Coping with model structural uncertainty in medium-term hydro-climatic forecasting

Ashish Sharma and Shahadat Chowdhury

ABSTRACT

This paper reviews two alternatives for reducing structural uncertainty in medium-term hydro-climatic forecasting. The first is a static ensemble average, illustrated here using the Multiple Reservoir Inflow Forecasting System, a nonparametric probabilistic forecasting model that relates streamflow to climate predictors, and generates monthly sequences of multi-site flow from the present for the coming 12 months. Instead of forming a single predictive relationship, multiple constituent models, each having their own unique predictor variable sets, are formed. A weighted probabilistic combination of these constituent models completes the static ensemble average. The second alternative is a dynamic ensemble average that allows constituent models to change importance with time, model weights evolving as a function of these weights at preceding time steps. Dynamic model combination is demonstrated here for first combining multiple sea surface temperature anomaly forecasts to produce a global sea surface temperature anomaly field, and then using the dynamically combined sea surface temperature anomaly (SSTA) field to concurrently ascertain inflows at multiple locations in a semi-arid Australian catchment. The paper concludes by identifying scenarios under which one would expect to see improvements as a result of static or dynamic model combination, and provides suggestions for further research in this area.

Key words | dynamic combination, model uncertainty, probabilistic forecasting, rainfall, seasonal forecasting, streamflow

Ashish Sharma (corresponding author)
School of Civil and Environmental Engineering,
The University of New South Wales,
Sydney, NSW 2052,
Australia
E-mail: *a.sharma@unsw.edu.au*

Shahadat Chowdhury
School of Civil and Environmental Engineering,
The University of New South Wales,
Sydney, NSW 2052,
Australia
(now at the NSW Department of Environment, Climate Change and Water)

INTRODUCTION

Medium- to long-term probabilistic forecasting of rainfall or streamflow involves predicting (often as a conditional probability distribution) the responses being studied as a function of lead time, conditional to relevant climatic and catchment forcings. Examples of such probabilistic forecast methods are many, including those reported in (Peichota *et al.* 1998; Sharma 2000*a*, *b*; Sharma *et al.* 2000; Chiew *et al.* 2003; Souza & Lall 2003; Regonda *et al.* 2006; Westra *et al.* 2008; Wang *et al.* 2009; Westra & Sharma 2009), most such approaches providing users with the probability with which the response may fall in a designated category (such as "low", "medium" or "high") or with which it may exceed a specified threshold. A limitation of such forecasts is that they cannot be used to ascertain the risks associated with undertaking longer-term water planning. For such planning to be done, the probabilistic forecasts must link across time (to the maximum lead time they are designed for) and across the multiple nodes of the multi-node water system they represent, the resulting forecasts being multivariate sequences of flows starting from the current for a fixed length of time. This allows the forecasts to be used as representative scenarios or realizations indicative of how inflows may evolve over the medium

doi: 10.2166/nh.2011.104

to long-term planning horizon, thereby enabling water managers to identify operating policies that maximize profits and minimize risks of system failure.

Another limitation of many of the approaches mentioned above is their need to resort to a simplistic representation of the climate variables that are assumed to drive local hydro-climatology. In most cases, the dominant modes of medium term climatic variability (such as the El Niño Southern Oscillation, the Pacific Decadal Oscillation or its variants, or the Indian Ocean Dipole, see Westra *et al.* (2010) for discussion and associated references), are assumed to represent the sole descriptors of the local climate being modeled, an assumption that becomes questionable in regions where other influences may be dominant. However, the reason why such simplistic representations are used is often the instability associated with specifying alternate predictors, often because the use of many such predictors or their combinations leads to similar variability in the responses being modeled. Specification of such a forecasting approach which is unstable in its structural representation often results in unstable forecasts of the future.

A clever way around the increase of model structural uncertainty due to the use of potentially noisy and unstable climate predictors, is ensemble averaging or model combination (Ajami *et al.* 2006; Devineni *et al.* 2008). Model combination involves pooling the probabilistic forecasts associated with the multiple component models that form the prediction ensemble. Alternatively, it can be thought of as a weighted average of the predicted conditional probability distributions of the responses by the various component models, or an importance sampling of stochastic predictions by component models with the importance (or weights) being specified to maximize the averaged forecast accuracy.

Model combination, as described above, can be static or dynamic. Static model combination involves specifying the weights or importance of individual component models in a static, or time-unvarying manner. Dynamic model combination, on the other hand, allows individual component models to assume greater or lower importance at each time step. The changes in component model weight are ascertained based on a combination of the persistence characteristics of the weight time series, or any exogenous climate information that may be found relevant.

This paper reviews recent development on the use of model combination in a static and a dynamic manner, with reference to medium-term forecasting of flows at multiple point locations in a catchment. The first of the two approaches presented illustrates the workings of a static model combination approach, the model combination resulting in stochastic sequences of 12 month long monthly flows at multiple point locations in the catchment, with individual constituent models that form the forecast ensemble being specified using a range of past sea surface temperature anomaly variables and prior flows. The specific case study for which results are presented here is the Hydro-Tasmania hydroelectric power generation system, where the relevance of good probabilistic forecasts is invaluable in the context of power allocation and commitment. The second of the two approaches presented illustrates the workings of a dynamic model combination approach, the end output being an expected valued forecast of flows at multiple nodes of a semi-arid Australian catchment, the forecasts being issued based on a dynamically combined sea surface temperature forecast, from which the flows are derived. It should be noted here that the second approach differs from the first in more ways than the type of model combination used. The dynamic combination model response is ascertained as an expected (or mean) value representing the current time step, instead of the sequence of flows for a length of 12 months that are issued in the static combination procedure. Additionally, as concurrent forecasts are issued using the dynamic combination model, a separate procedure must be adopted to forecast the predictor field (gridded sea surface temperature anomalies) to the current time step.

The outline of the paper is as follows. The next section presents the rationale behind the Multiple Reservoir Inflow Forecasting System, the ensemble averaging approach that is adopted, and the results attained from its application to medium- to long-term forecasting of the Hydro-Tasmania multiple reservoir system. This is followed by the dynamic model combination approach, applied first to predict a globally distributed field of sea surface temperature anomalies based on projections from a mixture of dynamical and statistical climate models, and the second to predict multi-site flows in a semi-arid catchment in Australia using the predicted sea surface temperature anomaly field. This is followed by a discussion of the ensemble combination approaches discussed, along with suggestions for improvements to follow.

STATIC MODEL COMBINATION IN A MULTIPLE SITE STREAMFLOW PROBABILISTIC FORECASTING SYSTEM

Multiple reservoir inflow forecasting system

As outlined earlier, there is merit in combining simulations from multiple models where model structural uncertainty is expected to be high. We present our first alternative for model combination in this section, for use in formulating a medium to long term probabilistic forecasting system for multi-variable responses. The structural uncertainty in this forecasting system results from the uncertainty in selecting alternate predictor variable sets useful in formulating the forecasts desired, the term structural uncertainty not referring to the form (linear or nonlinear) of the model, but of the predictor variables that are used. The multi-variable response being modeled here is reservoir inflow at multiple nodes of a large water supply/hydro-electric power generation system, and the predictor variables considered in formulating the model are lagged reconstructed sea surface temperature anomalies available over a globally distributed grid (Kaplan *et al.* 1998), along with commonly used climate indices representing dominant modes of climate variability, and lagged representations of the response variables being modeled. To add to the complexity of the modeling problem, the multivariable responses are to be predicted as stochastic sequences for a lead time of four seasons (one year), implying that the probabilistic forecast issued at any point of time consists of multiple realizations of the multi-variable response set for a length of four time steps.

The presence of a high number of plausible predictor variables leads to the possibility of identifying spurious predictor variables and predictor variable sets. To get around the high structural uncertainty that is introduced as a result, the approach adopted consists of formulating multiple predictive models which include the multiple predictor sets that are identified. The steps involved in the selection of predictor variables from sea surface temperature anomalies and catchment indicators are as follows:

1. For each individual response variable being modeled, to represent temporal dependence in the ensuing probabilistic forecasts, the first predictor (Markov order-1 predictor) is selected as the preceding response value. Use of the previous response imparts a Markov order-1 dependence in the forecasts issued. Denoting the ith response as Y_t^i, the predictive model now assumes the form $Y_t^i \equiv Y_t^i | Y_{t-1}^i$. Note that "$t$" here represents a season, and "$t-1$" the previous season of the year, under the assumption that each response exhibits Markov order-1 dependence.
2. The second-order predictors are selected from climate indices, sea surface temperature anomalies, and lagged representations of the response, using a nonparametric measure of partial information (akin to partial correlation but in a nonlinear setting) known as the Partial Mutual Information (PMI) (Sharma 2000*b*). Consideration is also given to any nonstationary in the standard error associated with the data being used, by weighting individual observations that form the PMI on the basis of their associated standard error. To account for the fact that there is considerable uncertainty in selecting a unique predictor variable for the response being studies, a set of five predictors are identified conditional to them exhibiting a cross-PMI score lower than a pre-specified threshold. The predictive model now becomes

$$Y_t^i \equiv Y_t^i | [Y_{t-1}^i, X_t^{i,j}] \tag{1}$$

where $X_t^{i,j}$ denotes one of the five predictor variables $\{X_t^{i,1}, X_t^{i,2}, X_t^{i,3}, X_t^{i,4}, X_t^{i,5}\}$, which represent lagged sea surface temperature anomalies, climatic indices, or past responses, with locations (of the SSTAs) and lags being selected using the PMI criterion, conditional to the $\mathrm{PMI}(X_t^{i,j}, X_t^{i,(-j)} | Y_t^i, Y_{t-1}^i)$ exceeding a fixed threshold. In the results presented later, a threshold equivalent to a nonlinear partial correlation of 0.5 was used. Readers should note that too high a threshold can result in four unique second-order predictors not being identified.
3. Higher-order predictors are selected using the PMI, based on the order-1 common predictor, and the five order-2 predictors that are identified. Predictor identification continues until a maximum of four predictor variables have been identified. At this stage, five predictor sets for modeling the same response have been formulated, and represent the uncertainty in selecting the drivers for the response being studied. Note that while it is assumed

that the use of pseudo-independent order-2 predictors leads to near-independent predictor sets, one often finds the variables included in the alternate sets that are identified as common (the result of which is the assignment of a zero model selection weight, see step (6) below, for one of the two common constituent models). The five predictive models now appear as

$$
\begin{aligned}
&M_1 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,1}]\\
&M_2 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,2}]\\
&M_3 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,3}]\\
&M_4 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,4}]\\
&M_5 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,5}]
\end{aligned}
\qquad (2)
$$

where predictor vectors $X_t^{i,j}$ consist of at most 3 variables (making the full predictor vector 4 variable long), with the first of these being the second-order predictors $X_t^{i,j}$ identified in step 2. It should be noted that one often ascertains redundant predictors through the stepwise selection procedure, which need to be identified and discarded, as described in step (5).

4. The identification of the 5 predictor sets in step (3) is performed for (a) each response variable in the multivariable response vector being modeled, (b) each season of the year, and (c) a lead-time of 1 to 4 seasons ahead. Identification of the predictor vector for lead-times greater than one season still uses the lag-one response variable as the order one predictor. However, the higher-order predictor variables have lags greater than the forecast lead-time being considered.
5. In addition to the above, each predictor variable is associated with an influence weight (Souza & Lall 2003; Mehrotra & Sharma 2006) that dictates the relevance of the predictor variable in modulating the response, this influence weight collapsing to zero when the partial utility of a variable is negligible. Use of this influence weight helps eliminate the predictors that are redundant, forcing the predictor vector to assume a length shorter than the length of 4 formulated in step (3).
6. The last step involves the identification of the probability of selecting any one of the five predictor sets (or constituent models) that are being used to formulate the combined probabilistic forecast. This probability is estimated based on the performance of the model averaged forecasts in a leave-one-out cross-validation setting using a nonlinear constrained optimization algorithm with an objective of minimizing the leave-one-out cross-validation forecast mean square error. A leave-one-out cross-validation forecast error here refers to the difference between the observed and the forecast value when the forecast model has been developed using all data except the observation that is being forecast. The final predictive models now appear as

$$
\begin{aligned}
&M_1 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,1}];\ P(M_1) = p_1\\
&M_2 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,2}];\ P(M_2) = p_2\\
&M_3 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,3}];\ P(M_3) = p_3\\
&M_4 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,4}];\ P(M_4) = p_4\\
&M_5 \equiv Y_t^i|[Y_{t-1}^i, \boldsymbol{X}_t^{i,5}];\ P(M_5) = 1 - p_1 - p_2 - p_3 - p_4
\end{aligned}
\qquad (3)
$$

where p_j represent the probabilities with which each respective model can be selected for each probabilistic forecast issued. It should be noted that these probabilities are static, or, they do not evolve with time (hence the term "static combination"). In contrast, in the dynamic combination approach described in the next section, these model probabilities (or weights) are allowed to change with time, using a predictive procedure based on the values of the weights at preceding time steps.

Once the model predictor sets for all responses, seasons and lead times have been identified, the tasks that remain are (a) the formulation of the constituent model that uses these sets and generates the probabilistic forecasts needed, and (b) a procedure that ensures that the responses being modeled exhibit the spatial dependence expected across them. The first of the two tasks uses nonparametric kernel methods for formulate a conditional probability density function (PDF) estimate for each response variable. Details on the nonparametric probabilistic forecasting model are available in Sharma (2000*a*), and are not repeated here for brevity. The only difference here in contrast to the above referenced study is that instead of the response being conditionally generated using a defined conditional PDF, one of the five predictor sets that define the conditional density are first selected randomly as per their associated probabilities discussed in step (6) of the predictor selection algorithm above. This is then followed by the conditional generation of the response using the selected predictor set and associated conditional PDF. The generated value then becomes the order one predictor for the probabilistic forecast for the next time-step, leading to a stochastic replicate representing the response for the oncoming four seasons from the time it is issued.

Dependence across the multiple responses being modeled is maintained using a procedure proposed by Wilks (1998), which involves seeding the conditional PDFs for each of the multi-variable responses though uniform random numbers that exhibit a spatial dependence structure (correlation matrix) that is ascertained using trial and error. More details on this approach can be obtained from Wilks (1998) or Mehrotra & Sharma (2009).

The probabilistic forecasting system outlined above has been used to probabilistically forecast reservoir inflows at multiple nodes of a water supply or hydro-electricity generation system, and is referred to as the multiple reservoir inflow forecasting system (MRFLO). An application of MRFLO to a multiple-reservoir hydro-electric system in the state of Tasmania, Australia, is presented next.

Application

The MRFLO was applied to predict inflows across multiple notes of the hydro-electric generation system in place in the state of Tasmania in Australia. The need for the probabilistic forecasting system here was in the context of allowing Hydro-Tasmania (the organization that manages the system) an insight into the power they can generate in the coming 12 months, thereby enabling them to use a risk-based procedure for committing electricity supplies to potential users in mainland Australia. The main water supply regions over which the reservoirs are located, along with the distribution of the annual rainfall they receive on average, are represented in Figure 1. Most of the rain for Tasmania occurs in the Winter and Spring seasons, with the Summer (DJF) rainfall being about one-third of the Winter rain.

The spatial cross-correlations evaluated on a seasonal basis across the six catchments in Figure 1 range from 0.5 to 0.97, with the cross-correlations between the catchments excluding the Great Lakes (number 3 in Figure 1) catchment falling in the high 0.8–0.97 range. It is felt that this is a result of the Great Lakes catchment having a smaller frontal rain contribution, which is a more dominant rainfall mechanism in the other five catchments being modeled.

As a result of the above observations, it was deemed appropriate to form two aggregate response variables representing the entire Tasmania reservoir system, these being denoted as “WCD” (catchments 1, 2, 4, 5, 6) and CGL (catchment 3) in the remainder of this paper. These aggregate variables form the multivariable response vector described in the algorithm described in the previous section.

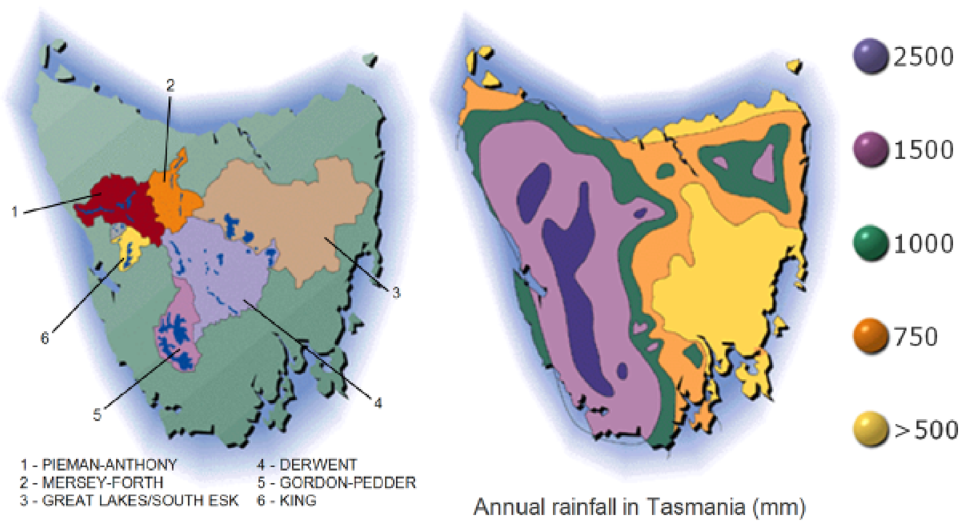

Figure 1 | Main water supply catchments for the Tasmanian hydroelectric system, along with their associated rainfall characteristics. Note that the two response variables represent aggregate inflows for catchment 3 (CGL) and rest (WCD).

The probabilistic forecasts were issued from four defined start points in the year (March, June, September, December) for the next 12 months. Predictions were issued on a seasonal basis. The results presented next are obtained in two application settings: leave-one-out cross-validation for inflow data from 1924 to 1998, and pure forecasts for inflow data between 1999 and 2008, pure forecasts referring to forecasts issued by a model that was developed using data prior to 1999. The pure forecasts were issued using the model that was developed based on data until 1998, hence representing a situation analogous to one in which the model can finally be used. If one were to assume that the predictions for the period represented by the pure forecasts will be of similar quality to the pre-1999 period, one should expect the model to perform similarly in both segments. All the results presented next are based on the issuance of 100 realizations, each of them 12 months long, from the four starting points indicated before. Results are assessed on the forecasting models ability to simulate the correct spatio-temporal dependence structure, as well as reduce the predictive uncertainty through the use of the additional climate information utilized in issuing the probabilistic forecasts.

Results

As mentioned before, predictor variables for the probabilistic forecasting models were selected separately for (a) each response (WCD and CGL), (b) each season (MAM, JJA, SON, DJF), and (c) each forecast lead time (lead times of 1 season to 4 seasons ahead). A total of five predictor sets were formulated to account for structural uncertainty, and the order-1 predictor defined as the lag-1 response being modeled. In general the predictors for the eastern response variable (CGL) included the major modes of climatic variability (El Niño Southern Oscillation or ENSO as characterized by the Southern Oscillation Index, the NINO3.4 index, or the sea-surface temperature anomalies for grid cells that fall in the NINO region), along with those at locations far away from the response and representing long lag times with respect to the current. Predictor variables for the western response (WCD) did not include ENSO (except for one predictor set in summer), and represented longer time lags from the current (4 to 6 years). As has been mentioned before, it was difficult assigning a physical meaning to the predictors, especially because they represent a modulation of the response after taking into consideration of the leading predictor variables, adding to the justification of the ensemble averaging approach that was adopted. We have refrained from presenting a list of predictors here because each predictive model (for a given response, season and lead time) can have a maximum of 20 predictors (five constituent models having four predictor variables each, of which the first variable is common), implying the presentation of the full predictor list will occupy much more space than is acceptable in a journal article. It should, however, be noted that the procedure of assigning influence weights and model probabilities, in general, resulted in the number of constituent models (maximum 5) as well as the number of predictors per model, reducing substantially as the forecast lead time became higher. Similarly, in many cases while order-2 predictors were forced to be different amongst each other, the order three predictors identified were often the ones that were selected as order-2 predictors for a different constituent model. When this was noted, the procedure automatically assigned a zero model selection probability for one of these two identical models.

Unless otherwise noted, the results below represent attributes of the probabilistic forecasts in a leave-one-out cross-validation setting (Tables 1 and 2, and Figure 2), with a concluding set of results (Table 3) provided for the pure forecast case to compare the similarity of the leave-one-out cross-validation with the pure forecasts.

A key advantage of using the probabilistic forecasting model described in the previous section is the ability to generate sequences that are capable of exhibiting the spatio-temporal dependence attributes present in the historical record. Without this in place, the use of the generated forecasts for water resources simulation and risk assessment would lead to significant biases in the policies developed.

The order-1 spatio-temporal dependence attributes of the probabilistic forecasts issued in February for the coming four seasons in a cross-validation mode are presented in Table 1. As can be noted from the table, the spatial and temporal dependence attributes in the probabilistic forecasts, correspond well with the observations, lending credibility for their use in a full-scale water resources simulation exercise.

The close representation of spatio-temporal dependence attributes leads to high skills when the forecasts are

Table 1 | Lag-1 autocorrelation and spatial cross-correlation attributes of the observed data (denoted "H" or "HIST") and the probabilistic forecasts issued in February (denoted "S" or "SIM"). Note "L1-2" implies "Lead 1-2" or a lead time of 3 or 6 months from the current. While the agreement between the observed and generated autocorrelation attributes is good, there is a greater deviation between the two for high cross-correlation in Winter (JJA). This deviation could be a result of poor modeling of the Wilk's (1998) spatial dependence approach, or because of the fact that lead-2 statistics depend strongly on the accuracy with which the lead-1 values are generated. Similar agreements were noted for the other three time-points the forecasts are issued from (results not presented)

Lag 1 Correlation

	L1-2	L2-3	L3-4	
H-WCD	0.22	0.38	0.1	
H-CGL	0.58	0.47	0.1	
S-WCD	0.21	0.36	0.05	
S-CGL	0.56	0.51	0.04	
Cross Correlation				
	L1	**L2**	**L3**	**L4**
HIST	0.31	0.64	0.45	0.23
SIM	0.34	0.38	0.42	0.25

Table 2 | The Nash–Sutcliffe Coefficient of Efficiency for individual and aggregated responses. The first set of results (top 8 rows) represent forecasts issued in February of each year. Note that the values marked "L1-3" represent probabilistic forecasts that have been aggregated to represent lead times of 1 to 3 seasons (flows for MAMJJASON). Note that the efficiencies reported above were ascertained by aggregating each of the 100 replicates issued as the probabilistic forecasts, and then comparing these aggregated values with those observed. Also note that the last 4 rows represent aggregated full system forecast efficiencies for forecasts issued in May, August and November respectively

Nash-Sutcliffe Coefficient of Efficiency – February

	L1	L2	L3	L4
WCD	0.47	0.57	0.53	0.49
CGL	0.56	0.59	0.49	0.51
	L1	**L1-2**	**L1-3**	**L1-4**
WCD	0.47	0.6	0.6	0.6
CGL	0.56	0.62	0.62	0.6
Both	0.58	0.65	0.63	0.62
Nash-Sutcliffe Coefficient of Efficiency – Other start months				
	L1	**L1-2**	**L1-3**	**L1-4**
May	0.55	0.46	0.48	0.49
August	0.44	0.51	0.47	0.51
November	0.40	0.57	0.50	0.42

aggregated over multiple lead times, or over space. Table 2 presents the Nash–Sutcliffe Coefficient of Efficiency (denoted $\mathrm{NS}=1-\frac{\sum (y_t^i-\hat{y}_t^i)^2}{\sum (\bar{y}_t^i-\hat{y}_t^i)^2}$, y_t^i, $\hat{y}_t^i$ and $\bar{y}^i$ representing the observation to be forecast, the expected probabilistic forecast and the historical mean for the current season and response variable i) of the probabilistic forecasts for individual lead-times and when aggregated over time and space. A NS score approaching 1 denotes a "perfect" forecasting model, and one close to zero (or negative) implies the model does not perform any better than the use of the historical mean as the forecast. While detailed results are presented for forecasts issued in February, results for the aggregated system inflow are presented for the other forecast start points (May, August, November).

As can be noted from the results in Table 2, the probabilistic forecasts are able to exhibit high efficiencies even when aggregated over time and space (for instance, note the one-year accumulated forecast efficiency of 0.62 for forecasts issued in February). Note that this would not be possible were the spatial and temporal dependence attributes not modeled acceptably.

Note that while the results illustrated in Figure 2 and Table 2 point to a high accuracy of the probabilistic forecasting system, certain limitations are also visible. The main limitation that can be spotted from the results in Figure 2 pertain to an overestimation of low observed flows and an underestimation of the high observed flows. This bias is a result of the ensemble averaging procedure that is adopted, which also results in wider prediction intervals associated with the forecasts, as compared to what would be obtained were a single model used. Readers should note that were a single forecast model used, the result would have been lower bias in the forecasts, but a significantly greater variance.

The best verification of a forecasting technique is its evaluation by the end users over time. The above described procedure was developed in 2005 using observed data until the end of 1998. The remaining data was set aside to validate the performance of the proposed approach. The results from this validation (using observed flows from 1999 to 2008) are presented in Table 3, for each of the four forecast starting months, and for lead times of 1 to 4 seasons. Aggregated results (representing accumulated values for longer time periods are not shown, but represent similar levels of accuracy as presented for the leave-one-out cross-validation

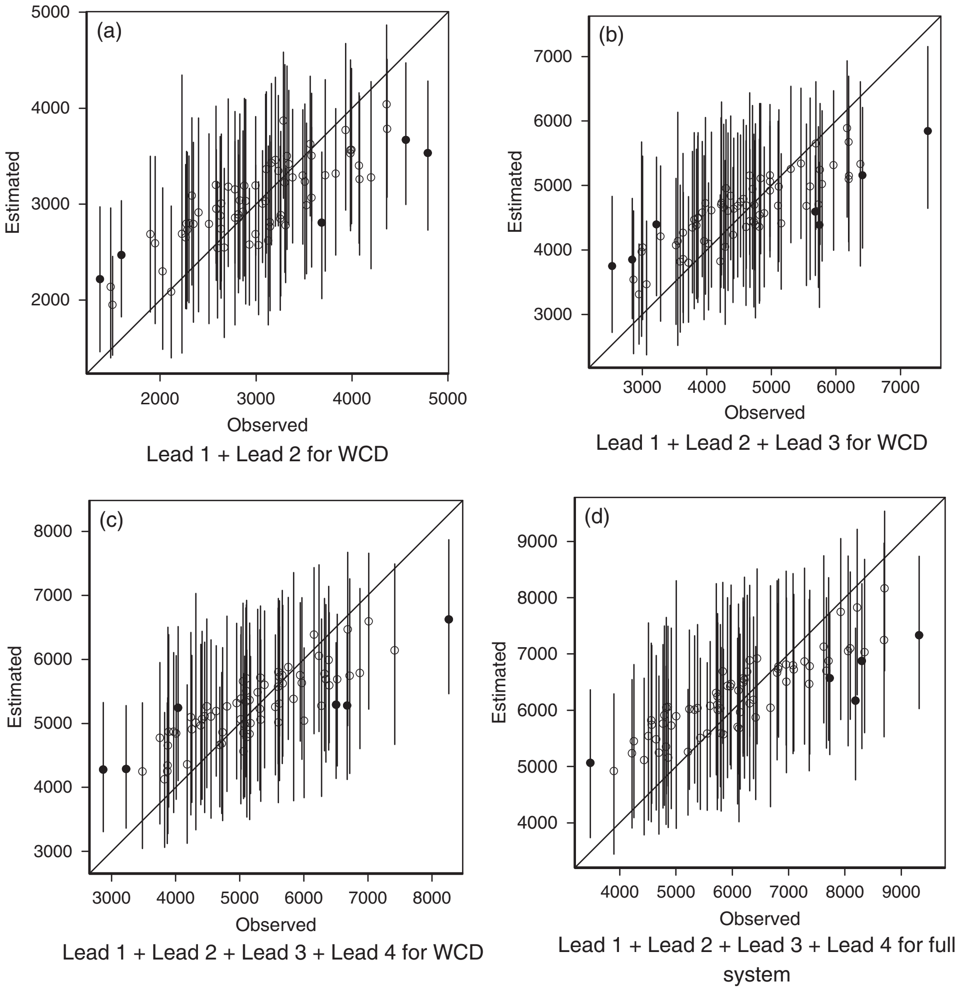

Figure 2 | Aggregate 3–12 month probabilistic forecast for Western and Central catchment inflows (WCD) and 12-month aggregate probabilistic forecasts for the entire system (in units of $12 \times m^3\ s^{-1}$). Note that these forecasts are issued in February, and represent leave-one-out cross-validation results. The vertical lines represent the 10th to 90th percentiles of the probabilistic forecasts issued. In cases where the observed flow falls within these limits (the line intersects the 45 degree line) the observed flow value is indicated by an open circle. In cases where this is not so, it is indicated by a closed circle.

results in Table 2. As can be noted, the model performs consistently well across the pure forecast period, with an efficiency that is statistically similar to that obtained in cross-validation. It should be pointed out that this similarity in the efficiencies is a result of the ensemble averaging procedure adopted, with significant variations in the efficiency that would be attained were an individual model used.

Discussion of the static model combination approach

This section presented an alternative for issuing medium- to long-term probabilistic forecasts at multiple nodes of a reservoir system up to a year ahead. These forecasts were issued as stochastic sequences of length 4 seasons (1 year), with predictors for each of the season selected from commonly used climate indices as well as globally distributed reconstructed sea surface temperature anomalies. An important aspect of the forecasts was the approach adopted to mitigate the high uncertainty present in any forecasting procedure that is formulated based on a large number of possible predictor variables. This uncertainty leads any such high-dimensional predictive model to exhibit high variability in their expected forecasts when applied to new data that was not part of the model specification process. To circumvent

Table 3 | Nash–Sutcliffe Coefficient of Efficiency for the full-system inflow in a pure-forecast setting, for the four starting months forecasts are issued from. It is notable that the efficiencies reported here are similar to those attained for the leave-one-out cross-validation results in Table 2, although the variability in the pure forecasts is more given the smaller sample size used. Note that the above results have been ascertained using observed flow data from 1999 to 2008

Nash-Sutcliffe Coefficient of Efficiency

	L1	L1-2	L1-3	L4
WCD				
February	0.85	0.11	0	0.66
May	0.28	0.36	−0.1	−0.3
August	0.49	0.35	0.36	0.27
November	0.51	0.65	0.63	0.45
CGL				
February	0.91	0.85	0.7	0.91
May	0.81	0.56	0.65	0.55
August	0.3	−0.2	−0.2	−0.3
November	0.75	0.49	0.45	0.22

this high predictive variance, five such models were formulated, and their results combined using a weighting procedure. The combination of these individual predictive models led to greater stability in the model predictions of the future.

Two additional points are important to note:

1. The use of the ensemble averaging procedure, while reducing the structural uncertainty that would otherwise be present, created an additional problem of introducing a bias in the estimation of the low and high flows across the system. This bias was a result of the use of multiple predictions, each with their own associated variability in the expected mean state being forecast. The averaging of these multiple predictions tended to pull the average ensemble forecast towards the mean of the data, resulting in an overestimation of the low flows and an underestimation of the high flows. Additionally, the uncertainty (or prediction intervals) associated with the probabilistic forecasts was greater than that associated with individual models being combined, as a result of the reduction in the structural uncertainty in the ensemble averaged forecast.
2. The ensemble averaging procedure reported above combined individual models without changing the associated probabilities in time. As a result, the weight associated with any single model was the same, irrespective of its performance at predicting high versus low flows, or predicting flows during an El Niño event versus, say, a La Niña event. This limitation has the potential to reduce the overall accuracy of the probabilistic forecasting system developed, especially because each of the individual forecast models tend to exhibit high accuracy in spurts for a sustained period of time. An approach that allowed for the model weight to change depending on the past accuracy of the forecast model used, could potentially address this limitation, and also help reduce the bias noted in the low and high flow tails. The next section presents another alternative for model combination that allows model weights to vary over time.

DYNAMIC MODEL COMBINATION FOR CLIMATE AND HYDROLOGIC PREDICTION

The previous section described an alternative for reducing structural uncertainty in medium- to long-term streamflow forecasts by formulating multiple constituent forecast models, and combining them to form an ensemble average. The model combination that was described was termed a static model combination, as the weights or selection probabilities associated with individual constituent models were not allowed to vary with time. In contrast to this, the model combination approach presented here represents a *dynamic* model combination, where individual model weights or selection probabilities are allowed to vary with time.

Some of the key differences between the model combination approaches presented in the previous section and the one presented here are as follows:

1. The MRFLO probabilistic forecasting alternative described previously issues stochastic sequences spanning multiple time steps from the time they are issued, whereas the forecasts here represent the expected forecast for a defined lead time. Furthermore, the probabilistic forecasts using MRFLO were issued based on predictors that represented lagged response and climate variables. Here, the forecasts occur in two stages, the first stage being a prediction of the climate variables to the next time step, and the second being an estimation of the streamflow based on the predicted climate. As a result, the dynamic combination presented here represents an indirect forecast of the

response variables (streamflow), whereas the static combination in the previous section presented a direct forecast of the multivariable flows.

2. The constituent models in the static ensemble averaging approach described in the previous section, were all developed separately using predictor sets that were distinctly different across the models being combined. The dynamic model combination approach presented in this section uses pre-developed models (analogous to a "black-box" or a model that cannot be altered or re-calibrated) for climate forecasts for stage one of the two-stage approach.
3. The main difference between the two approaches is the manner in which constituent models are combined. As stated before, the probabilistic forecasting approach of the previous section uses static model weights or selection probabilities, whereas the approach presented here uses weights that change with time.

As outlined above, the dynamic model combination is constructed in two stages. The first stage represents the use of the dynamic combination to forecast sea surface temperature anomalies distributed across the global ocean surface (Chowdhury 2009), while the second stage uses dynamic combination to formulate multiple predictive models of streamflow for a semi-arid region in Australia using the predicted sea surface temperature anomaly field (Chowdhury & Sharma 2009*a*). Details on each of these are presented in the sections below.

Dynamic model combination for medium-term global SSTA forecasts

As mentioned, the first stage of the dynamic combination logic presented here consists of a dynamic combination of multiple sea surface temperature anomaly (SSTA) forecast models. The forecasts are constrained to SSTA at a 5^ox5^o grid spanning the region 60^o N to 40^o S. The basis for the anomalies here is the Global Ocean Surface Temperature Atlas (GOSTA) from 1951 to 1980 (Bottomley *et al.* 1990), and the specific models combined are the (a) ECM model, a dynamical forecast model developed by the European Centre for Medium Range Weather Forecast (ECWMF) as part of the DEMETER project described in Wolff *et al.* (1997); (b) MetF, also a dynamical model originating from the DEMETER project but developed at Meteo-France (Madec *et al.* 1997); and (c) CPC, a statistically based predictive alternative originating from the Climate Prediction Centre of the National Oceanic and Atmospheric Administration, USA (van den Dool 2000), developed using a statistical technique known as the constructed analogue to forecast SSTA as linear combinations of all past observations for the current month. The common period of hindcasts (model simulations of the past or the observed climate) available for model formulation and evaluation was March 1958 to December 2001.

The dynamic model combination logic employed to forecast the distributed SSTA field three months ahead is summarized as follows:

1. Unlike the static model combination procedure described before, here constituent models are combined on a pairwise basis. The basis of choosing the constituent models to be paired is the dissimilarity in their respective outputs, the premise being that the greatest improvements will be when the most dissimilar alternatives are combined. The pairwise combination adopted here was [ECM + MetF] + [CPC]. Readers are referred to (Chowdhury 2009) for additional details.
2. For the two models to be paired, the first step is to compute the exact combination weight at each time step such that use of these weights results in a perfect forecast. Once the weight time series has been formulated, this becomes the basis for developing a predictive model for the dynamic weight.
3. The predictive model for the dynamic weight is formulated as an autoregressive approach, with lagged weights serving as predictors of the response being modeled. Readers are referred to Chowdhury & Sharma (2009*a*, *b*) for details on the specific formulation of the autoregressive model used. It should be noted that the predictors of the autoregressive model were identified using standard approaches such as the partial autocorrelation function of the observed weights, along with nonlinear formulations such as the partial mutual information (PMI) (Sharma 2000*b*). Note that while the method allows the inclusion of any identified exogenous predictors in the modeling structure, for simplicity the model did not use any exogenous predictors.
4. Spatial dependence across the SSTA field was not explicitly modeled. Instead the approach adopted involved

formulating a predictive model for each grid cell individually, allowing for weightings for specific regions to be different and representative of how those regions are modeled by the constituent models being combined. The predictor values from a grid cell to the next cell maintained a smooth transition that helped impart local dependence. It was found that this approach served to approximate historical dependence to an acceptable extent (Chowdhury 2009).

The improvements in the dynamically combined forecasts over the best performing constituent model are illustrated in Figure 3. The MetF is chosen for comparison due to its error variance being lower than that of CPC and ECM. The figure shows a considerable improvement of the SSTA forecast post combination; with most of the region returning a 25% to 75% reduction in mean squared error. This dynamically combined SSTA field now forms the basis for estimating flows at multiple locations in a semi-arid catchment in Australia. Details on this second stage of the predictive approach presented are discussed next.

Dynamic model combination and concurrent estimation of multisite streamflow

The second stage of the dynamic model combination approach presented involves using the SSTA forecasts described in the previous section, and formulating statistical models for estimating streamflow at multiple locations of the Namoi River catchment (total area 42000 km^2), located roughly 450 km north-west of Sydney in eastern Australia. The response variables being modeled represent five locations of the Namoi River catchment system, with high levels of spatial dependence (seasonal correlations in excess of 0.6), and markedly seasonal characteristics including significant segments of zero monthly flows (leading to coefficient of variations up to 2.9 with the mean flow being up to 10 times the median flow). Three statistical approaches are used to ascertain the seasonal flow response vector based on the concurrent (predicted) SSTA field, along with lagged representations of the response being modeled. A short description of the three statistical alternatives is provided below, with readers referred to Chowdhury & Sharma (2009*a*) for the detailed mathematical formulation used.

A. Generalized linear autoregressive model using SSTA principal components (GLM): The rationale behind the GLM approach for estimation of multisite flows across the Namoi catchment, was the use of functions of the first two principal components (Hastie *et al.* 2001) of the predicted SSTA field, along with selected lags to the aggregated flow across the five locations, as predictors. The predictive model used was a combination of a

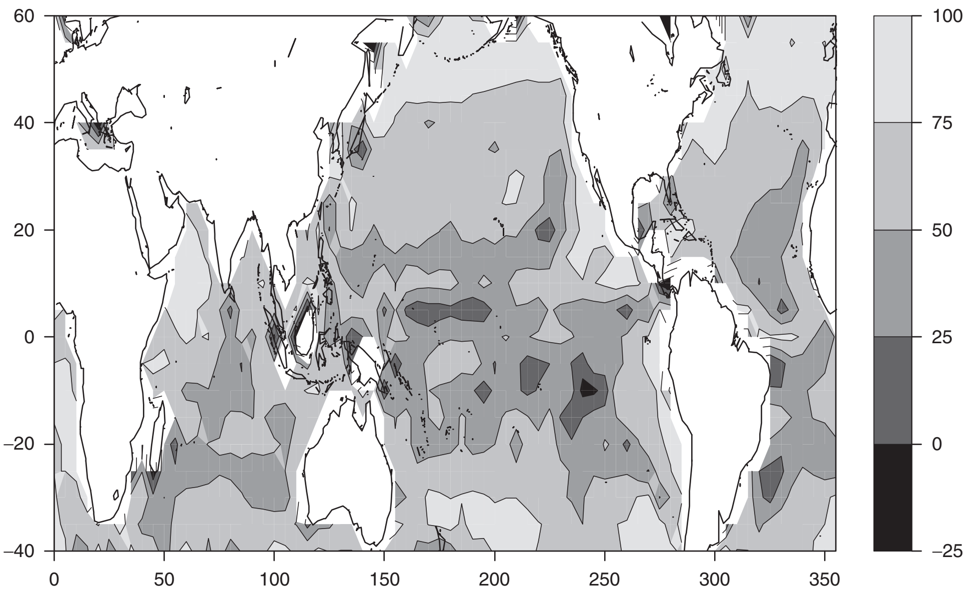

Figure 3 | Percentage reduction in mean squared error of dynamically combined ssSSTA forecast compared to that of single best SSTA forecasting model (MetF) (adapted from (Chowdhury & Sharma 2009*a*)).

Generalised Linear Model for predicting the aggregate response, and a multinomial model for disaggregating the aggregated flow to the five components. Additionally, a log-transformation to reduce the skewness associated with the aggregated response was employed first, and the GLM approach developed to predict the transformed flows.

B. Local polynomial regression using SSTA independent components (ICM): Independent component analysis (ICA) is a different approach compared to principal component analysis (PCA) for feature extraction in high dimensional data fields such as SSTA (Westra *et al*. 2007). The rationale behind ICA is to find independent components (ICs) such that they are orthogonal as well as independent to each other. While orthogonality is an expected attribute of PCs too, PCA formulates its components to ensure zero correlation, which does not automatically transform to independence. More details on ICA and how it has been used to forecast multivariable series is presented in (Westra *et al*. 2007, 2008). A good text that includes details on both PCA and ICA is (Hastie *et al*. 2001). The ICM model used here utilized three ICs of the residuals derived by formulating an autoregressive order one model for the individual site flows, which were regressed against three ICs of the predicted SSTA field. It should be noted that the ICM approach represented a significantly different modeling structure and associated predictor variable field as compared to the GLM approach described above.

C. Modified nearest-neighbor sampling using pre-selected SSTA regions (KNM): This approach involves using pre-selected regions in the SSTA field which exhibit high dependence with the responses being modeled, along with lagged responses to form the complete predictor field. The actual predictive model is a variation of the knn resampling algorithm by Lall & Sharma (1996) with importance weights for individual predictor variables as per Mehrotra & Sharma (2006).

It should be noted that the three statistical options to model the multivariable response were selected to ensure markedly different structural forms, as well as different predictor fields. The rationale behind this was to ensure each of the modeled responses will exhibit differences, creating a situation where model combination will lead to advantages over any single model.

Table 4 presents the outcomes of the constituent statistical models, along with their dynamic model combination. Results are also presented for the case where the SSTA field originates from the best performing SSTA prediction model (MetF), along with the case where the dynamic model combination is used in deriving the SSTA field. From the results it is evident that (a) the use of dynamic combination in both the SSTA and the flow modeling results in the best predictive performance amongst all cases compared, and (b) the improvement in the modeled flows using the dynamic model combination versus the better performing statistical options is small (but significant). Readers are referred to Chowdhury & Sharma (2009*a*, *b*) for details on the model evaluations including the seasonal and temporal dependence representations achieved using the dynamic combination approach.

Table 4 | Error variance of the transformed flow forecasts for the five dimensional response (adapted from Chowdhury & Sharma 2009*a*). The transformation applied here is the standardized log of non-zero flows. The final column shows the analysis where all the variables are pooled together. Note that the results presented here are evaluated in a cross-validation setting and hence the best performing models are the ones associated with the lowest error variance, irrespective of the number of parameters they contain

SSTA	Models	Man R	Namoi R	Peel R	Mooki R	Coxs C	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

Discussion

A question that arises from the results presented in the previous sections is, why do we observe significant improvements in the SSTA forecast as a result of dynamic combination, when equivalent improvements are not present when the flow estimation models are combined? To answer this issue we need to evaluate what exactly are we combining in both the cases mentioned. A pairwise scatter plot of the residual errors associated with the three SSTA forecast models and the three flow forecast models is presented in Figure 4. As is clear from the figure, the correlation (or similarity) between the flow residuals being combined is significantly higher than is the case for the SSTA residuals. This high correlation is partly due to the overbearing presence of the common order-one predictor which is the observed flow at the prior time step (the previous seasons flows), whereas no such commonality is present in the SSTA forecasts. Besides, the statistical models formulate a relationship with the same response (flow), enforcing some level of similarity in the constituent models developed. As a result, model combination of markedly different models (SSTA) leads to a stabilization of the errors when all models are combined, whereas no such stabilization is possible when the residuals are similar to each other to start off with (flow).

It should be noted though that the argument that dependence across the models being combined leads to reduced improvements in the combined output, does not consider the advantages introduced because of dynamic combination specifically, but more so the advantage because of model combination in a static sense. To evaluate whether dynamic combination is an added level of complexity that is worth introducing, depends significantly on the predictive power of the constituent model weights. In cases where the model weights exhibit a high level of persistence, or alternately a high level of dependence on an exogenous state variable (such as the NINO3.4 index), the advantages of using the dynamic combination approach should be easy to justify. It should be pointed out that the results presented here did not use any exogenous predictors for the dynamic model weights.

CONCLUSIONS

Medium-term prediction of hydroclimatic systems is fraught with significant structural uncertainty, mainly because of the

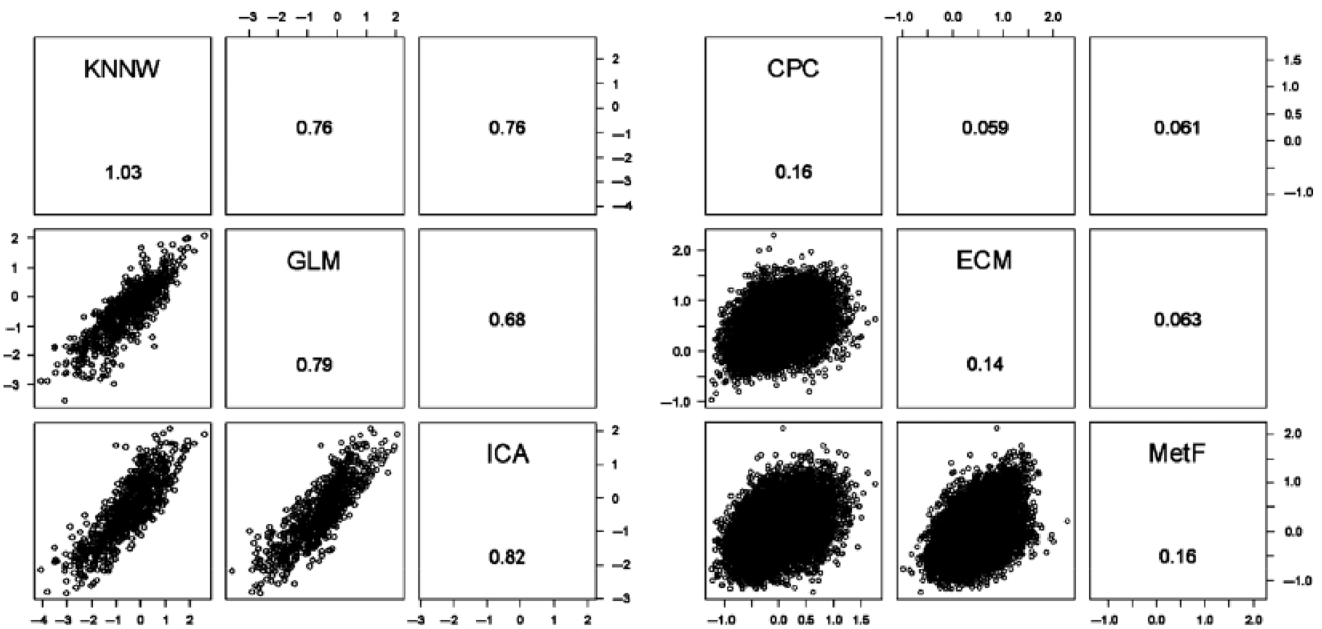

Figure 4 | Residual errors of the three component forecasts with the corresponding variance and covariance measures listed in the diagonal and upper triangle boxes (adapted from Chowdhury & Sharma 2009a). The left panel contains residual error of log flow, the right panel shows residual errors of sea surface temperature within the 175°–225°E by 10°–20°N box.

lack of an easy to formulate physical basis through which a predictive model can be defined. As a result, one can formulate multiple configurations of markedly different predictive models that all exhibit similar accuracies over the periods they are calibrated against, but high uncertainty when applied to future data. This paper summarized two alternatives for model combination as a means of reducing the structural uncertainty associated with medium-term hydroclimatic model forecasts. The first of the two approaches used a "static" model combination rationale, where constituent models were combined based on a weight or model selection probability associated with each model. The second approach allowed the model weight to vary with time, its variation being modeled using an additional predictive approach, the resulting model combination being termed a "dynamic" model combination.

Setting aside the specific differences between the examples of model combinations that were presented (such as the model outputs being stochastic sequences for the static combination, and expected values for the dynamic combination), the main advantage in the static combination over the dynamic case was the simplicity associated with it. This simplicity can be a factor in the combination choice especially when the length of the constituent model hindcasts available for defining the combination model is limited.

An additional advantage of the static combination case study was the flexibility available in defining the constituent models (and thereby choosing only those models that were somewhat independent of each other). Similar flexibility was not available in the dynamic model combination case (especially where SSTA forecasts were being combined) as the constituent models came pre-defined and developed by other research groups. It should, however, be pointed that the ideal scenario for model combination would be where the constituent models are defined keeping in mind that they will be used for combination later, as has been done in the context of streamflow simulation in a Hierarchical Mixtures of Experts framework by Marshall *et al.* (2006, 2007*a*, *b*).

We believe the examples presented here represent initial attempts at model combination in hydrology, and research in this area is likely to grow as computing power increases. Some of the directions that need further research include the issue of specifying accurate confidence and prediction intervals that are cognizant of the uncertainty introduced by the combination algorithm used, a better way in which constituent models are defined such that their use in a model combination sense leads to an explanation of additional variability in the responses being modeled, and, a better way or representing spatial dependence in multi-variable responses while still allowing model combination for each variable separately to the full vector. Work on these issues is in progress and shall be presented at a later stage.

ACKNOWLEDGEMENTS

This research was supported over the years by the Hydro-electric Corporation Tasmania, the Sydney Catchment Authority, and the Australian Research Council. The first author is grateful to Fiona Ling, James Bennett, David Wilson, Stephen Stolp, Rowan Murray and Jai Allison, from HEC Tasmania, for their support and inputs to the results presented in the static model combination part of the paper. We acknowledge the NSW Office of Water for supplying the unimpaired Namoi River flow estimates for this research. We also gratefully acknowledge the detailed comments received by the two anonymous reviewers for the journal, which greatly improved the presentation of the paper.

REFERENCES

Ajami, N. K., Duan, Q. Y., Gao, X. G. & Sorooshian, S. 2006 Multimodel combination techniques for analysis of hydrological simulations: Application to Distributed Model Intercomparison Project results. *J. Hydrometeorol.* **7**, 755–768.

Bottomley, M., Folland, C. K., Hsiung, J., Newell, R. E. & Parker, D. E. 1990 *Global Ocean Surface Temperature Atlas*. Meteorological Office, Bracknell, UK and the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA.

Chiew, F. H. S., Zhou, S. L. & McMahon, T. A. 2003 Use of seasonal streamflow forecasts in water resources management. *J. Hydrol.* **270**, 135–144.

Chowdhury, S. 2009 *Mitigating Predictive Uncertainty in Hydroclimatic Forecasts: Impact of Uncertain Inputs and Model Structural Form*. School of Civil and Environmental Engineering, University of New South Wales, Sydney. 217 pp.

Chowdhury S. & Sharma, A. 2009*a* Multi-site seasonal forecast of arid river flows using a dynamic model combination approach. *Water Resour. Res.* **45**, W10428.

Chowdhury, S. & Sharma, A. 2009*b* Long-range NINO3.4 predictions using pair wise dynamic combinations of multiple models. *J. Clim.* **22**, 793–805.

Devineni, N., Sankarasubramanian, A. & Ghosh, S. 2008 Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations. *Water Resour. Res.* **44**.

Hastie, T., Tibshirani, R. & Friedman, J. 2001 *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, New York.

Kaplan A., Cane, M. A., Kushnir, Y., Clement, A. C., Blumenthal, M. B. & Rajagopalan, B. 1998 Analyses of global sea surface temperature 1856–1991. *J. Geophys. Res. Oceans* **103**, 18567–18589.

Lall, U. & Sharma, A. 1996 A nearest neighbor bootstrap for time series resampling. *Water Resour. Res.* **32**, 679–693.

Madec, G., Delecluse, P., Imbard, M. & C. Levy. 1997 *OPA Release 8, Ocean General Circulation Model Reference Manual*, Available from LODYC/IPSL, France.

Marshall, L., Nott, D. J. & Sharma, A. 2007*a* Towards dynamic catchment modeling: A Bayesian Hierarchical Mixtures of Experts framework. *Hydrol. Process.* **21**, 847–861.

Marshall, L., Sharma, A. & Nott, D. J. 2006 Modelling the catchment via mixtures: issues of model specification and validation. *Water Resour. Res.* **42**.

Marshall, L., Sharma, A. & Nott, D. J. 2007*b* A single model ensemble versus a dynamic modeling platform: semi-distributed rainfall runoff modeling in a Hierarchical Mixtures of Experts framework. *Geophys. Res. Lett.* **34**.

Mehrotra, R. & Sharma, A. 2006 Conditional resampling of hydrologic time series using multiple predictor variables: A *k*-nearest neighbour approach. *Adv. Water Resour.* **29**, 987–999.

Mehrotra, R. & Sharma, A. 2009 Evaluating spatio-temporal representations in daily rainfall sequences from three stochastic multisite weather generation approaches. *Adv. Water Resour.* **32**.

Peichota, T. C., Dracup, J. A., Chiew, F. H. S. & Mcmahon, T. A. 1998 Seasonal streamflow forecasting in Eastern Australia and the El Nino Southern Oscillation. *Water Resour. Res.* **34**, 3035–3044.

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

Sharma, A. 2000a Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3 – A nonparametric probabilistic forecast model. *J. Hydrol.* **239**, 249–258.

Sharma, A. 2000b Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 – A strategy for system predictor identification. *J. Hydrol.* **239**, 232–239.

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

Souza, F. A. & Lall, U. 2003 Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm. *Water Resour. Res.* **39**, 1307–1320.

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 Bull.* **9**, 38–41.

Wang, Q. J., Robertson, D. E. & Chiew, F. H. S. 2009 A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resour. Res.* **45**, 1–18.

Westra, S., Brown, C., Lall, U., Koch, I. & Sharma, A. 2010 Interpreting variability in global SST data using independent component analysis and principal component analysis. *Intl J. Climatol.* **30**(3), 333–346.

Westra, S., Brown, C., Lall, U. & Sharma, A. 2007 Modelling multivariable hydrological series: Principal Component Analysis or Independent Component Analysis? *Water Resour. Res.* **43**.

Westra, S., Brown, C., Lall, U. & Sharma, A. 2008 Multivariate streamflow forecasting using independent component analysis. *Water Resour. Res.* **44**, W02437.

Westra, S. & Sharma, A. 2009 Probabilistic estimation of multivariate streamflow using independent component analysis and climate information. *J. Hydrometeorol.*

Wilks, D. S. 1998 Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.* **210**, 178–191.

Wolff, J. E., Maier-Reimer, E. & Legutke, S. 1997 *The Hamburg ocean primitive equation model*. Tech. Rep.13, Dtsch. Klimarechenzent., Hamburg, Germany.

First received 8 September 2009; accepted in revised form 7 September 2010. Available online February 2011