as independent univariate time series. Say $\mathbf{V}=\{v_{i,t};\ i=1,2..I;\ t=1,2,..T\}$ denotes I number of ICs of flow residuals $\mathbf{Q}=\{q_{l,t}\ ;\ l=1,2..L;\ t=1,2,..T\}$ where $I \leq L$. Next **V** is modelled using local regression as described below.

The use of local regression to forecast flow is not new (Grantz et al. 2005, Lall et al. 2006). Local regression blends much of the simplicity of linear least squares regression with the flexibility of nonlinear regression providing a convenient tool to ascertain complex nonlinear relationship between **V** and the SSTA predictor field. Prediction of **V** is done using a locally weighted polynomial regression, called *Loess*, originally proposed by Cleveland (1979) and further developed later (Cleveland et al. 1988, Cleveland and Grosse 1991). At each point in the data set a quadratic curve is fit to a local subset of the data. The polynomial is fit using weighted least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away. The use of the weights is based on the idea that points near the explanatory variable space are more likely to be related to each other in a simple way than points that are further apart. The following second degree local polynomial regression $G_{i,j}(.)$ is used to predict the ICs of the flow residual matrix.

$$v_{i,t} = G_{i,j}(\varphi_j X_t + \psi_j X_t^2) + \varepsilon_{i,t} \qquad [6]$$

X_t : multiple predictor vector that includes persistence and climatic indices

φ_j; ψ_j : locally weighted regression parameters where j is within the neighbourhood of X_t.

$\varepsilon_{i,t}$: regression error of i^{th} IC at time t.

The inverse of ICA (and PC) weight matrices transform back the predicted $\hat{v}_{i,t} \xrightarrow{inv} \hat{q}_{l,t}$, the substitution of estimated regression residual in Equation [5] gives the flow forecast for the ICM method: $\hat{y}_{l,t}^{(I)}$. Note that as each IC is independent, their associated predictors must be independent too. Hence, separate predictor vectors X_t are identified in the forecasting model for each IC.

This case study first deduced 3 ICs of five flow residuals (I=3, L=5). The 3 flow ICs, shown in Figure 6.4, retained 98% of variance. Five ICs of Indian and Pacific Ocean SSTA were used as candidate predictors. The final model also retained three ICs of SSTA, one SSTA IC forms predictors to locally regress one flow IC. The spatial loadings of three SSTA ICs are shown in Figure 6.5.

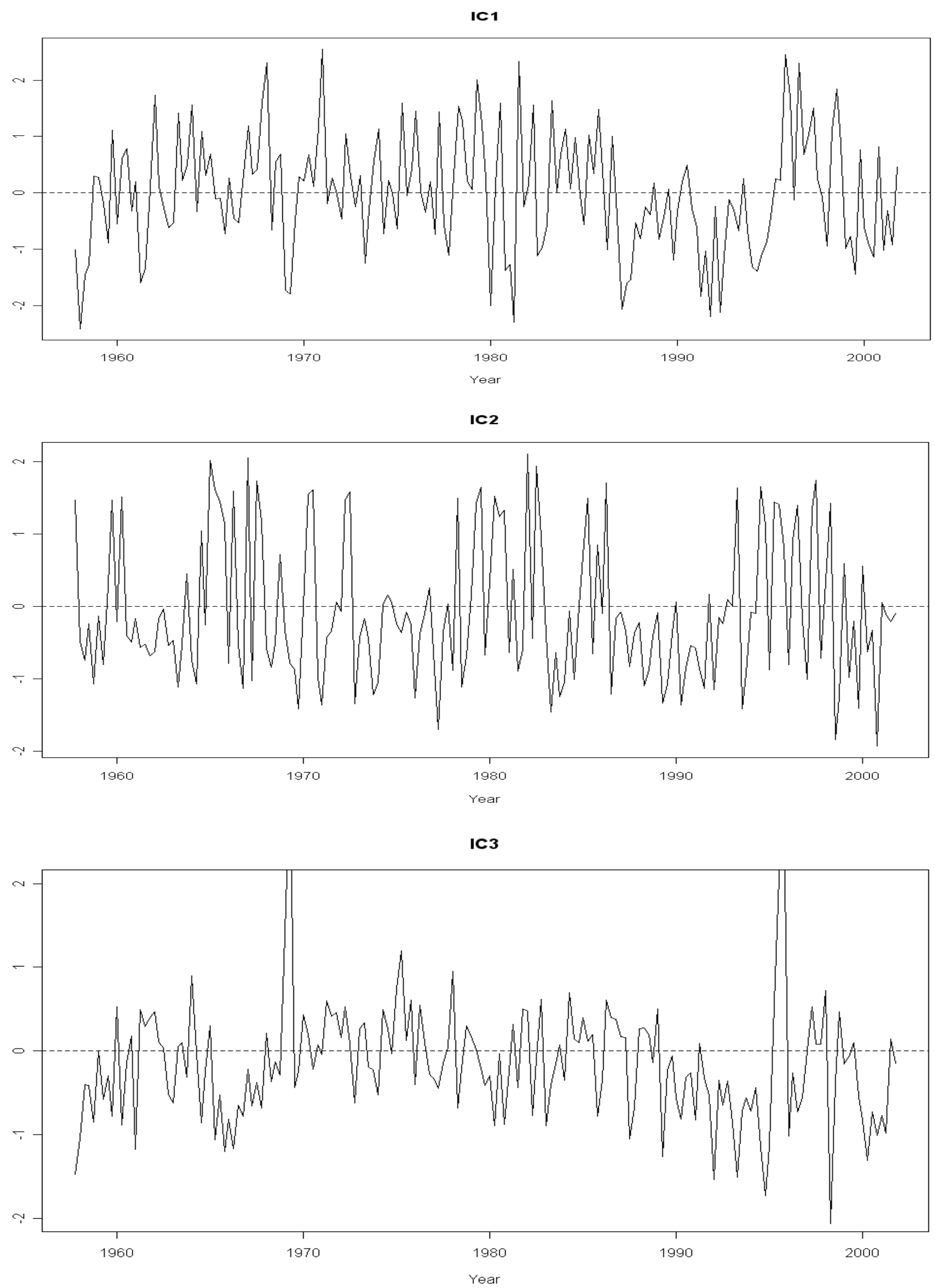


Figure 6.4. Three independent component of the five river flow, together they retain 98% of total variance.

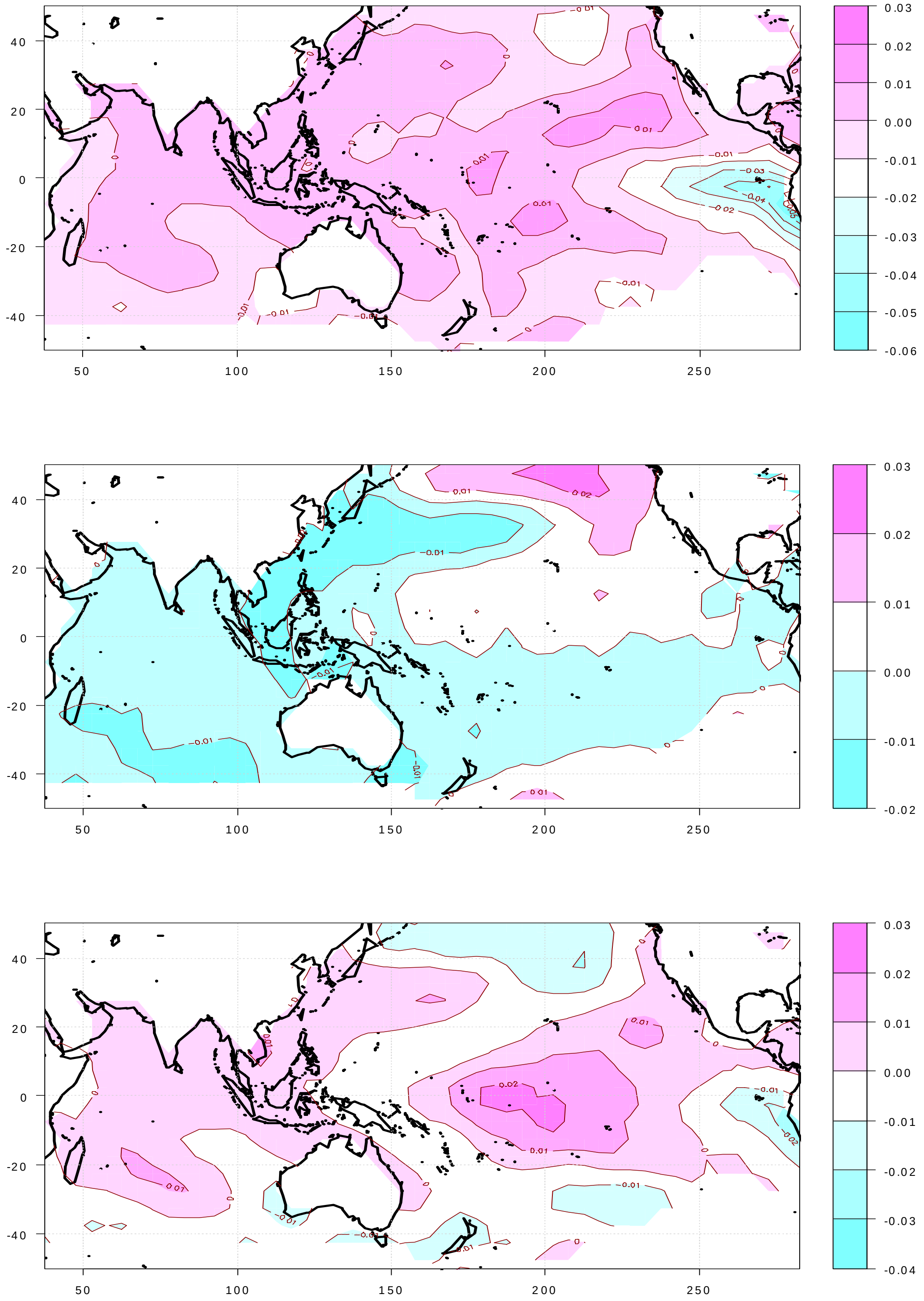


Figure 6.5 Loadings of three independent components used as predictors in the IC forecast model.

6.3.5 Modified nearest neighbour sampling (KNM)

The non-parametric method allows modelling with fewer assumptions on the nature of dependence or the nature of the probability distribution of the response series to be modelled, and hence a suitable alternative to a parametric or semi-parametric regression model (GLM and ICM here) to forecast the multi-site seasonal flow vector. The nonlinearity of the underlying dynamics associated with hydrological processes and the availability of large data sets to develop the models from, favour nonparametric alternatives over corresponding parametric equivalents in regression and simulation modelling studies. Nearest neighbour methods use the similarity (neighbourhood) between current observations of predictors (SSTA field in this study) and similar sets of historical observations to obtain the best estimate for a dependent variable (multi-site flow) (Karlsson and Yakowitz 1987, Lall and Sharma 1996). We have applied a weighted nearest neighbour approach (Souza Filho and Lall 2003, Mehrotra and Sharma 2006) in formulating the forecasting model reported here. We refer this K nearest neighbour modelling approach as KNM in this paper. The KNM approach aims to ascertain the conditional dependence of multi-site seasonal flow $\mathbf{y}_t = \{y_{l,t}\ ,\ l=1,2,\ \ldots L\}$ on a weighted set of predictors by identifying K nearest neighbours in the historical record. Identification of the nearest neighbours proceeds by ranking historical responses using a modified squared Euclidean distance (ξ) metric:

$$\xi_\tau^{(t)} = \sum_p \beta_p \,(x_{p,\tau} - x_{p,t})^2 \qquad [7]$$

where, $x_{p,\tau}$:the scaled p^{th} predictor at a past time τ,

p :1, 2, … index of multiple predictor vectors,

τ : t-1, t-2, … index of past time,

β_p :the influence load to p^{th} predictor vector. This load can be approximated as the coefficients of the linear relationship of $(Y_t \sim \mathbf{x}_t)$ where $\mathbf{x}_t=\{x_{p,t};\ p=1,2,..\}$ and $Y_t = \sum_l y_{l,t}$

The response time series $\{\mathbf{y}_\tau;\ \tau = t\text{-}1,\ t\text{-}2,\ t\text{-}3\ \ldots\}$ is ranked based on the order of the current $\xi_\tau^{(t)}$. If k_τ is the sorted rank of $\mathbf{y}_\tau$ then $k_\tau \in \{1,\ 2,\ 3,\ \ldots K,\ \infty\}$, where K is the farthest neighbour considered for ascertaining the prediction in the KNM approach. We recommend K equals the nearest integer of $\sqrt{T}$ (Lall and Sharma 1996). A forecast is then expressed as an expected value of the conditional probability distribution formed based on the nearest neighbours. The probability of re-sampling a past observation at time t is then specified as follows (Lall and Sharma 1996):

$$Pr(\hat{\mathbf{y}}_t = \mathbf{y}_\tau \mid \mathbf{X}) = k_\tau^{-1} / (1 + 2^{-1} + 3^{-1} \ldots + K^{-1}) \qquad [8]$$

where $\mathbf{X}$ is the multiple predictor vector $\{x_{p,\tau}\}$. Note that flows at all sites are sampled together from past observation $\mathbf{y}_\tau = \{y_{l,\tau}.; l=1,2\ldots L\}$. This concurrent multisite sampling reflects historical spatial dependence. The expected value of the forecast by KNM method is $\hat{y}_{l,t}^{(K)}$.

This case study selected KNM model predictors in few steps. First the total flow time series (Y_t) is scaled by removing the mean and normalising the variance to one. Alternatively, variance stabilising transformation like taking logarithm can also be used. The seasonal mean and the lag one scaled flow forms first two predictor vectors. Then we globally explored the spread of correlation and mutual information scores (Sharma 2000a) (see Appendix for mutual information formulation) of Y_t to the SSTA field. Seven SSTA zones with high dependence scores are selected as potential predictors. A partial F test of linear association of Y_t to the seven zones retained four zones (see Figure 6.6). Hence $\mathbf{x}_t$ = {seasonal mean flow, Y_{t-1}, average SSTA zones at Pacific Ocean north, central, south and Indian Ocean north}. The influence loads (β_p) of the predictors are {0.55, 0.07, 0.08, 0.08, 0.11, 0.11}. This approach of using zone averaged SSTA is not new in hydrological studies. (Maity, R. and Kumar, D. N., 2009: Sharma, A., Luk, K.C., Cordery, I., Lall, U., 2000b: Verdon, D. C. and Franks, S. W., 2005: Verdon, D. C. *et al.*, 2004).

6.3.6 Dynamic combination of forecasts

The three component forecasts GLM, ICM and KNM are dynamically combined to reduce the structural uncertainty of any single forecast. It should be noted that the approach presented assumes that the component models being used are pre-defined and non-alterable. This is especially so when the component models are conceptual water balance models that have been developed by different groups and agencies, and hence are difficult to modify at each time-step of the simulation.

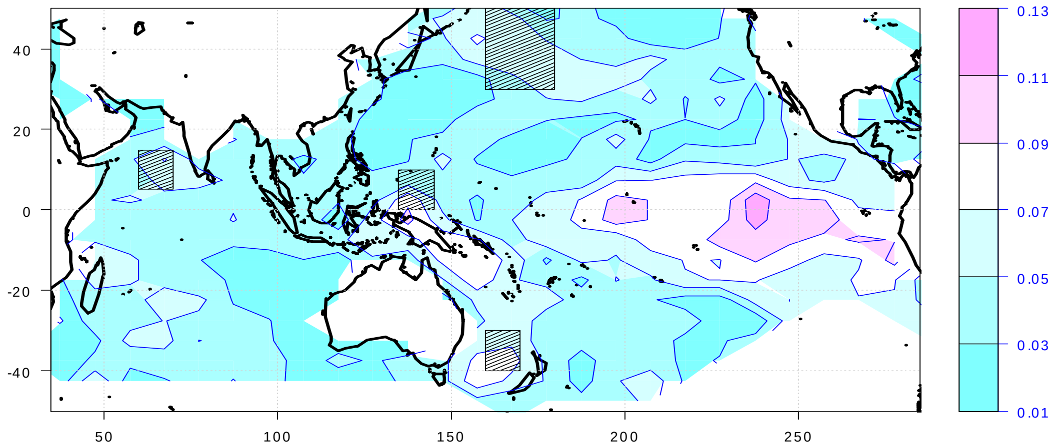


Figure 6.6. Predictors of the KNM model as the mean seasonal SSTA at 4 zones shown as hatched boxes in this graph. The mutual information score of total flow volumes to mean seasonal SSTA is shown in the background.

The pair wise dynamic weight requires pre sorting the model pairs and the design of a hierarchical tree structure. In case of a univariate response, Chowdhury and Sharma (2009a) recommended models to be paired such that the covariance of the paired responses was the minimum across all possible pairs. The guideline to optimal design of the hierarchical tree for multivariate response remains a potential future research topic. In this case study, the residual co variances are all of similar magnitudes. Hence any variation of the tree architecture has a minor influence on the final result. After a few evaluations we decided on the tree structure as shown in Figure 6.7. For simplicity, the first level of combination used static weight with dynamic combination assigned to the highest level, an approach similar to multivariate SSTA combination of Chowdhury and Sharma (2009b). Accordingly, KNM is combined to GLM using static weights. The static weight assigned to KNM is the ratio of precision of KNM forecast to the sum of precision of KNM and GLM (Granger and Newbold 1977, McLeod et al. 1987). Then the KNM+GLM are dynamically combined to ICM. Note that the tree structure keeps the number of weights to two, same number as a static combination of three forecasts.

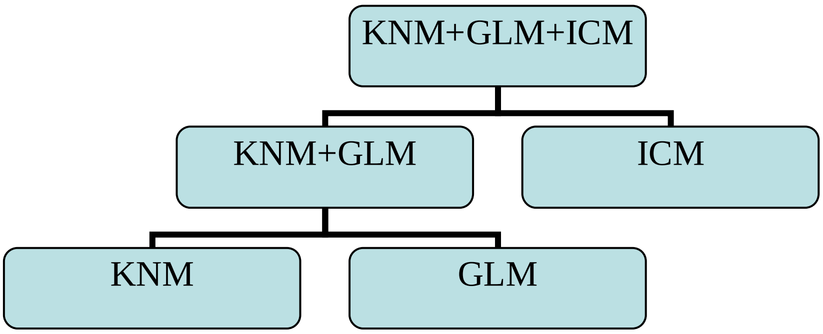
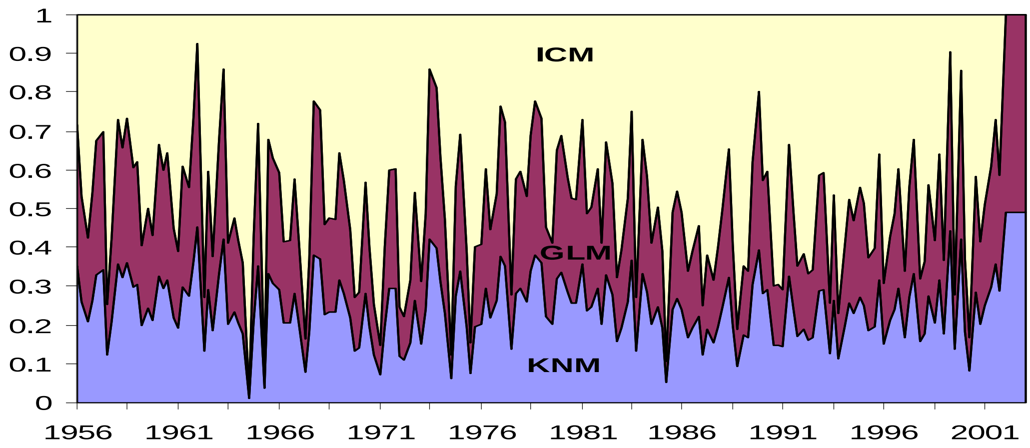


Figure 6.7. Pair wise forecast combination tree.



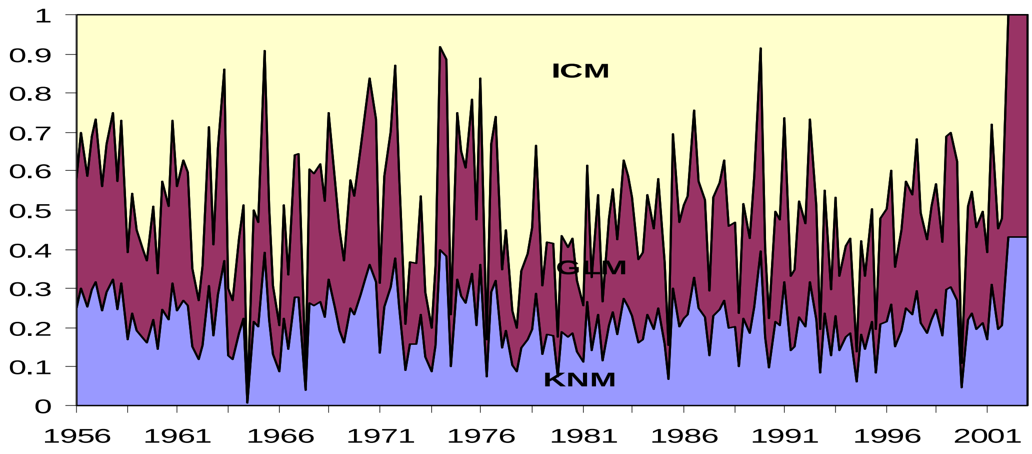
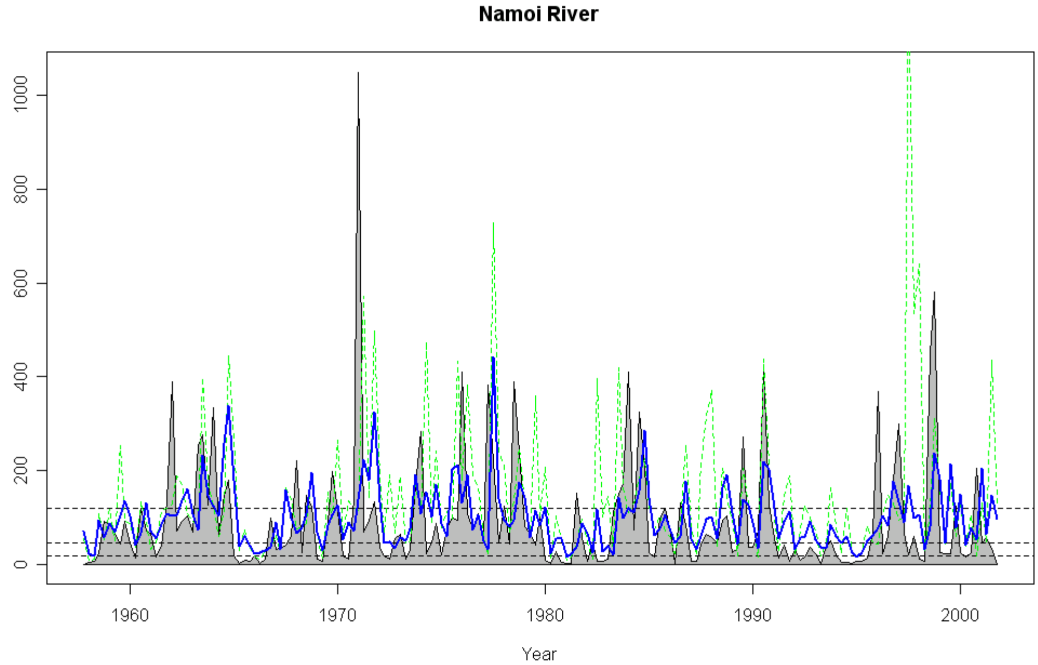


Figure 6.8. The stacked time series of weights corresponding to three models. The forecast time series are extracted from ±6 month cross validation results. The Namoi River is on top and the Coxs Creek is the figure below.

The predictors of the time variant weights are $\{\omega_{t-1}$; $PC1^-$ and $PC1^+\}$, where PC1 represents the first principal component of the Indian and Pacific Ocean SSTA field, and $PC1^+$ and $PC1^-$ their respective positive and negative components. The weights imply that the comparative forecast strength of next season of any single model is reflected by this season's performance and the current state of climate. Figure 6.8 shows the observed weight time series and the fitted weight prediction. The forecast flow time series is drawn against the observation in Figure 6.9. Note that only the time series of the wettest and the driest rivers are shown in this manuscript for sake of brevity.



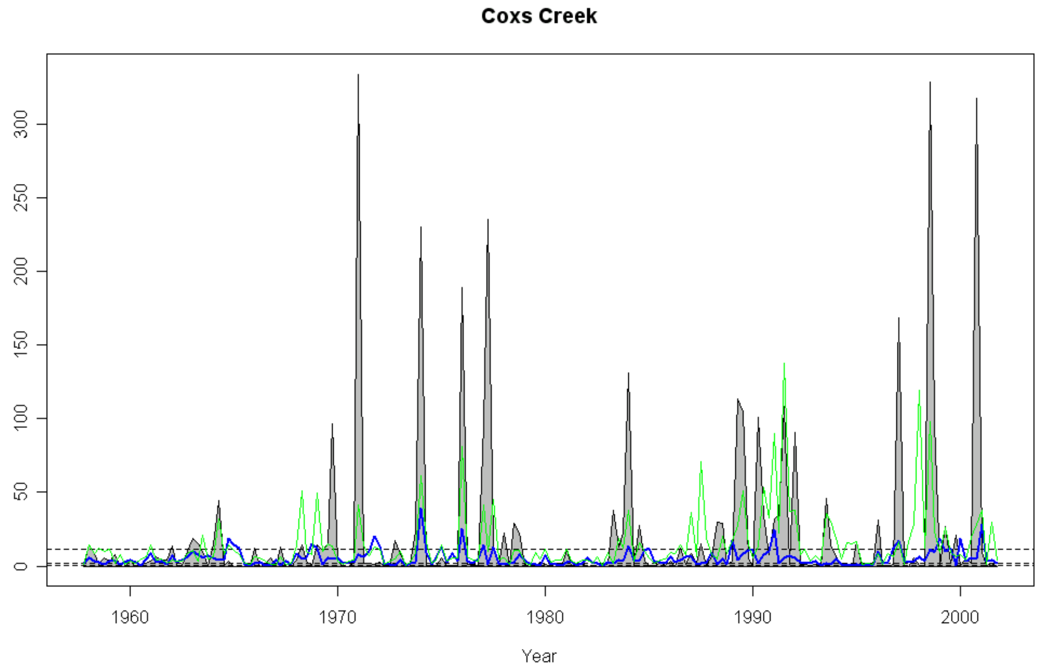


Figure 6.9. Forecast of the next season's flow total (in 10^6 m^3) is drawn against the actual observation (in black shade). The blue solid line is the combined forecast and the green broken line is the single model (GLM) forecast. The forecast time series are extracted from ±6 month cross validation results. The horizontal broken lines are showing first, median and third quarter of the 100 year flow records. The Namoi River is on top and the Coxs Creek is the figure below.

6.4 Results and Discussion

6.4.1 Model calibration and validation

All three models (GLM, ICM, KNM) forecast the total flow volume of the next season at the end of current season. For example the autumn flow volume is forecasted using summer flow volume and the autumn SSTA forecast. The results presented here represent concurrent model forecasts using three SSTA fields. The first field is the observed SSTA used to calibrate the model for a period of 1893 to 2003 (excluding a validation window). The second set is the MetF SSTA forecast. The third set is the DW SSTA forecast derived in this research by dynamically combining SSTA forecasts from two additional models to MetF as mentioned in Section 6.2.3. Predictors derived from the second and third sets contain forecast error. Our decision of using observed SSTA to calibrate the component models avoids biased parameter estimates due to error in the SSTA forecasts (Kavetski et al. 2002, Huard and Mailhot 2006, Chowdhury and Sharma 2007). MetF and DW forecasts are used to validate (out of sample) the performance of the calibrated model. We used a leave ±2 seasons out cross validation to assess our results. For example, the winter 1980 validation uses the parameters calibrated by removing the full 1980 data set. The validation period spans from 1958 to 2001, decided by the length of SSTA forecasts. Note that the long data set allows us to carefully choose the predictor space that remains static while moving the validation window along the time series (Mason 2008). Assuming minimal artificial skill beyond one year, the validation may be a reflection of true forecast skill instead of calibration accuracy.

We have chosen three component flow forecast models: KNM, GLM and ICM. There are four alternative sources of predictors (SSTA forecasts) to these models as described in Section 6.2.3. They are MetF, CPC, ECM and DW (or. MetF+CPC+ECM). The first validation trial used predictors derived from the best SSTA forecast model MetF, out of available three single SSTA forecasts. Note that the absence of any non SSTA flow predictor implies that flow forecast skill depends strongly on SSTA forecast accuracy. The MetF forecast exhibits the smallest forecast error variance compared to that of CPC and ECM. Consequently, MetF is used to illustrate the performance of the model that would be attained if a single SSTA forecast approach were used instead of the dynamic combination approach. The second validation trial is based on predictors derived from combined SSTA forecast: DW. Finally we investigated incremental skill post combination of the three flow forecast models where the predictors came from DW SSTA forecast.

Model combination, due to extra parameters, reduces degrees of freedom. The method relies on due diligence of the modellers in maintaining parsimony, using minimal parameters and testing extensive validation. We chose not to present the calibration results as this does not reflect the true predictive errors. Validation results are mainly unaffected by model complexity and hence better represent predictive performance. Accordingly, it is emphasised that the next two sections analysed validation results only.

6.4.2 Measures of forecast strength

One of the most commonly used performance criteria is the mean of squared errors (MSE). MSE is a measure of squared deviation of forecast from the observation. Note that the mathematical formulations of the error measures discussed in this section are included in the appendix. The MSE can also be expressed as the summation of systematic error (squared-bias) and random error (variance) (Vazquez, 2003, Hastie and Tibshirani 2000). The MSE for all the cases evaluated are presented in Table 6.2. The KNM model achieved higher reductions in MSE (25-30%), by using DW instead of MetF SSTA forecast, due to its higher initial error base. In general, the ICM came out to be the best model (based on the MSE measure). While forecast combination further reduced MSE, the reduction is minor and inconsistent across the rivers.

The MSE is a poor indicator of a models ability to forecast both high and low flow events, with the squared residual terms tending to enhance the significance of the larger errors which are usually associated with high flow events. The research on evaluating forecast performance by Armstrong (2001, pp. 460) concluded against using MSE in comparative evaluation. The MSE is aimed at cases where residuals are randomly distributed with zero mean and constant variance which is unlikely in case of seasonal flow of arid river. There are few ways to address the MSE bias caused by flood outliers and boundary values such as sequences of low flow or zeros. One method is to stabilise the variance is by a prior transformation of flow using a logarithmic or a Box-Cox transformation (Box and Cox 1964). Accordingly, a log transformation and standardisation was performed on the data to allow an improved characterisation of the associated errors. The entire flow series matrix was increased by one unit (10^3 m^3), prior to transformation, in order to avoid any zero flow seasons.

Table 6.2. Mean of squared errors (MSE) of the flow forecast. The units are $(10^6\ m^3)^2$ per season. The last column includes forecasts of total flow.

SSTA	Models	Manila	Namoi	Peel	Mooki	Coxs	Total
MetF	KNM	253	2344	1258	1009	457	19740
	GLM	193	3763	1779	445	238	18780
	ICM	168	1959	893	663	333	13850
DW	KNM	188	1652	929	674	327	13670
	GLM	170	2435	1316	459	242	14760
	ICM	158	1594	805	768	324	12850
Combined	K+G+I	160	1488	806	610	310	12180

The MSE of the transformed flows is presented in Table 6.3. Overall, the reduction in transformed flow MSE when using the DW SSTA is evident. Due to the transformation process the errors are now independent of the river size, hence all the five rivers can be pooled together to get an overall measure of performance. The combined forecast illustrates a drop in transformed flow MSE in the pooled time series; however the reduction is not consistent across the five rivers. A comparative analysis of MSE of multiple river flows may not account for differing level of forecast difficulty of each river satisfactorily. The relative potency of the models for each river requires different measures as discussed next.

The Nash-Sutcliffe efficiency (Nash and Sutcliffe 1970) measures the improvement in forecast error variance with respect to the null model which is the sample mean. In case of seasonal flow the null model can be extended to a set of four seasonal means, a proposition after Garrik *et al.* (1978). We refer this measure as coefficient of efficiency (CE) here. The advantage of CE over MSE is that it is dimensionless. While the CE includes the variance of observations in addition to systematic and random error (Vazquez 2003), similar to the MSE, it is sensitive to large random error associated to flood peaks (Leagates and McCabe 1999, V´azquez et al. 2008). Notably, a prior transformation (as used in Table 6.3) may not be always robust in case of a flow series with a long drought sequence (series of zero bounded entries). Hence we search for measures, as an alternative of CE, which is suitable for arid river flow (time series with unstable variance).

Table 6.3. Error variance of the transformed flow forecasts. The transformation applied here is the standardised log of non zero flows. The final column shows the analysis where all the rivers are pooled together.

SSTA	Models	Manila	Namoi	Peel	Mooki	Coxs	Pooled
MetF	KNM	1.04	1.06	1.14	1.09	1.04	1.07
	GLM	0.83	0.84	1.03	0.70	0.72	0.83
	ICM	0.80	0.73	0.81	0.91	1.05	0.87
DW	KNM	1.07	0.92	1.02	1.03	1.05	1.03
	GLM	0.77	0.79	0.99	0.70	0.71	0.79
	ICM	0.77	0.67	0.75	0.89	0.93	0.82
Combined	K+G+I	0.78	0.67	0.78	0.80	0.86	0.78

The relative absolute error (RAE), which is the ratio of absolute residual error to that of null model error, serves as a useful alternative to CE. The RAE returns 1 or a higher value for no model skill to 0 for a perfect forecast. In order to avoid outliers due to near zero null error and rivers with long dry seasons, either a Winsorised or a median value of RAE (MdRAE) is recommended. Winsorising trims time series at an upper and lower boundary (Wu 2006, Jose and Winkler 2008). Winsorising relies on selection of appropriate boundary limit, requiring external deliberation. For simplicity we chose the ultimate trimming by accepting MdRAE as presented in Table 6.4.

Table 6.4. Median relative absolute error (MdRAE) of the flow forecast. The final column shows the analysis where all the rivers are pooled together.

SSTA	Models	Manila	Namoi	Peel	Mooki	Coxs	Pooled
MetF	KNM	0.92	1.14	1.13	1.09	1.09	1.07
	GLM	0.87	1.03	1.19	0.38	0.44	0.77
	ICM	0.90	1.08	0.98	0.40	0.51	0.81
DW	KNM	0.90	1.04	1.01	1.00	0.92	0.98
	GLM	0.69	0.97	1.04	0.34	0.40	0.66
	ICM	0.81	0.93	0.94	0.56	0.22	0.78
Combined	K+G+I	0.74	0.92	0.95	0.49	0.30	0.74

New insights to the forecast strength of the two driest rivers (Mooki and Coxs Rivers) are evident from MdRAE, as they are most prone to outliers in MSE based measure. The MdRAE measures of Mooki and Coxs rivers are 0.5 and 0.3, which is lower (hence superior skill) than the other three rivers. The combined flow forecast yields a narrower range of MdRAE across all the five rivers (0.30 to 0.95) than either GLM (0.22 to 0.94) or ICM (0.34 to 1.04). It is important to note that if the combined model results are compared with any one of the models on an individual basis, the combined model outperforms the other model on a majority of the rivers analysed. Similar conclusion can be reached using other statistics (not shown here) such as the linear error in probability space skill score (Potts et al. 1996).

In summary, the improvement in predictor field by using DW rather than MetF SSTA forecast advances the flow forecast accuracy. The performance of single forecast model ICM and GLM are similar while KNM is a weaker alternative. There is an overall reduction of errors after combining the three forecast models. However the incremental reduction in error due to the combination is small and not consistent across all the five rivers. Note that these remarks are based on univariate measures only. These measures do not consider spatial dependence of the rivers.

6.4.3 Representation of spatial dependence

One of the key requirements for forecasting a multivariate response is the representation of spatial dependence across the river system. A Pearson's correlation can be used as a simple measure of spatial dependence. The correlation of the raw flows of the arid river system would reflect linear dependence among larger rivers during major flooding period only. Hence, the flow time series is transformed using a log transformation (after adding a unit volume to avoid zeroes) and standardised to reduce the influence of outliers. The correlations of each transformed river flow against the four other rivers are estimated next. Accordingly an observed vector of 10 correlation statistic is established. Figure 6.10 shows the comparative graph of observed versus modelled correlation values. For clarity we only presented the set derived using DW SSTA forecast. The figure provides some interesting insights into the differing capabilities of each model. The simple GLM structure is able to model the correlation in six out of the ten cases. The structure of GLM is weak in reproducing low correlation or interdependence between dry rivers. On the other hand linear regression tends to inflate dependency between the larger rivers. KNM and ICM structures are primarily aimed at forecasting multivariate response variable. The KNM model maintained the inter site correlation better than GLM, however there is an apparent bias of underestimation. Conversely ICM returns a pattern of overestimated correlation. The dynamically combined flow forecasts illustrate superior replication of inter site correlation reducing the scatter present in the GLM estimates and mitigating the bias in KNM and ICM.

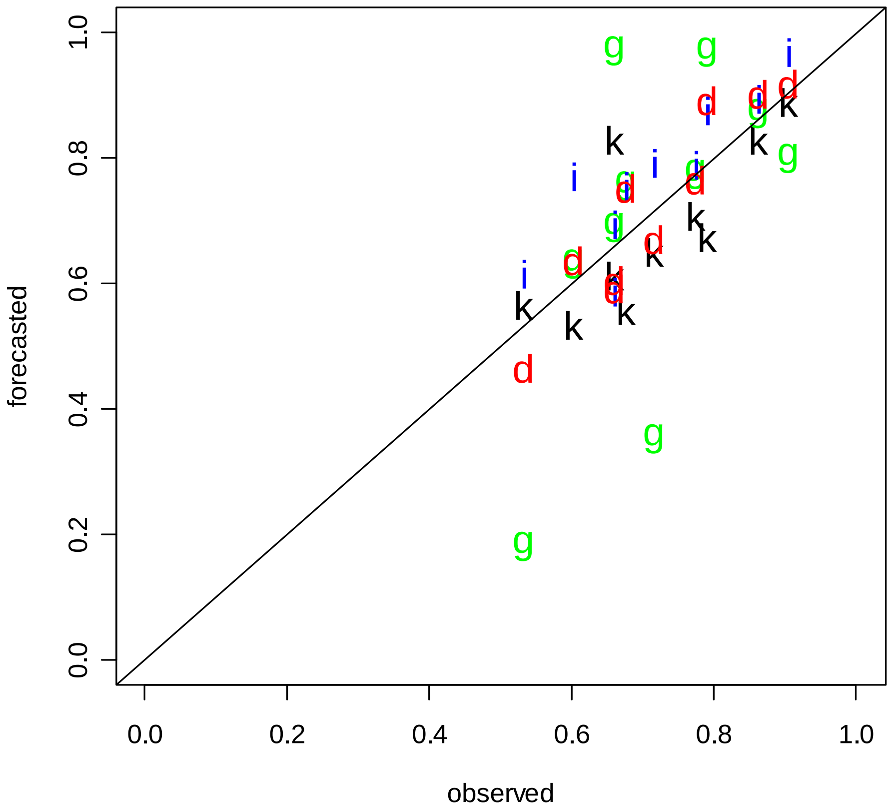


Figure 6.10. Spatial dependence presented as paired correlation of log flow of two rivers. Component models are KNM (k), GLM (g), ICM (i) and the combined model legend is (d).

6.5 Discussion

Combining flow forecast in this case study yields improvement as compared to forecasts where no combination was performed. However, the improvements in the combined flow forecasts are smaller compared to what was achieved through a similar combination in order to arrive at a multivariate SSTA forecast. What factors determine the potential and extent of enhancement possible as a result of combination? There are two major prior indicators of likely improvement post combination. First and foremost, the component model that has the lowest associated residual error variance is likely to influence the final model combination the most. Theoretically, if we have a model with a zero error variance, the combined model should effectively collapse to this. Secondly, the improvement in each combination (at all levels of the hierarchical tree) depends on the covariance of the errors between the two combined models. If this covariance is small in comparison to either of the model error variance, mixing the two models is likely to result in greater improvements. An inspection of SSTA forecasts in Figure 6.11 for the Central North Pacific shows that the minimum covariance is only 43% of the best forecast variance. On the other hand, the covariance of flow forecasts is a high 86% of the best forecast variance. The high covariance ratio indicates that the three different flow models (each having their own distinct predictor set) are unable to contribute enough "new" information so as to result in better combined forecasts. Note that we intentionally attempted to exert independence and dissimilarity among different flow forecasting models. However, the resulting weak independence in the forecast despite the use of different model structures and predictors may be attributed to the fact that the unexplained variances of all component forecasts are relatively high. The high unexplained variance masks any independence across the forecasts.

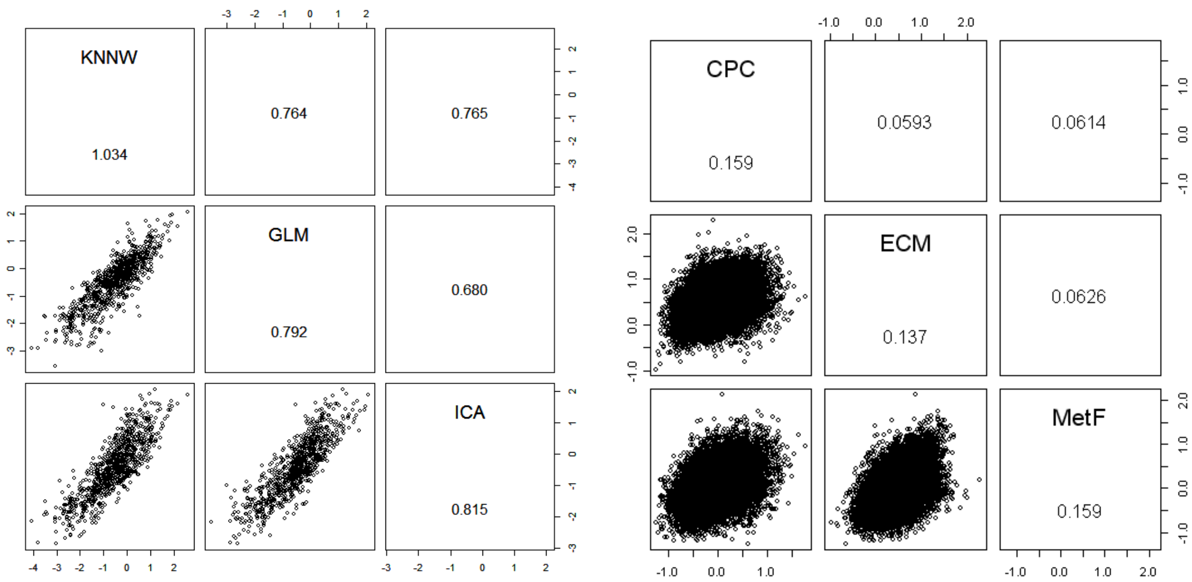


Figure 6.11. Residual errors of the three component forecasts with the corresponding variance and covariance measures listed in the diagonal and

upper triangle boxes. The left panel contains residual error of log flow, the right panel shows residual errors of sea surface temperature within the 175°-225°E by 10°-20°N box.

A comparative analysis of the relative merits of individual predictor variables in the three forecast schemes (GLM, ICM, KNM) is complicated due to prior transformations and different modelling hierarchies. As a result, our comparison (Table 6.5) is based on a backward removal of each predictor from the full model and then assessing the overall increase in forecast error variance. The approach is similar to computing chi square statistic by removing a predictor and estimating the increase in residual deviance while analysing generalised linear models. In general SSTA predictors did not aid prediction skill significantly and as a result narrows any difference between the three forecast schemes. Another reason of lack of independence may be the similarity between the SSTA principle components and independent components. A closer inspection reveals that the zones of the KNM predictors coincide with regions having high loadings in both ICs and PCs. Future studies should explore inclusion of different predictor sources such as wind speed and geo potential heights in forming the models to be combined.

Table 6.5. The percentage increase of forecast error variance after backward removal of a predictor. The forecast is issued at the end of last season for this season. The variance is computed after log transformation of the flows.

GLM Predictors	Manilla	Namoi	Peel	Mooki	Coxs
Total flow: last season	21	23	20	11	4
Tot flow: this season, last year	8	15	14	2	2
SSTA PC1: this season	8	14	13	2	2
SSTA PC2: this season	12	18	15	3	2
Namoi flow fraction: last season	8	23	18	7	15

ICM Predictors	Manilla	Namoi	Peel	Mooki	Coxs
Total flow: last season	20	26	15	19	2
Mean total flow: this season	8	8	7	7	6
SSTA IC1: this season	8	8	7	4	5
SSTA IC2: this season:	5	5	5	19	5
SSTA IC3: this season:	12	11	13	12	16
KNM Predictors	**Manilla**	**Namoi**	**Peel**	**Mooki**	**Coxs**
Total flow: last season	19	14	4	5	4
Mean total flow: this season	4	11	19	4	7
SSTA Zone N Pacific: this season	16	9	14	2	2
SSTA Zone C Pacific: this season	6	13	14	2	6
SSTA Zone S Pacific: this season	12	10	20	15	12
SSTA Zone I Ocean: this season	11	12	15	7	4

It is important to note that all three forecasting schemes used in this paper are formulated on statistical platform. Statical models are forced into a relationship with the response to maximize the performance returning higher covariance among the forecasts. One way to address this is to reduce the forecast (such as using PCs or ICs) and then combing the reduced forecast only. Similarly, the inclusion of the physics based forecasting platform (dynamical model) has the potential of further improving the combined forecast skill.

In addition to the results reported above, we have further scrutinised the performance of a static weight combination for the three forecasts. The static weights are derived as the ratio of the precision (inverse of variance) of each forecast to the summation of all three precisions (Kim et al. 2006). As expected, the static weight did not offer any overall improvement over a dynamic weight combination. We have not included the analysis here for the sake of brevity. Interested readers are referred to our earlier publications on the dynamic combination logic (Chowdhury and Sharma 2009a, 2009b) for a discussion of the improvements obtained over the static combination in the context of SSTA forecasting. It is relevant to state that the derivation of static weight often involves the objective function that minimises squared error of combined forecast. Such objective function may artificially dampen forecast error variance without necessarily enhancing any forecast skill (Zhang and Casey 2000). In contrast, the analytical derivation of dynamic weight (Equation [2]) does not force minimisation of squared error and hence any resulting improvement is more reliable.

6.6 Conclusion

This study presented a rationale for combining multiple model forecasts using a dynamic combination rationale, an approach that has been documented to improve forecast accuracy in the context of forecasting globally distributed SSTA fields forward in time. It used a dynamically combined SSTA forecast field as the basis of deriving concurrent flow forecasts using three alternate approaches, and proceeded to assess the improvements that result when these three flow forecasts were dynamically combined. The three forecasting methodologies were applied to forecasts inflows at five locations in the upper Namoi Catchment in Eastern Australia, the approaches being: (1) a mixture of generalised log normal and multinomial models, named GLM; (2) the non parametric nearest neighbour method KNM; and (3) the ICM scheme, based on a local regression of the independent components of the five inflows. Notable to each of these approaches was the different predictor base used for each, these predictors being derived from various transformations of the dynamically combined forecasts of the Indian and Pacific Ocean SSTA.

Improvements in flow forecast were sought in two stages. First, the SSTA forecasts were improved by dynamically combining three alternative SSTA forecasts. It was demonstrated that the improved SSTA forecast in turn improves all three flow forecasts scheme. In terms of a univariate assessment of forecast accuracy, the GLM and ICM exhibited an overall superior forecast skill to that of the KNM. Spatial dependence, however, was better replicated by KNM and ICM compared to that from the GLM. Secondly, a dynamic combination of the GLM, KNM and ICM forecast was performed as an added step to reduce forecast structural uncertainty. This dynamic combination was found to result in a small though noticeable improvement in terms of the overall flow forecast; however this improvement was not consistent across all five rivers. The combined flow forecast exhibited superior representation of spatial dependence than any of the single forecasts. Our reasoning for the improvements in flow forecasts being smaller in proportion to the improvements noted in case of the dynamically combined SSTA forecasts lies with the strong dependence that exists across the three flow forecasts. Lack of such dependence across the three SSTA forecasting models resulted in the combination being significantly superior to any of the individual forecasting models.

6.7 Acknowledgement

This research was funded by the Australian Research Council and the Sydney Catchment Authority. We acknowledge the NSW Department of Water and Energy for the natural river flow data simulated using Namoi IQQM. The catchment is drawn by Don Stazic. The computation was performed using the freely available R statistical computing platform (R Development Core Team 2008). All SSTA time series were downloaded from the data library of International Research Institute for Climate and Society, New York.

6.8 Appendix

MSE: Mean squared error;

CE: Coefficient of efficiency;

MdRAE: Median relative absolute error;

MI: mutual information score.

y_t : observed response at time *t*

$\hat{y}_t$: forecasted response at time *t*

$\bar{y}$: mean observed response

s: season where, *s*=1,2,3 and 4.

t: time index where *t*=1,2,3,4,5, ….*T*;

f(.) =density.

f(.;.)=joined density.

i = 1,2,..*N*, ordinate of discrete points of density estimation.

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

$$CE = 1 - \frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{T}(y_t - \bar{y}_s)^2}$$

$$MdRAE = Median\left\{\frac{|y_t - \hat{y}_t|}{|y_t - \bar{y}_s|}\right\}_{t=1}^{T}$$

$$MI = \frac{1}{N}\sum_{i=1}^{N}\log\frac{f(y_i;\hat{y}_i)}{f(y_i)\,f(\hat{y}_i)}$$

6.9 Reference

Agresti, A. 1996. An Introduction to Categorical Data Analysis, New York.

Ajami, N. W., Q. Duan, X. Gao, and S. Sorooshian. 2006. Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. Journal of Hydrometeorology **7**:755-768.

Anderson, P. L., and M. M. Meerschaert. 1998. Modeling river flows with heavy tails. Water Resources Research **34**:2271-2280.

Araghinejad, S., D. H. Burn, and M. Karamouz. 2006. Long-lead probabilistic forecasting of streamflow using ocean atmospheric and hydrological predictors. Water Resources Research **42**:W03431.

Armstrong, J. S. 2001. Combining forecasts. Pages 417-439 *in* J. S. Armstrong, editor. Principles of Forecasting: A Handbook for Researchers and Practitioners. Kluwer Academic Publishers, Norwell.

Augustin, N. H., L. Beevers, and W. T. Sloan. 2008. Predicting river flows for future climates using an autoregressive multinomial logit model. Water Resources Research **44**:W07403.

Box, G. E. P., and D. R. Cox. 1964. An Analysis of Transformations. Journal of Royal Statistical Society (Series B) **26**:211-252.

Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen. 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. Journal of Hydrology **298**:242-266.

Cardoso, A. O., R. T. Clarke, and P. L. S. Dias. 2005. A case study of the sea surface temperatures (SSTs) to obtain predictors of river flow. Pages 231-238 *in* Regional Hydrological Impacts of Climatic Variability and Change - Imapact Assessment and Decision Making. IAHS, Foz do Iguacu, Brazil.

Cardoso, A. O., and P. L. S. Dias. 2006. The relationship between ENSO and Paran´a River flow. Advances in Geosciences **6**:189-193.

Chambers, J. M. 1992. Linear Models. Pages 95-144 *in* J. M. Chambers and T. Hastie, editors. Statistical Models in S. Wadsworth & Brooks, Cole.

Chandler, R. E. 2005. On the use of generalized linear models for interpreting climate variability. Environmetrics **16**:699-715.

Chiew, F. H. A., T. C. Piechota, J. A. Dracup, and T. A. McMahon. 1998. El Nino Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. Journal of Hydrology **204**:138-149.

Chiew, F. H. S., and T. A. McMahon. 2002. Global ENSO–streamflow teleconnection, streamflow forecasting and interannual variability. Hydrological Sciences **47**:505-522.

Chiew, F. H. S., S. L. Zhou, and T. A. McMahon. 2003. Use of seasonal streamflow forecasts in water resources management. Journal of Hydrology **270**:135-144.

Chowdhury, S., and A. Sharma. 2007. Mitigating Parameter Bias in Hydrological Modelling due to Uncertainty in Covariates. Journal of Hydrology **340**:197-204.

Chowdhury, S., and A. Sharma. 2009a. Long Range NINO3.4 Predictions Using Pair Wise Dynamic Combinations of Multiple Models. Journal of Climate **22**:793-805.

Chowdhury, S., and A. Sharma. 2009b. Global sea surface temperature forecasts using a pair wise dynamic combination approach. Journal of Climate **under review**.

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting **5**:559-583.

Cleveland, W. S. 1979. Robust locally weighted regression and smoothing scatterplots. Journal of American Statistical Association **74**:829-836.

Cleveland, W. S., S. J. Devlin, and E. Grosse. 1988. Regression by Local Fiting Methods, Properties, and Computational Algorithms. Journal of Econometrics **37**:87-114.

Cleveland, W. S., and E. Grosse. 1991. Computational methods for local regressions. Statistics and Computing **1**:47-62.

Common, P. 1991. Independent Component Analysis, A New Concept? Signal Processing **24**:1-10.

Crapper, P. F., S. G. Beavis, and L. Zhang. 1999. The relationship between climate and streamflow in the Namoi Basin. Environment International **25**:827-839.

CSIRO. 2007. Water availability in the Namoi. A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project. CSIRO, Australia.

Devineni, N., A.Sankarasubramanian, and S. Ghosh. 2008. Multimodel Ensembles of Streamflow Forecasts: Role of Predictor State in Developing Optimal Combinations. Water Resources Research **44**:W09404.

Drosdowsky, W., and L. E. Chambers. 2001a. Near-Global Sea Surface Temperature Anomalies as Predictors of Australian Seasonal Rainfall. Journal of Climate **14**:1677-1687.

Dutta, S. C., J. W. Ritchie, D. M. Freebairn, and G. Y. Abawi. 2006. Rainfall and streamflow response to El Niño Southern Oscillation: a case study in a semiarid catchment, Australia. Hydrological Sciences **51**:1006-1020.

ECMWF. 2004. Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER). 434, England.

Garrick, M., C. Cunnane, and J. E. Nash. 1978. A criterion of efficiency for rainfall runoff modelling. Journal of Hydrology **36**:375-381.

Georgakakos, K. P., D.-J. Seo, H. Gupta, J. Schaake, and M. B. Butts. 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. Journal of Hydrology **298**:222-241.

Girolami, M., and C. Fyfe. 1996. Negentropy and Kurtrosis as projection persuit indices provide generalized ICA algorithms. *in* Advances in Neural Information Processing Systems Workshops: Blind Signal Processing, Snowmaas, CO.

Goswami, M., and K. M. O'Connor. 2007. Real-time flow forecasting in the absence of quantitative precipitation forecasts: A multi-model approach. Journal of Hydrology **334**:125-140.

Granger, C. W. J., and P. Newbold. 1977. Forecasting economic time series. Academic Press, New York.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona. 2005. A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. Water Resources Research **41**:W10410.

Guetter, A., and K. Georgakakos. 1996. Are the El Nino and La Nina Predictors of the Iowa River Seasonal Flow? Journal of Applied Meteorology **35**:690-705.

Hameed, T., and G. Podger. 2001. Use of IQQM simulation model for planning and management of a regulated system. Pages 83-89 *in* IAHS Red Book 2001.

Hamlet, A. F., and D. P. Lettenmaier. 1999. Columbia River streamflow forecasting based on ENSO and PDO climate signals. Journal of Water Resources Planning and Management-Asce **125**:333-341.

Hastie, T., and D. Pregibon. 1992. Generalised Linear Models. Pages 195-248 *in* J. M. Chambers and T. Hastie, editors. Statistical Models in S. Wadsworth & Brooks, Cole.

Hastie, T., R. Tibshirani, and J. Friedman. 2000. The Elements of Statistical Learning, Data Mining, Inference and Prediction. Springer Series in Statistics.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T. 1999. Bayesian Model Averaging: A Tutorial (with discussion). Statistical Science **14**:382-417.

Hsieh, W. W., Yuval, J. Li, A. Shabbar, and S. Smith. 2003. Seasonal Prediction with Error Estimation of Columbia River Streamflow in British Columbia. Journal of Water Resources Planning and Management-Asce **129**:146-149.

Huard, D., and A. Mailhot. 2006. A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc". Water Resources Research **42**:WR004661.

Hyvarinen, A., and E. Oja. 2000. Independent Component Analysis: Algorithms and Applications. Neural Networks **13**:411-430.

Jose, V. R. R., and R. L. Winkler. 2008. Simple robust averages of forecasts: Some empirical results. International Journal of Forecasting **24**:163-169.

Kachroo, R. K. 1992. River flow forecasting. Part 1. A discussion of the principles. Journal of Hydrology **133**:1-15.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan. 1998. Analyses of global sea surface temperature 1856-1991. Journal of Geophysical Research (C. Oceans). **103**:8,567-518,589.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal. 1997. Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. Journal of Geophysical Research (C. Oceans). **102**:27,835-827,860.

Karlsson, M., and S. Yakowitz. 1987. Nearest neighbor methods for non parametric rainfall runoff forecasting. Water Resources Research **23**:1300-1308.

Kavetski, D., S. Franks, and G. Kuczera. 2002. Confronting Input Uncertainty in Environmental Modelling. Pages 49-68 *in* H. V. Gupta, Sorooshian, S., Rousseau, A. N. and R. Turcotte, editor. Calibration of Watershed Models, Duan.

Kim, Y. O., D. Jeong, and I. H. Ko. 2006. Combining rainfall-runoff model outputs for improving ensemble streamflow prediction. Journal of Hydrologic Engineering **11**:578-588.

Lall, U., Y.-I. Moon, H.-H. Kwon, and K. Bosworth. 2006. Locally weighted polynomial regression: Parameter choice and application to forecasts of the Great Salt Lake. Water Resources Research **42**.

Lall, U., and A. Sharma. 1996. A nearest neighbor bootstrap for time series resampling. Water Resources Research **32**:679-693.

Leagates, D. R., and G. J. McCabe. 1999. Evaluating the use of 'goodness of fit' measures in hydrologic and hyroclimatic model validation. Water Resources Research **35**:233-241.

Letcher, R. A., F. H. S. Chiew, and A. J. Jakeman. 2004. An Assessment of the Value of Seasonal Forecasts in Australian Farming Systems. Pages 1511-1516 *in* C. Pahl-Wostl, S. Schmidt, and A. J. Jakeman, editors. International Environmental Modelling and Software Society (iEMSs) Conference. University of Osnabruck, Germany.

Madec, G., P. Delecluse, M. Imbard, and C. Levy. 1997. OPA Release 8, Ocean General Circulation Model Reference Manual. LODYC/IPSL, Paris.

Maity, R., and D. N. Kumar. 2009. Hydroclimatic influence of large-scale circulation on the variability of reservoir inflow. Hydrological Processes **23**:934-942.

Marshall, L., D. Nott, and A. Sharma. 2007. Towards Dynamic Catchment Modelling: A Bayesian Hierarchical Modelling Framework. Hydrological Processes **21**:847-861.

Mason, S. J. 2008. Understanding forecast verification statistics. Meteorological Applications **15**:31-40.

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models, 2nd ed. edition. Chapman & Hall, London.

McLeod, A. I., D. J. Noakes, K. W. Hipel, and R. M. Thompstone. 1987. Combining hydrologic forecast. Journal of Water Resources Planning Management **113**:29-41.

McMahon, T. A. 1979. Hydrological characteristics of arid zones. The hydrology of areas of low precipitation. Proc. Canberra symposium, December 1979, (International Association of Hydrological Sciences, Washington DC; IAHS-AISH Publication 128):105-123.

McMahon, T. A., R. M. Vogel, M. C. Peel, and G. G. S. Pegram. 2007. Global streamflows - Part 1: Characteristics of annual streamflows. Journal of Hydrology **347**:243-259.

Mehrotra, R., and A. Sharma. 2006. Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach. Advances in Water Resources **29**:987-999.

Menezes, L. M. d., D. W. Bunn, and J. W. Taylor. 2000. Review of guidelines for the use of combined forecasts. European Journal of Operational Research **120**:190-204.

Muluye, G. Y., and P. Coulibaly. 2007. Seasonal reservoir inflow forecasting with low-frequency climatic indices: a comparison of data-driven methods. Hydrological Sciences **52**:508-522.

Nash, J. E., and J. V. Sutcliffe. 1970. River flow forecasting through conceptual models, Part I - a discussion of principles. Journal of Hydrology **10**:282-290.

NSW Government. 2003. Water Sharing Plan for the Upper Namoi and Lower Namoi Regulated River Water Sources 2003. Pages p. 2843 *in* Gazette No. 49, Sydney, Australia.

Piechota, T. C., F. H. S. Chiew, J. A. Dracup, and T. A. McMahon. 1998. Seasonal streamflow forecasting in eastern Australia and the El Nifio-Southern Oscillation. Water Resources Research **34**:3035-3044.

Podbury, T., T. C. Scheales, I. Hussain, and B. S. Fischer. 1998. Use of El Nino climate forecast in Australia. American Journal of Agriculture Economics **80**:1096-1101.

Potts, J. M., C. K. Folland, I. T. Jolliffe, and D. Sexton. 1996. Revised 'LEPS' Score for Assessing Climate Model Simulations and Long-Range Forecasts. Journal of Climate **9**:34-53.

Preece, R. M., and H. A. Jones. 2002. The effect of Keepit Dam on the temperature regime of the Namoi River, Australia. River Research and Applications **18**:397-414.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing. *in*. R Foundation for Statistical Computing, Vienna, Austria.

Ragonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona. 2006. A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. Water Resources Research **42**:W09404.

See, L., and R. J. Abrahart. 2001. Multi-model data fusion for hydrological forecasting. Computers & Geosciences **27**:987-994.

Shamseldin, A. Y., O'Connor, K.M., Liang, G.C. 1997. Methods for combining the outputs of different rainfall-runoff models. Journal of Hydrology **197**:203-229.

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

Sharma, A. 2000c. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 - A nonparametric probabilistic forecast model. Journal of Hydrology **239**:249-258.

Sharma, A., Luk, K.C., Cordery, I., Lall, U. 2000b. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 2 - Predictor identification of quarterly rainfall using ocean-atmosphere information. Journal of Hydrology **239**:240-248.

Shephard, N. 1995. Generalized Linear Autoregressions, February 6 edition. Nuffield College, Oxford.

Simons, M., Podger, G., Cooke, R. 1996. IQQM--A hydrologic modelling tool for water resource and salinity management. Environmental Software **11**:185-192.

Souza Filho, F. A., and U. Lall. 2003. Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. Water Resources Research **39**:1307.

Sveinsson, O. G. B., U. Lall, V. Fortin, L. Perrault, J. Gaudet, S. Zebiak, and Y. Kushnir. 2008. Forecasting Spring Reservoir Inflows in Churchill Basin in Québec, Canada. Journal of Hydrologic Engineering **13**:426-437.

V´azquez, R. F., P. Willems, and J. Feyen. 2008. Improving the predictions of a MIKE SHE catchment-scale application by using a multi-criteriapproach. Hydrological Processes **22**:2159-2179.

van den Dool, H. 2000. Constructed analogue prediction of the east central tropical Pacific SST and the entire world ocean for 2001. Experimental Long-Lead Forecast Bulletin **9**:38-41.

van den Dool, H., J. Huang, and Y. Fan. 2003. Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. Journal of Geophysical Research **108**:8617.

Vazquez, R. F. 2003. Assessment of the performance of physically based distributed codes simulating medium size hydrological systems. PhD. K. U. Leuven, Leuven, Belgium.

Verdon, D. C., and S. W. Franks. 2005. Indian Ocean sea surface temperature variability and winter rainfall: Eastern Australia. Water Resources Research **41**:W09413.

Verdon, D. C., A. M. Wyatt, A. S. Kiem, and S. W. Franks. 2004. Multidecadal variability of rainfall and streamflow: Eastern Australia. Water Resources Research **40**:W10201.

Ward, R. C. 1984. Some aspects of river flow in northern New South Wales, Australia. Journal of Hydrology **71**:31-51.

Westra, S., C. Brown, U. Lall, and A. Sharma. 2007. Modeling multivariable hydrological series: Principal component analysis or independent component analysis? Water Resources Research **43**:W06429.

Westra, S., A. Sharma, C. Brown, and U. Lall. 2008. Multivariate streamflow forecasting using independent component analysis. Water Resources Research **44**:W02437.

Wolff, J. E., E. Maier-Reimer, and S. Legutke. 1997. The Hamburg Ocean Primitive Equation Model. Technical report no. 13, Deutsches Klimarechenzentrum, Hamburg.

Wu, M. 2006. Trimmed and winsorized estimators. PhD Thesis. Michigan State University.

Xiong, L., A. Y. Shamseldin, and K. M. O'Connor. 2001. A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. Journal of Hydrology **245**:196-217.

Yang, C., R. E. Chandler, and V. S. Isham. 2005. Spatial-temporal rainfall simulation using generalized linear models. Water Resources Research **41**:W11415.

Yu, L., S. Wang, and K. K. Lai. 2005. A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. Computers & Operations Research **32**:2523-2541.

Zhang, H., and T. Casey. 2000. Verification of Categorical Probability Forecasts. Weather and Forecasting **15**:80-89.

(blank page)

CHAPTER SEVEN

7 CONCLUSION

The motivation of this research is to reduce predictive uncertainty of models aiding superior water resources management. It explores techniques applicable beyond exhaustion of current model improvement practices such as rigorous calibration and continual structural adjustment. The reduction in uncertainty is achieved by a two step statistical approach. Firstly it proposes shrinking the parameter bias of individual models by addressing input error using the method ‘Simulation Extrapolation’ or SIMEX. Secondly it advocates combining a number of bias corrected individual models using pair wise dynamic weights (PDW) as a way to reduce overall uncertainty of any single modelling structure.

The research demonstrates that SIMEX mitigates the parameter bias of hydrologic models as long as the error distribution and any associated non-stationarity are known. Two types of hydroclimatic models are used representing increasing complexity. First, Chapter 2 demonstrates a reduction of error in Southern Oscillation Index reconstruction using linear regression of NINO3.4 observations with non-stationary noise. Secondly, Chapter 3 exhibits a reduction of error in flow series resulting from erroneous rainfall input in the Sacramento model.

The PDW model combination method is used in three different settings. First of all the PDW is applied to combine three univariate models, the response being the three month ahead forecast of the monthly NINO3.4 SSTA derived index. The study demonstrates that the proposed PDW formulation is an improvement over the use of model combination logic where combination weights do not change with time. PDW exhibits a small but consistent increase in prediction skill over that of current practice of static weight method. The case study concludes that the potential of improvement is real if multiple predictions are combined using the proposed dynamic weight rationale.

The multivariate case study uses three month ahead forecasts of globally grided monthly sea surface temperature anomalies (SSTA) from three predictive models. The prediction skill from static weight combination is used as the base case to compare the merits of using dynamic weights instead. The predicted SSTA using the dynamic combination algorithm consistently exhibits notable improvement, both in space and time, to that from static combination. This constitutes a numerical validation of PDW in a multivariate forecasting context.

Finally, three seasonal forecasts of inflows at five neighbouring locations in Eastern Australia are combined using PDW. The customary high skew associated with arid Australian rivers results in flow volumes many times higher than their respective medians along with significant periods of zero flow. The flow forecasts require specification of predictor variables which are derived from various transformations of concurrent forecasts of Indian and Pacific Ocean SSTA. The combined flow forecast exhibits superior replication of spatial dependence than any single forecast. The PDW approach achieves a minor improvement in the overall flow forecast. However, the combination of flow forecasts does not yield similar enhancement of skill as compared to the SSTA forecast combination. This is due to limited independence among all three flow forecast models. The extent of independence among component forecasts dictates the extent of improvement post combination. This case study explores the strengths and weaknesses of the PDW method in hydrological modelling context.

The original research contributions from this study are reiterated below, while details can be sought in the earlier chapters.

7.1 Original Research Contributions

7.1.1 Mitigation of input error

The SIMEX method mitigates parameter bias caused by inaccurate input variables. It is a functional estimator whereas existing limited research on input error modelling in hydrology is based on the use of structural estimators. Structural estimators often resorts to assumptions regarding the distribution of unobserved true covariates leading to certain likelihood measures. SIMEX offers advantages over structural estimators when model structures are complex with nonlinear and non-stationary error distributions as in hydrologic systems. This is the first reported research on the use of SIMEX in hydrological and climate modelling.

7.1.2 Mitigation of model error

This doctoral work presents a methodology for combining forecasts from multiple models in a dynamic manner. Multiple models are mixed in pairs based on importance weights that are allowed to vary in time reflecting the persistence of individual model skills and of any relevant exogenous variable. The weights are structured in a hierarchical pair wise combination tree.

The optimal weight time series are based on analytical solution (Equation [3], Chapter 5) of the combination that returns error free prediction. The analytical weights do not require minimisation of squared error, contrary to the current practice of estimating weights. While the current practice imposes the reduction of squared error however the reduction may not necessarily correspond to improved predictive skill.

Two approaches for forecasting pair wise dynamic weights (PDW) are introduced. The first of the two approaches uses a mixture of two basis distributions which are three category ordered logistic regression model and a generalised linear autoregressive model. The second method uses a modified nearest neighbour approach to forecast the future weights. The proposed methodologies constitute an original contribution to the literature on forecast combination in hydrological and climate modelling.

7.2 Limitations and Future Work

This research presents case studies that demonstrates the alternatives developed to mitigate model input uncertainty as well as model structural uncertainty. Consequently, the manuscript does not follow a single case study from the beginning to the end, but separate case studies for each of the chapters presented.

The input uncertainty reduction alternative (SIMEX) outlined in Chapters 2 and 3 is used to assess the changes in the formulation of a model to predict El Nino Southern Oscillation indices (Chapter 2) and the hydrological response to rainfall (Chapter 3). It should be noted that it is not applied to the NINO3.4 or global SSTA forecast models included in Chapter 4 or Chapter 5. The forecast models assume the inputs to not be uncertain, an acceptable assumption given the limited information that is available to quantify the extent of uncertainty that could be present. Later in Chapter 6, the structural details of the river flow forecast models are included. The structural accuracy of these flow forecast models are low compared to the uncertainty of the SSTA covariates. Hence SIMEX has a limited potential of improving parameter estimates where bulk of the parameter inaccuracy can be contributed to the model structure.

The three component models used in NINO3.4 case study (Chapter 4) are not followed through to flow forecast model in Chapter 6. The component model UCLA (see Chapter 4.3) is predominantly calibrated for the NINO region and no suitable UCLA hindcast has been available at the time of research progression to multivariate application presented in the later chapters. Consequently, the later chapters replace the UCLA model by the MetF model. The reader may also note a difference between the reductions in squared error post combination presented in Figure 5.7 and Figure 6.1. The difference is caused by different forecast lead time and different comparative bench marks. Chapter 5 uses three months ahead forecast of monthly SSTA, whereas Chapter 6 uses forecast of next three monthly SSTA. The comparative bench mark in the first case is the best single forecast at each grid point and in the second case the benchmark is MetF forecast at every grid point.

Future progression of this research will be to apply all the tools within a single case study of hydroclimatic model simulation. The limitations and future work related to the two main tools (SIMEX, PDW) are presented below.

7.2.1 Simulation extrapolation (SIMEX)

This section first discusses the limitations of the case studies presented to introduce SIMEX. The synthetic rainfall-runoff study assumes that the three model parameters are independently related to the variance inflation factor. This may not be the case in a more general setting, necessitating a multivariate regression relationship that enables all parameters to be modelled jointly. Another limitation of SIMEX is an implied assumption of high structural accuracy of the model compared to the noise in covariates. Hence SIMEX is recommended after satisfying any remaining perfection of the model structure. The case study in Chapter 2 (SOI prediction) is presented as a proof of concept only, while the partitioners are recommended to first check sampling variability (by tools such as bootstrapping) prior to embark into bias correction exercise.

While current research exposes the SIMEX rationale to a hydrological audience, there are several aspects of it that need further work and investigation as listed next.

a) Practice guideline of SIMEX application in hydrology needs attention to following issues. They are: the number of simulations required, the incremental steps of the variance inflation factor needed, more directives on extrapolation methods, and the confidence intervals associated with SIMEX estimated parameters.

b) Another issue to be investigated in greater detail is the specification of the error distribution (with possible co dependence) of various hydroclimatic variables (e.g. rainfall, evaporation). There are several sources of uncertainty in catchment rainfall measurement that needs attention. For example non stationary instrumentation error, conversion of radar rainfall record, downscaling of GCM simulations at a local catchment scale or spatial and temporal transformation of point rainfalls. Reporting of standard error of statistical conversion of point to aerial rainfall is a prerequisite of SIMEX. Practical application of SIMEX recommends future work on proper specification of error of rainfall data.

c) These analyse can be extended to investigate the role of SSTA data error in influencing non stationarity of SOI versus NINO3 relationship providing important insights into ENSO behaviour. The flexibility of allowing non stationary noise band in SIMEX implies that it may be used to transport parameter estimate to future climate change scenario.

d) Current research is limited to the parametric form of SIMEX. Alternative research using a non parametric SIMEX may be useful in data intensive settings.

e) Finally, SIMEX is essentially presented as a robust alternative to existing orthogonal or structural estimators of input error analysis in hydrologic models. A comparative study in future to the orthogonal regression method using historical hydrologic time series will help develop a comprehensive picture on input error modelling.

7.2.2 Pair wise dynamic weight (PDW)

The success of the PDW method does not nullify the strength of simpler static combination methods. The complexity of the PDW combination increases with higher number of component models. Increased number of hyper parameters (parameters external to the component models) reduces the degrees of freedom which may eventually compromise the strength of combination. Hence this method is suitable for component models that come with a long period of hind cast for sake of calibration and validation. In general, this PhD work envisages following future research studies.

a) The optimal number of component models is unclear as past research has been case specific. Generic research on optimal number (as a function of dependence of candidate forecasts) is necessary. A follow on investigation should explore the ways to reduce number of models to the optimal quantity.

b) The case study combines mean forecasts only. Natural progression of the application will be combining full set of component realisations in order to reflect full probability distribution that is likely.

c) One useful application of PDW is to include null model forecast (e.g. climatology) as one of the component models. This will guard against any period of poor forecast accuracy to fall below mean forecast.

d) The design of the tree architecture of PDW needs further attention. This is especially true for combination of multivariate response field. The simplification of applying PDW in only one node (possibly at the highest level) requires simultaneous attention while researching the tree architecture.

e) While the PDW combination logic presented was used to estimate a combined forecast for a seasonal time step, how such a combination will lead to a complete forecast hydrograph with finer time step is a question that needs to be suitably addressed. This is important for time series with long memory for example hourly flow hydrograph or daily storage volume. A combined forecast hydrograph will require a combination of multiple models at multiple points in time (and possibly in space), which will lead to a discontinuity in the mass-balance in the hydrograph whenever such a combination is made. Future research needs to investigate on how this problem can be suitably addressed and mass balance conditions appropriately satisfied.

7.3 List of Peer Reviewed Manuscripts

This PhD work is released in 5 journal papers which are progressively developed over prior presentations in 7 peer reviewed international conferences as listed next.

Peer reviewed journal papers:

Chowdhury, S. and Sharma, A. (2007), Mitigating Parameter Bias in Hydrological Modelling due to Uncertainty in Covariates. *Journal of Hydrology*, 340: 197-204. DOI:10.1016/j.jhydrol.2007.04.010

Chowdhury, S. and Sharma, A. (2008), A simulation based approach for representation of rainfall uncertainty in conceptual rainfall runoff models. *Hydrological Research Letters*, 2: 5-8. DOI: 10.3178/HRL.2.2.

Chowdhury, S. and Sharma, A. (2009) Long Range NINO3.4 Predictions Using Pair Wise Dynamic Combinations of Multiple Models. *Journal of Climate*, **22(3)**, 793-805. DOI: 10.1175/2008JCLI2210.1.

Chowdhury, S. and Sharma, A. (2009a). Global sea surface temperature forecasts using a pair wise dynamic combination approach, *Journal of Climate*, under review.

Chowdhury, S. and Sharma, A. (2009b). Dynamic combination of multi site seasonal flow forecasts, *Water Resources Research*, under review.

Peer reviewed conferences and abstracts:

Chowdhury, S. and Sharma, A. (2005), Measurement error in hydrological data and their effect on water resources modelling, In: *International conference MTERM*, 6-10 June, Bangkok.

Chowdhury, S. and Sharma, A. (2005), Errors in Hydrological Variables and Their Effect on Model Parameters, In: *International congress on modelling simulation, MODSIM05*, December, Melbourne.

Chowdhury, S. and Sharma, A. (2005). Input Uncertainty and its Implications on Parameter Assessment in Hydrologic and Hydroclimatic Modelling Studies, *American Geophysics Union Fall Meeting*, December, San Francisco.

Chowdhury, S. and Sharma, A. (2006). Combining Climate Prediction Models Using Dynamic Weights, In: *30th Water Resources and Hydrology Symposium*. 4-7 December, Launceston.

Chowdhury, S. and Sharma, A. (2007). Pair-wise Dynamic Ensemble of Multiple Models for Hydroclimatic Applications, In: *Asia Oceania Geophysics Society Meeting*, August, Bangkok.

Chowdhury, S. and Sharma, A. (2007). Dynamic Model Mixing for Enhancing the Predictability of Hydroclimatic Variables, In:

International congress on modelling simulation, MODSIM07, December, Christchurch.

Chowdhury, S. and Sharma, A. (2008). Integrating Multiple Forecast Models Using a Dynamic Combination Algorithm: Application to Multi-site Seasonal Inflows, In: *Western Pacific Geophysics Meeting of American Geophysics Union*, August, Cairns.

(blank page)

CHAPTER EIGHT

8 BIBLIOGRAPHY

Agresti, A., 1996: *An Introduction to Categorical Data Analysis*. *Wiley Series in Probability and Statistics*.

Ajami, N. W., Q. Duan, X. Gao, and S. Sorooshian, 2006: Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. *Journal of Hydrometeorology*, **7,** 755-768.

Aksu, C., Gunter, S., 1992: An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecasts. *International Journal of Forecasting*, **8,** 27-43.

Anderson, P. L. and M. M. Meerschaert, 1998: Modeling river flows with heavy tails. *Water Resources Research*, **34,** 2271-2280.

Andreassian, V., C. Perrin, C. Michel, I. Usart-Sanchez, and J. Lavabre, 2001: Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology*, **250,** 206– 223.

Araghinejad, S., D. H. Burn, and M. Karamouz, 2006: Long-lead probabilistic forecasting of streamflow using ocean atmospheric and hydrological predictors. *Water Resources Research*, **42,** W03431.

Armstrong, J. S., 1989: Combining forecasts: The end of the beginning or the beginning of the end? *International Journal of Forecasting*, **5,** 585-588.

——, 2001: Combining forecasts. *Principles of Forecasting: A Handbook for Researchers and Practitioners*, J. S. Armstrong, Ed., Kluwer Academic Publishers, 417-439.

Augustin, N. H., L. Beevers, and W. T. Sloan, 2008: Predicting river flows for future climates using an autoregressive multinomial logit model. *Water Resources Research*, **44,** W07403.

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,** 1783-1796.

Barry, B. A., 1978: *Errors in practical measurement in science, engineering, and technology*. Wiley.

Bates, J. M. and C. W. J. Granger, 1969: The combination of forecasts. *Operational Research Quarterly*, **20,** 451-468.

Berry, S. M., R. J. Carroll, and D. Ruppert, 2002: Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, **97,** 160-169.

Beven, K., 2006: On undermining the science? *Hydrological Processes, HPToday*, **20,** 3141-3146.

Beven, K. and J. Freer, 2001: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, **249,** 11-29.

Bottomley, M., C. K. Folland, J. Hsiung, R. E. Newell, and D. E. Parker, 1990: Global Ocean Surface Temperature Atlas.

Boughton, W., 2005: Catchment water balance modelling in Australia 1960–2004. *Agricultural Water Management*, **77,** 91-116.

Box, G. E. P. and D. R. Cox, 1964: An Analysis of Transformations. *Journal of Royal Statistical Society (Series B)*, **26,** 211-252.

Brown, B. W. and R. S. Mariano, 1993: *Stochastic simulations for inference in nonlinear errors-in-variables models.* Vol. 11, *Handbook of Statistics*, North Holland, 611-627 pp.

Bruhn, J. A., W. E. Fry, and G. W. Fick, 1980: Simulation of Daily Weather Data Using Theoretical Probability Distributions. *Journal of Applied Meteorology***,** 1029-1036.

Burnash, R. J., 1975: The NWS river forecast system–catchment modelling. *Computer Models of Watershed Hydrology*, V. P. Singh, Ed., Water Resources Publications.

Burnash, R. J., Ferral, R.L., McGuire, R.A., 1973: A Generalized Streamflow Simulation System–Conceptual Modelling for Digital Computers.

Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen, 2004: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology***,** 242-266.

——, 2004: An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology*, **298,** 242-266.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C., 1995: A limited memory algorithm for bound constrained optimization. *Scientific Computing*, **16,** 1190-1208.

Cardoso, A. O. and P. L. S. Dias, 2006: The relationship between ENSO and Paran´a River flow. *Advances in Geosciences*, **6,** 189-193.

Cardoso, A. O., R. T. Clarke, and P. L. S. Dias, 2005: A case study of the sea surface temperatures (SSTs) to obtain predictors of river flow. *Regional Hydrological Impacts of Climatic Variability and Change - Imapact Assessment and Decision Making*, Foz do Iguacu, Brazil, IAHS, 231-238.

Carpenter, T. M. and K. P. Georgakakos, 2004: Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrologic model. *Journal of Hydrology*, **298,** 202-221.

Carrasco, J. A. and J. D. Ortuzar, 2002: Review and assessment of the nested logit model. *Transport Reviews*, **22,** 197-218.

Carroll, R. J., D. Ruppert, and L. A. Stefanski, 1995: *Measurement Error in Nonlinear Models*. *Monographs on Statistics and Applied Probability, 63*, Chapman and Hall, 305 pp.

Carroll, R. J., J. D. Maca, and D. Ruppert, 1999: Nonparametric regression in the presence of measurement error. *Biometrika*, **86,** 541-554.

Chambers, J. M., 1992: Linear Models. *Statistical Models in S*, J. M. Chambers and T. Hastie, Eds., Wadsworth & Brooks, 95-144.

Chandler, R. E., 2005: On the use of generalized linear models for interpreting climate variability. *Environmetrics*, **16,** 699-715.

Chen, D., M. A. Cane, A. Kaplan, S. E. Zebiak, and D. Huang, 2004: Predictability of El Niño over the past 148 years. *Nature*, 733 - 736.

Chiew, F. H. A., T. C. Piechota, J. A. Dracup, and T. A. McMahon, 1998: El Nino Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology*, **204,** 138-149.

Chiew, F. H. S. and T. A. McMahon, 2002: Global ENSO–streamflow teleconnection, streamflow forecasting and interannual variability. *Hydrological Sciences*, **47,** 505-522.

Chiew, F. H. S., S. L. Zhou, and T. A. McMahon, 2003: Use of seasonal streamflow forecasts in water resources management. *Journal of Hydrology*, **270,** 135-144.

Chowdhury, S. and A. Sharma, 2007: Mitigating Parameter Bias in Hydrological Modelling due to Uncertainty in Covariates. *Journal of Hydrology*, **340,** 197-204.

——, 2008: A simulation based approach for representation of rainfall uncertainty in conceptual rainfall runoff models. *Hydrological Research Letters*, **2,** 5-8.

——, 2009: Multi-site Seasonal Forecast of Arid River Flows Using a Dynamic Model Combination Approach. *Water Resources Research*, **under review**.

——, 2009a: Long Range NINO3.4 Predictions Using Pair Wise Dynamic Combinations of Multiple Models. *Journal of Climate*, **22,** 793-805.

——, 2009b: Global sea surface temperature forecasts using a pair wise dynamic combination approach. *Journal of Climate*, **under review**.

Chumchean, S. and A. Sharma, 2006: An integrated approach to error correction for real-time radar-rainfall estimation. *Journal of Atmospheric and Oceanic Technology*, **23,** 67-79.

Clemen, R. T., 1989: Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, **5,** 559-583.

Cleveland, W. S., 1979: Robust locally weighted regression and smoothing scatterplots. *Journal of American Statistical Association*, **74,** 829-836.

Cleveland, W. S. and E. Grosse, 1991: Computational methods for local regressions. *Statistics and Computing*, **1,** 47-62.

Cleveland, W. S., S. J. Devlin, and E. Grosse, 1988: Regression by Local Fiting Methods, Properties, and Computational Algorithms. *Journal of Econometrics*, **37,** 87-114.

Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson, 2004: Forecast Calibration and Combination: A Simple Bayesian Approach for ENSO. *Journal of Climate*, **17,** 1504-1516.

Colman, A. W. and M. K. Davey, 2003: Statistical prediction of global sea-surface temperature anomalies. *International Journal of Climatology*, **23,** 1677-1697.

Common, P., 1991: Independent Component Analysis, A New Concept? *Signal Processing*, **24,** 1-10.

Cook, J. R. and L. A. Stefanski, 1994: Simulation-Extrapolation Estimation in Parametric Measurement Error Models. *Journal of the American Statistical Association*, **89,** 1314-1328.

Cordery, I., R. Mehrotra, A. Sharma, and M. J. Nazemosadat, 2004: Hydrological models in a non stationary environment. *Hydrology, Science and practice for the 21st century*, Imperial College, London, British Hydrological Society, 103-107.

Crapper, P. F., S. G. Beavis, and L. Zhang, 1999: The relationship between climate and streamflow in the Namoi Basin. *Environment International*, **25,** 827-839.

CSIRO, 2007: *Water availability in the Namoi. A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project.* CSIRO, 154 pp.

Dawes, R., R. Fildes, M. Lawrence, and K. Ord, 1994: The past and the future of forecasting research. *International Journal of Forecasting*, **10,** 151-159.

de Menezes, L. M., D. W. Bunn, and J. W. Taylor, 2000: Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, **120,** 190-204.

DelSole, T. and J. Shukla, 2009: Artificial Skill Due to Predictor Screening. *Journal of Climate*, **22,** 331-345.

Deutsch, M., C. W. J. Granger, and T. Terasvirta, 1994: The combination of forecasts using changing weights. *International Journal of Forecasting*, **10,** 47-57.

Devanarayan, V. and L. A. Stefanski, 2002: Empirical simulation extrapolation for measurement error models with replicate measurements. *Statistics & Probability Letters*, **59,** 219-225.

Devineni, N., A.Sankarasubramanian, and S. Ghosh, 2008: Multimodel Ensembles of Streamflow Forecasts: Role of Predictor State in

Developing Optimal Combinations. *Water Resources Research*, **44,** W09404.

Dieck, R. H., 2002: *Measurement Uncertainty: Methods and Application*. 3rd edition ed. Instrumentation, Systems and Automation Society (ISA).

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

Drosdowsky, W. and L. E. Chambers, 2001a: Near-Global Sea Surface Temperature Anomalies as Predictors of Australian Seasonal Rainfall. *Journal of Climate*, **14,** 1677-1687.

Duan, Q., S. Sorooshian, and V. Gupta, 1992: Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, **28,** 1015-1031.

Dunteman, G. H. and M. R. Ho, 2006: *An Introduction to Generalized Linear Models*. Vol. 145, *Quantitative Applications in the Social Sciences*, Sage Publications.

Dutta, S. C., J. W. Ritchie, D. M. Freebairn, and G. Y. Abawi, 2006: Rainfall and streamflow response to El Niño Southern Oscillation: a case study in a semiarid catchment, Australia. *Hydrological Sciences*, **51,** 1006-1020.

ECMWF, 2004: Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER) 434.

Feiring, M. B., 1967: *Streamflow Synthesis*. Harvard University Press.

Ferrari, S. L. P. and F. Cribari-Neto, 2004: Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, **31,** 799-815.

Fraedrich, K. and N. Smith, 1989: Combining Predictive Scheme in Long-Range Forecasting. *Journal of Climate*, **2,** 291-294.

Fritsch, J. M., J. Hilliker, and J. Ross, 2000: Model Consensus. *Wether Forecasting*, **15,** 571-582.

Fuller, W. A., 1987: *Measurement Error Models*. John Wiley & Sons.

Garrick, M., C. Cunnane, and J. E. Nash, 1978: A criterion of efficiency for rainfall runoff modelling. *Journal of Hydrology*, **36,** 375-381.

Gelman, A. and J. Hill, 2006: *Data Analysis Using Regression and Multilevel/Hierarchical Models*. 1st edition ed. Cambridge University Press.

Georgakakos, K. P., D.-J. Seo, H. Gupta, J. Schaake, and M. B. Butts, 2004: Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, **298,** 222-241.

Girolami, M. and C. Fyfe, 1996: Negentropy and Kurtrosis as projection persuit indices provide generalized ICA algorithms. *Advances in Neural Information Processing Systems Workshops: Blind Signal Processing*, Snowmaas, CO.

Goddard, L., Mason, S.J., 2002: Sensitivity of seasonal climate forecasts to persisted SST anomalies. *Climate Dynamics*, **19,** 619-631.

Gooijer, J. G. D. and R. J. Hyndman, 2006: 25 years of time series forecasting. *International Journal of Forecasting*, **22,** 443-473.

Goswami, M. and K. M. O'Connor, 2007: Real-time flow forecasting in the absence of quantitative precipitation forecasts: A multi-model approach. *Journal of Hydrology*, **334,** 125-140.

Granger, C. W. J. and P. Newbold, 1977: *Forecasting economic time series*. Academic Press.

Grantz, K., B. Rajagopalan, M. Clark, and E. Zagona, 2005: A technique for incorporating large-scale climate information in basin-scale ensemble streamflow forecasts. *Water Resources Research*, **41,** W10410.

Greene, A. M., L. Goddard, and U. Lall, 2005: Probabilistic multimodel climate change scenarios for the evolution of regional temperatures.

Guetter, A. and K. Georgakakos, 1996: Are the El Nino and La Nina Predictors of the Iowa River Seasonal Flow? *Journal of Applied Meteorology*, **35,** 690-705.

Hall, D. B., 2000: Zero-inflated Poisson binomial regression models with random effects: a case study. *Biometrics*, **56,** 1030-1039.

Hameed, T. and G. Podger, 2001: Use of IQQM simulation model for planning and management of a regulated system. *IAHS Red Book 2001*, 83-89.

Hamlet, A. F. and D. P. Lettenmaier, 1999: Columbia River streamflow forecasting based on ENSO and PDO climate signals. *Journal of Water Resources Planning and Management-Asce*, **125,** 333-341.

Harrold, T. I., A. Sharma, and S. J. Sheather, 2003: A nonparametric model for stochastic generation of daily rainfall amounts. *Water Resources Research*, **39**.

Hastie, T. and R. Tibshirani, 1986: Generalized Additive Models. *Statistical Science*, **1,** 297-310.

Hastie, T. and D. Pregibon, 1992: Generalised Linear Models. *Statistical Models in S*, J. M. Chambers and T. Hastie, Eds., Wadsworth & Brooks, 195-248.

Hastie, T., R. Tibshirani, and J. Friedman, 2000: *The Elements of Statistical Learning, Data Mining, Inference and Prediction*. Springer Series in Statistics.

Helsel, D. R. and R. M. Hirsch, 2002: Statistical Methods in Water Resources. *Book 4, Hydrologic Analysis and Interpretation*, United States Geological Survey.

Hibon, M. and T. Evgeniou, 2005: To combine or not to combine: selecting among forecasts and their combinations. *International Journal of Forecasting*, **21,** 15-24.

Hoeting, J. A., Madigan, D., Raftery, A. E. and Volinsky, C. T., 1999: Bayesian Model Averaging: A Tutorial (with discussion). *Statistical Science*, **14,** 382-417.

Holcomb, J., 1999: Regression with covariates and outcome calculated from a common set of variables measured with error: Estimation using the SIMEX method. *Statistcs in Medicine*, **18,** 2847-2862.

Hsieh, W. W., Yuval, J. Li, A. Shabbar, and S. Smith, 2003: Seasonal Prediction with Error Estimation of Columbia River Streamflow in British Columbia. *Journal of Water Resources Planning and Management-Asce*, **129,** 146-149.

Huang, M. and X. Liang, 2006: On the assessment of the impact of reducing parameters and identification of parameter uncertainties for a hydrologic model with applications to ungauged basins. *Journal of Hydrology*, **320,** 37-61.

Huard, D. and A. Mailhot, 2006: A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model "abc". *Water Resources Research*, **42,** WR004661.

Huerta, G. and B. Sansó, 2007: Time-varying models for extreme values. *Environ Ecol Stat*, **14,** 285-299.

Hyvarinen, A. and E. Oja, 2000: Independent Component Analysis: Algorithms and Applications. *Neural Networks*, **13,** 411-430.

Jaech, J. L., 1985: *Statistical analysis of measurement errors*. Wiley.

Jonsdottir, H., H. Madsen, and O. P. Palsson, 2006: Parameter estimation in stochastic rainfall-runoff models. *Journal of Hydrology*, **326,** 379-393.

Jose, V. R. R. and R. L. Winkler, 2008: Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, **24,** 163-169.

Kachroo, R. K., 1992: River flow forecasting. Part 1. A discussion of the principles. *Journal of Hydrology*, **133,** 1-15.

Kamstra, M. and P. Kennedy, 1998: Combining qualitative forecasts using logit. *International Journal of Forecasting*, **14,** 83-93.

Kaplan, A., Y. Kushnir, M. A. Cane, and M. B. Blumenthal, 1997: Reduced space optimal analysis for historical data sets: 136 years of Atlantic sea surface temperatures. *Journal of Geophysical Research (C. Oceans).* **102,** 27,835-27,860.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan, 1998: Analyses of global sea surface temperature 1856-1991. *Journal of Geophysical Research (C. Oceans).* **103,** 8,567-18,589.

Karlsson, M. and S. Yakowitz, 1987: Nearest neighbor methods for non parametric rainfall runoff forecasting. *Water Resources Research*, **23,** 1300-1308.

Kavetski, D., S. Franks, and G. Kuczera, 2002: Confronting Input Uncertainty in Environmental Modelling. *Calibration of Watershed Models*, H. V. Gupta, Sorooshian, S., Rousseau, A. N. and R. Turcotte, Ed., 49-68.

Kavetski, D., G. Kuczera, and S. W. Franks, 2006a: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, **42**.

——, 2006b: Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. *Water Resources Research*, **42**.

Khadam, I. M. and J. J. Kaluarachchi, 2004: Use of soft information to describe the relative uncertainty of calibration data in hydrologic models. *Water Resources Research*, **40,** W11505.

Kim, C., C. Hong, and M. Jeong, 2000: Simulation-extrapolation via the Bezier curve in measurement error models. *Communications in Statistics-Simulation and Computation*, **29,** 1135-1147.

Kim, J. and L. J. Gleser, 2000: SIMEX approaches to measurement error in ROC studies. *Communications in Statistics-Theory and Methods*, **29,** 2473-2491.

Kim, Y. O., D. Jeong, and I. H. Ko, 2006: Combining rainfall-runoff model outputs for improving ensemble streamflow prediction. *Journal of Hydrologic Engineering*, **11,** 578-588.

Kondrashov, D., S. Kravtsov, A. W. Robertson, and M. Ghil, 2005: A Hierarchy of Data-Based ENSO Models. *Journal of Climate*, **18,** 4425-4444.

Kuczera, G., 1997: Efficient subspace probabilistic parameter optimization for catchment models. *Water Resources Research*, **33,** 177-185.

Lall, U. and A. Sharma, 1996: A nearest neighbor bootstrap for time series resampling. *Water Resources Research*, **32,** 679-693.

Lall, U., Y.-I. Moon, H.-H. Kwon, and K. Bosworth, 2006: Locally weighted polynomial regression: Parameter choice and application to forecasts of the Great Salt Lake. *Water Resources Research*, **42**.

Laplace, P. S. d., 1818: Deuxieme Supplement a la Theorie Analytique des Probabilities (Courcier Paris). *reprinted (1847) in Oeuvres Completes de Laplace*, **7,** 531-580.

Larrick, R. P. and J. B. Soll, 2006: Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, **52,** 111-127.

Latif, M., D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, E. Schneider, 1998: A review of the predictablility and prediction of ENSO. *Journal of Geophysical Research*, **103,** 14,375-14,393.

Leagates, D. R. and G. J. McCabe, 1999: Evaluating the use of 'goodness of fit' measures in hydrologic and hyroclimatic model validation. *Water Resources Research*, **35,** 233-241.

Lee, S. J., K. P. Georgakakos, and J. L. Schnoor, 1990: Effects of Parameter Uncertainty on Long-Term Simulations of Lake Alkalinity. *Water Resources Research*, **26,** 459-467.

Letcher, R. A., F. H. S. Chiew, and A. J. Jakeman, 2004: An Assessment of the Value of Seasonal Forecasts in Australian Farming Systems. *International Environmental Modelling and Software Society (iEMSs) Conference*, Germany, University of Osnabruck, 1511-1516.

Liang, K. Y. and X. H. Liu, 1991: Estimating equations in generalized linear models with measurement error. *Estimating Functions*, V. P. Godambe, Ed., Clarendon Press.

Liden, R. and J. Harlin, 2000: Analysis of conceptual rainfall-runoff modelling performance in different climates. *Journal of Hydrology*, **238,** 231-247.

Lin, X. H. and R. J. Carroll, 2000: Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, **95,** 520-534.

Linder, E., G. P. Patil, and G. S. II, 1993: Errors-in-variables analysis of extrapolation procedures in environmental toxiology. *Multivariate Environmental Statistics*, G. P. Patil and C. R. Rao, Eds., Elsevier Science Publishers B.V., 227-254.

Lundberg, S., T. Tera¨svirta, and D. van Dijk, 2000: Time-varying smooth transition autoregressive models.376.

Madec, G., P. Delecluse, M. Imbard, and C. Levy, 1997: OPA Release 8, Ocean General Circulation Model Reference Manual, 200 pp.

Madsen, H., 2000: Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology*, **235,** 276-288.

——, 2003: Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, **26,** 205-216.

Maity, R. and D. N. Kumar, 2009: Hydroclimatic influence of large-scale circulation on the variability of reservoir inflow. *Hydrological Processes*, **23,** 934-942.

Marcus, A. H. and R. W. Elias, 1998: Some useful statistical methods for model validation. *Environmental Health Perspectives*, **106,** 1541-1550.

Marshall, L., 2006: Bayesian analysis of rainfall-runoff models: Insights to parameter estimation, model comparison and hierarchical model development, School of Civil and Environmental Engineering, University of New South Wales, 222.

Marshall, L., D. Nott, and A. Sharma, 2007: Towards Dynamic Catchment Modelling: A Bayesian Hierarchical Modelling Framework. *Hydrological Processes*, **21,** 847-861.

Martin, T., B. Wintle, J. Rhodes, P. Kuhnert, S. Field, S. Low-Choy, A. yre, and H. Possingham, 2005: Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, **8,** 1235-1246.

Mason, S. J., 2008: Understanding forecast verification statistics. *Meteorological Applications*, **15,** 31-40.

McCullagh, P. and J. A. Nelder, 1989: *Generalized Linear Models*. 2nd ed. ed. Chapman & Hall.

McLeod, A. I., D. J. Noakes, K. W. Hipel, and R. M. Thompstone, 1987: Combining hydrologic forecast. *Journal of Water Resources Planning Management*, **113,** 29-41.

McMahon, T. A., 1979: Hydrological characteristics of arid zones. *The hydrology of areas of low precipitation. Proc. Canberra symposium, December 1979, (International Association of Hydrological Sciences, Washington DC; IAHS-AISH Publication 128)***,** 105-123.

McMahon, T. A., R. M. Vogel, M. C. Peel, and G. G. S. Pegram, 2007: Global streamflows - Part 1: Characteristics of annual streamflows. *Journal of Hydrology*, **347,** 243-259.

Mehrotra, R. and A. Sharma, 2004: Conditional resampling of hydrologic time series using multiple predictor variables: A k-nearest neighbour approach. *Water Resources Research*.

——, 2006: A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. *Journal of Geophysical Research - Atmospheres*, **111**.

——, 2006: Conditional resampling of hydrologic time series using multiple predictor variables: A K-nearest neighbour approach. *Advances in Water Resources*, **29,** 987-999.

Menezes, L. M. d., D. W. Bunn, and J. W. Taylor, 2000: Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, **120,** 190-204.

Miller, C. M., R. T. Clemen, and R. L. Winkler, 1992: The effect of nonstationarity on combined forecasts. *International Journal of Forecasting*, **7,** 515-529.

Montanari, A., 2007: What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrological Processes, HPToday*, **21,** 841-845.

Montanari, A. and G. Grossi, 2008: Estimating the uncertainty of hydrologic forecasts: a statistical approach. *Water Resources Research*, **44,** W00B08.

Muluye, G. Y. and P. Coulibaly, 2007: Seasonal reservoir inflow forecasting with low-frequency climatic indices: a comparison of data-driven methods. *Hydrological Sciences*, **52,** 508-522.

Nash, J. E. and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models, Part I - a discussion of principles. *Journal of Hydrology*, **10,** 282-290.

NSW Government, 2003: Water Sharing Plan for the Upper Namoi and Lower Namoi Regulated River Water Sources 2003. *Gazette No. 49*, p. 2843.

Palmer, T. N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Delecluse, Deque, M. , 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.

Pavan, V. and F. J. Doblas-Reyes, 2000: Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. *Climate Dynamics*, **16,** 611-625.

Pena, M. and H. van den Dool, 2008: Consolidation of Multi Model Forecasts by Ridge Regression: Application to Pacific Sea Surface Temperature. *Journal of Climate*, **21,** 6521-6538.

Peng, P., A. Kumar, and H. van den Dool, 2002: An analysis of multimodel ensemble predictions for seasonal climate anomalies. *Journal of Geophysical Research*, **107,** 4710.

Phillips-Wren, G. E., E. D. Hahn, and G. A. Forgionne, 2004: A multiple-criteria framework for evaluation of decision support systems. *Omega (Oxford)*, **32,** 323.

Piechota, T. C., F. H. S. Chiew, J. A. Dracup, and T. A. McMahon, 1998: Seasonal streamflow forecasting in eastern Australia and the El Nifio-Southern Oscillation. *Water Resources Research*, **34,** 3035-3044.

Podbury, T., T. C. Scheales, I. Hussain, and B. S. Fischer, 1998: Use of El Nino climate forecast in Australia. *American Journal of Agriculture Economics*, **80,** 1096-1101.

Potts, J. M., C. K. Folland, I. T. Jolliffe, and D. Sexton, 1996: Revised 'LEPS' Score for Assessing Climate Model Simulations and Long-Range Forecasts. *Journal of Climate*, **9,** 34-53.

Preece, R. M. and H. A. Jones, 2002: The effect of Keepit Dam on the temperature regime of the Namoi River, Australia. *River Research and Applications*, **18,** 397-414.

R Development Core Team, 2004. R Foundation for Statistical Computing.

——, 2006. R Foundation for Statistical Computing.

——, 2008. R Foundation for Statistical Computing.

Rabinovich, S. G., 1993: *Measurement errors : theory and practice*. American Institute of Physics.

Raftery, A. E., T. Geneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian Model Averaging to Calibrate Forecast Ensembles. *Monthly Weather Review*, **133,** 1155-1174.

Ragonda, S. K., B. Rajagopalan, M. Clark, and E. Zagona, 2006: A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River Basin. *Water Resources Research*, **42,** W09404.

Rajagopalan, B., U. Lall, and S. E. Zebiak, 2002: Categorical Climate Forecasts through Regularization and Optimal Combination of Multiple GCM Ensembles{ast}. *Monthly Weather Review*, **130,** 1792-1811.

Rajaram, H. and K. P. Georgakakos, 1989: Recursive parameter estimation of hydrologic models. *Water Resources Research*, **25,** 281-294.

Reynolds, R. W. and T. M. Smith, 1995: A High-Resolution Global Sea Surface Temperature Climatology. *Journal of Climate*, **8,** 1571-1583.

Robertson, A. W., U. Lall, S. E. Zebiak, and L. Goddard, 2004: Improved Combination of Multiple Atmospheric GCM Ensembles for Seasonal Prediction. *Monthly Weather Review*, **132,** 2732-2744.

Ruiz, J. E., I. Cordery, and A. Sharma, 2005: Integrating Ocean Subsurface Temperatures in Statistical ENSO Forecasts. *Journal of Climate*, **18,** 3571.

Sanders, F., 1963: On subjective probability forecasting. *Journal of Applied Meteorology*, **2,** 191-201.

See, L. and R. J. Abrahart, 2001: Multi-model data fusion for hydrological forecasting. *Computers & Geosciences*, **27,** 987-994.

Shamseldin, A. Y., O'Connor, K.M., Liang, G.C., 1997: Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology*, **197,** 203-229.

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

——, 2000c: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 - A nonparametric probabilistic forecast model. *Journal of Hydrology*, **239,** 249-258.

Sharma, A. and U. Lall, 1999: A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation*, **48,** 361-371.

——, 2004: Model averaging and its use in probabilistic forecasting of hydrologic variables. *Hydrology, Science and practice for the 21st century*, Imperial College, London, British Hydrological Society, 372-378.

Sharma, A., Luk, K.C., Cordery, I., Lall, U., 2000b: Seasonal to interannual rainfall probabilistic forecasts for improved water supply management:

Part 2 - Predictor identification of quarterly rainfall using ocean-atmosphere information. *Journal of Hydrology*, **239,** 240-248.

Shephard, N., 1995: *Generalized Linear Autoregressions*. February 6 ed. *Economics Working Paper 8*, Nuffield College.

Simons, M., Podger, G., Cooke, R., 1996: IQQM--A hydrologic modelling tool for water resource and salinity management. *Environmental Software*, **11,** 185-192.

Smith, T. M. and R. W. Reynolds, 2002: Extended reconstruction of global seas surface temperatures based on COADS data 1854 to 1997. *Journal of Climate*, **16,** 1495-1510.

Sorooshian, S., Q. Duan, and V. K. Gupta, 1993: Calibration of rainfall runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. *Water Resources Research*, **29,** 1185-1194.

Souza Filho, F. A. and U. Lall, 2003: Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. *Water Resources Research*, **39,** 1307.

Srikanthan, R. and T. A. McMahon, 2001: Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and Earth System Sciences*, **5,** 653-670.

Staudenmayer, J. and D. Ruppert, 2004: Local polynomial regression and simulation-extrapolation. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, **66,** 17-30.

Stefanski, L. A., 2000: Measurement error models. *Journal of the American Statistical Association*, **95,** 1353-1358.

Stefanski, L. A. and J. R. Cook, 1995: Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association*, **90,** 1247-1256.

Sveinsson, O. G. B., U. Lall, V. Fortin, L. Perrault, J. Gaudet, S. Zebiak, and Y. Kushnir, 2008: Forecasting Spring Reservoir Inflows in Churchill Basin in Québec, Canada. *Journal of Hydrologic Engineering*, **13,** 426-437.

Terui, N. and H. K. van Dijk, 2002: Combined forecasts from linear and nonlinear time series models. *International Journal of Forecasting*, **18,** 421-438.

Thompson, P. D., 1977: How to Improve Accuracy by Combining Independent Forecast. *Monthly Weather Review*, **105,** 228-229.

Tolson, B. A. and C. A. Shoemaker, 2008: Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resources Research*, **44,** W04411.

Troutman, B. M., 1982: An analysis of input errors in precipitation-runoff models using regression with errors in independent variables. *Water Resources Research*, **18,** 947–964.

V´azquez, R. F., P. Willems, and J. Feyen, 2008: Improving the predictions of a MIKE SHE catchment-scale application by using a multi-criteriapproach. *Hydrological Processes*, **22,** 2159-2179.

van den Dool, H., 2000: Constructed analogue prediction of the east central tropical Pacific SST and the entire world ocean for 2001. *Experimental Long-Lead Forecast Bulletin*, **9,** 38-41.

van den Dool, H., J. Huang, and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981–2001. *Journal of Geophysical Research*, **108,** 8617.

Vazquez, R. F., 2003: Assessment of the performance of physically based distributed codes simulating medium size hydrological systems, Department of Civil Engineering, K. U. Leuven.

Verdon, D. C. and S. W. Franks, 2005: Indian Ocean sea surface temperature variability and winter rainfall: Eastern Australia. *Water Resources Research*, **41,** W09413.

Verdon, D. C., A. M. Wyatt, A. S. Kiem, and S. W. Franks, 2004: Multidecadal variability of rainfall and streamflow: Eastern Australia. *Water Resources Research*, **40,** W10201.

Viney, N. R. and B. C. Bates, 2004: It Never Rains on Sunday: The prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set. *International Journal of Climatology*, **24,** 1171-1192.

Vrugt, J. A., C. G. H. Diks, H. V. Gupta, W. Bouten, and J. M. Verstraten, 2005: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, **41,** W01017.

Wagener, T. and H. S. Wheater, 2006: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty. *Journal of Hydrology*, **320,** 132-154.

Ward, R. C., 1984: Some aspects of river flow in northern New South Wales, Australia. *Journal of Hydrology*, **71,** 31-51.

West, M. and J. Harrison, 1997: *Bayesian Forecasting and Dynamic Models*. 2nd ed. Springer-Verlag.

Westra, S., C. Brown, U. Lall, and A. Sharma, 2007: Modeling multivariable hydrological series: Principal component analysis or independent component analysis? *Water Resources Research*, **43,** W06429.

Westra, S., A. Sharma, C. Brown, and U. Lall, 2008: Multivariate streamflow forecasting using independent component analysis. *Water Resources Research*, **44,** W02437.

Wolff, J. E., E. Maier-Reimer, and S. Legutke, 1997: The Hamburg Ocean Primitive Equation Model. Technical report no. 13.

Wu, M., 2006: Trimmed and winsorized estimators, Probability and Statistics Department, Michigan State University, 97.

Xiong, L., A. Y. Shamseldin, and K. M. O'Connor, 2001: A non-linear combination of the forecasts of rainfall-runoff models by the first-order Takagi-Sugeno fuzzy system. *Journal of Hydrology*, **245,** 196-217.

Yan, Z., S. Bate, R. E. Chandler, V. Isham, and H. Wheater, 2002: An Analysis of Daily Maximum Wind Speed in Northwestern Europe Using Generalized Linear Models. *Journal of Climate*, **15,** 2073-2088.

Yang, C., R. E. Chandler, and V. S. Isham, 2005: Spatial-temporal rainfall simulation using generalized linear models. *Water Resources Research*, **41,** W11415.

Yapo, P. O., H. V. Gupta, and S. Sorooshian, 1996: Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology*, **181,** 23-48.

——, 1998: Multi-objective global optimization for hydrologic models. *Journal of Hydrology*, **204,** 83-97.

Yu, L., S. Wang, and K. K. Lai, 2005: A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. *Computers & Operations Research*, **32,** 2523-2541.

Zhang, H. and T. Casey, 2000: Verification of Categorical Probability Forecasts. *Weather and Forecasting*, **15,** 80-89.

Zhu, C., R. H. Byrd, P. Lu, and J. Nocedal, 1997: Algorithm 778: L-BFGS-B: Fortran Subroutines for Large-Scale Bound-Constrained Optimization. *ACM Transactions on Mathematical Softwares*, **23,** 550-556.

Zou, H. and Y. Yang, 2004: Combining time series models for forecasting. *International Journal of Forecasting*, **20,** 69-84.