ICIWRM – 2000, Proceedings of International Conference on Integrated Water Resources
Management for Sustainable Development, 19 – 21 December, 2000, New Delhi, India

Seasonal to interannual hydrologic prediction: a case study from Sydney, Australia

ASHISH SHARMA
School of Civil and Environmental Engineering, The University of New South Wales, Sydney, NSW 2052, Australia

Abstract

Effective use of available water resources is a serious problem facing the world as it enters the 21st century. An increasing demand and an ever-uncertain supply in the form of precipitation is a cause of concern to water resources managers. Another concern is the occurrence of severe and sustained droughts that deplete reservoir storage to dangerous levels, forcing operators to enforce water supply restrictions. Such droughts are often associated with long-term or low frequency climatic fluctuations, such as the El Niño Southern Oscillation (ENSO).

A recently completed research project at the University of New South Wales, Sydney, evaluated the feasibility of predicting the seasonal rainfall at Warragamba dam, a large water supply reservoir 70km west of Sydney. The study had three main aims: (a) to develop a criterion that was capable of quantifying the utility of a predictor when used for probabilistic forecasts of seasonal rainfall; (b) to formulate an effective approach for making probabilistic forecasts for selected lead times from the present; and, (c) to compare the effectiveness of using atmospheric indices (such as the Southern Oscillation Index) rather than a broader suit of hydro meteorological variables for predicting seasonal rainfall. Presented here is a summary of some of the methods that were developed and the results attained.

INTRODUCTION

Effective use of available water resources is a serious problem facing the world as it enters the 21st century. An increasing demand for water and an ever-uncertain supply in the form of precipitation has always been a cause of concern to water resource managers. Another concern in the Australasian and South East Asian region is the occurrence of severe and sustained droughts that deplete reservoir storage to dangerous levels, forcing operators to enforce water supply restrictions. Such droughts are often associated with long-term or low frequency climatic fluctuations, such as the El Niño Southern Oscillation (ENSO). Reservoir operation for regions not affected by such fluctuations is relatively simple. Reservoir operators have no means of knowing when such a drought may occur, how long it would last, and how severe it may be. Were answers to such questions available, strategies to conserve water for use during droughts could be devised. This paper is part of a study to answer some of the questions raised above, with the aim of providing reservoir operators probabilistic estimates of the rainfall and reservoir inflows that could be expected in the future.

Forecasting future flows or precipitation requires an understanding of the nature and causes of climatic variability. While there have been significant advances in physically based models of the climatic and hydrologic systems in recent decades, their operational utility beyond a few days or weeks, and the spatial specificity and accuracy of their forecasts remains limited. Consequently, where long historical records of the variables of interest are available, statistical approaches that relate "at-site" hydrology to large scale ocean-atmosphere state variables could provide a basis for useful seasonal to interannual flow forecasts.

Identification of the oceanic or atmospheric variables that form useful predictors of rainfall is an important step in developing a long-term forecast model. If a linear relationship characterises the underlying system (the dependent variable being a linear function of all model predictors) linear dependence measures such as the coefficient of correlation may suffice. If the underlying system is more complex, as is usually the case with any real physical system, linear methods are likely to result in misleading predictors and a badly formulated forecast approach. A generalised measure of dependence, denoted the partial mutual information criterion, was developed to remove some of the problems mentioned above. Details on this measure of dependence are provided in section 2 of this paper.

Formulation of a prediction model that uses the variables identified as the useful predictors of rainfall, is the next step of the rainfall prediction exercise. The prediction model was developed using nonparametric kernel methods for probability density estimation. The model for developed to provide outputs in three different forms: a specified number of simulations of the likely rainfall value for use in Monte Carlo studies; a probability density function of the predicted rainfall for comparing with the unconditional or marginal probability density function; and, a cumulative probability distribution function of the predicted rainfall that indicated the quantiles (1^{st}, 5^{th}, 10^{th}, 90^{th}, 95^{th}, 99^{th}) that may be of interest in a water management context. Details on this prediction model are provided in section 3 of the paper.

The predictor identification and prediction approach were applied to quarterly rainfall data from the catchment of the Warragamba dam, a large water supply reservoir 70km west of Sydney. The Warragamba dam supplies nearly 70% of the Sydney's water, and hence is of great importance in the water management of Australia's largest city. Two sets of predictors were considered. In the first case, simplistic descriptors of an ongoing ENSO anomaly were used as the predictors of the Warragamba quarterly rainfall. Next, predictors were selected from amongst sea surface temperature anomalies averaged over 5° x 5° latitude-longitude grid cells spread over the world's ocean bearing surface. The results from the use of both these prediction scenarios are presented and discussed in section 4 of this paper.

Predictor identification using partial mutual information

Consider the following model:

$$\begin{aligned} X &= \sin(2\ \ t/T) + w_t \\ Y &= \cos(2\ \ t/T) + w_t \end{aligned} \tag{1}$$

where T equals 20, t ranges from 1 to 200, and w_t is a "noise" term chosen from a Gaussian distribution with zero mean and a standard deviation of 0.1. Figure 1 illustrates the

nonlinear relationship that exists between the two variables. A good predictor identification method should find X as a strong predictor of Y.
The sample coefficient of correlation between variables X and Y equals –0.008. If a linear model were fit between X and Y, it would thus explain close to 0% of the overall variance in Y. Even if more efficient nonlinear models were fit between these variables, they would collapse to the case of a simple linear model passing through the centroid of Figure 1. Hence, any measures of dependence that are based on fitting a regression model or a line (or curve) of "best" fit would not indicate the strong dependence between the two variables. This is a major drawback behind the use of any predictor identification criterion that is based on measures of error such as the residual sum of square of errors.

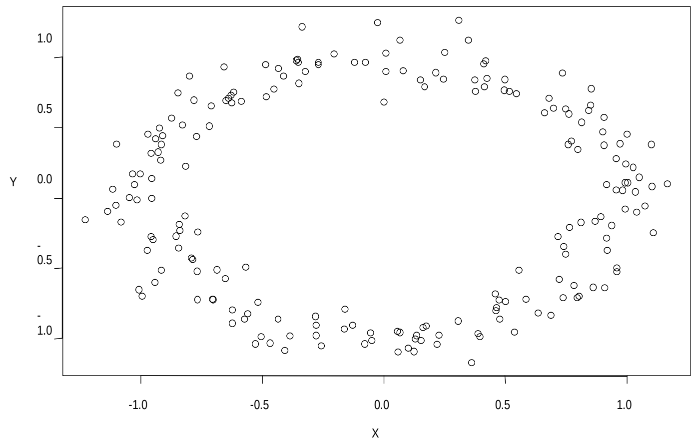

Figure 1. Two hundred data points from the model in equation 1.

The mutual information (MI) criterion [*Fraser and Swinney*, 1986] is a measure of dependence between any two variables. The MI score between X and Y is defined as:

$$MI = \iint f_{X,Y}(x,y)\log_2\left[\frac{f_{X,Y}(x,y)}{f_X(x)\,f_Y(y)}\right]dxdy \tag{2}$$

where:
$f_X(x)$ and $f_Y(y)$ are the marginal PDF's of X and Y respectively, and,
$f_{X,Y}(x,y)$ is the joint (bivariate) PDF of X and Y

The rationale behind mutual information is the basic definition of dependence. The joint probability of occurrence of two variables is equal to the product of the individual probabilities if they are independent. Hence the joint PDF $f_{X,Y}(x,y)$ would equal $(f_X(x)\times f_Y(y))$ is X was independent of Y. The MI score in (2) would, in that case, equal a value of 0 (the ratio of the joint and marginal densities being one, giving the logarithm a value of zero). A high MI score would indicate a strong dependence between the two variables.

The key to an accurate estimate of the MI is the accurate estimation of the marginal and joint PDF's in (2). Some of the earlier versions of the MI function used crude measures of the PDF such as a histogram. A more stable and efficient PDF estimator is based on the use of kernel density estimation techniques [*Scott*, 1992; *Silverman*, 1986]. The kernel density estimator adopted in this study uses a Gaussian kernel function [*Scott*, 1992], and is expressed as:

Error! Bookmark not defined. $\hat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(2\pi)^{d/2}\lambda^d \det(S)^{1/2}}\exp\left(-\frac{(x-x_i)^T S^{-1}(x-x_i)}{2\lambda^2}\right)$ (3)

where

$\hat{f}_X(\mathbf{x})$ is the mutivariate kernel density estimate of the d-dimensional variable set **X** at coordinate location **x**,

$\mathbf{x}_i$ is the i'th multivariate data point, for a sample of size n,

S is the sample covariance of the variable set **X**,

det( ) represents the determinant operation, and,

λ is a smoothing parameter, known as the "bandwidth" of the kernel density estimate

This study has used a relatively simple and computationally efficient rule for estimating the bandwidth λ, known as the Gaussian reference bandwidth λ_{ref} [*Scott*, 1992; *Silverman*, 1986].

$$\lambda_{ref} = \left(\frac{4}{d+2}\right)^{1/(d+4)} n^{(-1/(d+4))} \quad (4)$$

where n and d refer to the sample size and dimension of the multivariate variable set respectively.

Identification of multiple system predictors necessitates the use of a "partial" measure of dependence between the dependent and independent variable set. The partial dependence between an independent and a dependent variable depends on the pre-existing predictors of the system. The term "partial" implies the partial or additional dependence the new predictor can add to the existing prediction model.

The MI criterion presented earlier cannot be used to quantify the partial dependence between two variables. Hence, a partial mutual information criterion was developed. The partial mutual information between the dependent variable *y* and the independent variable *x*, for a set of pre-existing predictors **z**, can be defined as:

$$PMI = \iint f_{X',Y'}(x',y')\log_2\left[\frac{f_{X',Y'}(x',y')}{f_{X'}(x')f_{Y'}(y')}\right]dx'dy' \quad (5)$$

where

$$\begin{aligned} x' &= x - E[x \mid \mathbf{z}] \\ y' &= y - E[y \mid \mathbf{z}] \end{aligned} \quad (6)$$

where E[.] denotes the expectation operation. Use of the conditional expectations in (6) ensures that the resulting variables *x*' and *y*' represent the residual information in variables x and y, once the effect of the existing predictor(s) $\boldsymbol{z}$ has been taken into consideration.

A sample estimate of the PMI criterion in (5) can be formulated as:

$$PMI = \frac{1}{n}\sum_{i=1}^{n}\log_2\left[\frac{f_{X',Y'}(x_i', y_i')}{f_{X'}(x_i')\, f_{Y'}(y_i')}\right] \tag{7}$$

where

x'_i and y'_i are the i'th residuals in the sample data set of size n, and,

$f_X'(x'_i)$, $f_Y'(y'_i)$, and $f_{X',Y}'(x'_i,y'_i)$ are the respective marginal and joint PDF.

Accurate estimation of the PMI score in (7) involves estimation of the conditional expectation needed to estimate the variables x' and y'. This conditional expectation is estimated based on the joint probability density of the variables involved, using the multivariate kernel density estimator of (3). The conditional probability density $f_{X|\mathbf{Z}}(x|\boldsymbol{z})$ can be estimated as the ratio $f_{(X,\mathbf{Z})}(x, \boldsymbol{z})/f_{\mathbf{Z}}(\boldsymbol{z})$, where the denominator refers to the marginal probability density of the pre-existing variable set $\boldsymbol{z}$. The mean of this conditional density is the conditional expectation required in estimating the variables x' and y'. Using the multivariate probability density in (3), the conditional expectation $E[x|\boldsymbol{z}]$ can be estimated as:

$$E[x \mid \boldsymbol{z}] = \frac{1}{n}\sum_{i=1}^{n} w_i \left(x_i + (\boldsymbol{z}-\boldsymbol{z}_i)^T \boldsymbol{S}_{xz} \boldsymbol{S}_{zz}{}^{-1}\right) \tag{8}$$

where

$\boldsymbol{S}_{xz}$ is the sample cross-covariance between x and $\boldsymbol{z}$,

$\boldsymbol{S}_{zz}$ is the sample covariance of $\boldsymbol{z}$

$$w_i = \exp\left(-\frac{(\boldsymbol{z}-\boldsymbol{z}_i)^T \boldsymbol{S}_{zz}{}^{-1}(\boldsymbol{z}-\boldsymbol{z}_i)}{2\lambda^2}\right) \Bigg/ \sum_{j=1}^{n}\exp\left(-\frac{(\boldsymbol{z}-\boldsymbol{z}_j)^T \boldsymbol{S}_{zz}{}^{-1}(\boldsymbol{z}-\boldsymbol{z}_j)}{2\lambda^2}\right) \tag{9}$$

The conditional expectation in (8) is sensitive to the choice of the smoothing parameter λ. A high value of λ results in an oversmoothed regression function, whereas a low value results in the function being “rough". An ideal value of λ is often derived using appropriately constructed error measures, which are usually related to the residuals resulting from the fitted regression line. Data based measures of the bandwidth include cross validatory measures such as generalised cross validation (GCV), details of which can be obtained in standard statistical text books on kernel density and regression function estimation. The present study has used the Gaussian reference bandwidth of (4) as the bandwidth λ in equation (9).

A stepwise predictor selection algorithm can now be formulated for identifying the predictors of the system being modelled using the PMI criterion described above:

Identify the set of variables that could be useful predictors of the system being modelled. Denote this variable set as the vector $\boldsymbol{z_{in}}$. Denote the vector that will store the final predictors of the system as $\boldsymbol{z}$. This is a null vector at the start of the algorithm.

Estimate the PMI between the dependent variable y and each of the plausible new predictors in $\boldsymbol{z_{in}}$, conditional to the pre-existing predictor set $\boldsymbol{z}$.

Identify the variable in $\boldsymbol{z_{in}}$ having the highest PMI score in step 2.
Estimate a 95th percentile randomised sample PMI score for the variable identified in step 3, the randomisation being performed to ensure independence between the two residual variables, such that the 95th percentile score now represents the 95% confidence PMI value when there exists no dependence between the variables.

If the PMI score for the identified variable is higher than the 95th percentile randomised sample PMI score of step 4, include the variable in the predictor set $\boldsymbol{z}$, and remove from $\boldsymbol{z_{in}}$. If the dependence is not significant, go to step 7.

Repeat steps 2 to 5 as many times as are needed.

This step will be reached only when all the predictors have been identified.

Nonparametric conditional prediction

Seasonal prediction of rainfall can proceed through a kernel estimate of the conditional probability density of the rainfall conditional to the current values of the associated predictors. Let us denote the rainfall as the variable Y and the predictors as the d-dimensional vector **X**. The conditional probability can then be written as:

$$f_{Y|\mathrm{X}}(y\,|\,\mathrm{x}) = \frac{f(y,\mathrm{x})}{f(\mathrm{x})} \tag{10}$$

where the numerator represents the joint probability of (y, **x**) and the denominator the marginal probability of **x** (a constant for a given **x**). Using the kernel estimator in (3), the conditional probability in (10) reduces down to the following expression:

$$\hat{f}_{Y|\mathrm{X}}(y\,|\,\mathrm{x}) = \sum_{i=1}^{n} \frac{1}{(2\pi\lambda^2 S')^{1/2}} w_i \exp\left(-\frac{(y-b_i)^2}{2\lambda^2 S'}\right) \tag{11}$$

where

$\hat{f}_{Y|\mathrm{X}}(y\,|\,\mathrm{x})$ is the estimated conditional probability density,

S' is a measure of spread of the conditional density, estimated as:

$$S' = S_{yy} - \mathrm{S}_{xy}{}^{T}\mathrm{S}_{xx}{}^{-1}\mathrm{S}_{xy}$$

where the covariance of (y, **x**) is written as:

$$Cov(y,\mathrm{x}) = \begin{bmatrix} S_{yy} & \mathrm{S}_{xy}{}^{T} \\ \mathrm{S}_{xy} & \mathrm{S}_{xx} \end{bmatrix}$$

w_i is the weight associated with each kernel, representing the contribution that kernel has in forming the conditional probability density:

$$w_i = \frac{\exp\left(-\frac{1}{2\lambda^2}\left[\mathrm{x}-\mathrm{x}_i\right]^T\left[\mathrm{S}_{xx}\right]^{-1}\left[\mathrm{x}-\mathrm{x}_i\right]\right)}{\sum_{j=1}^{n}\exp\left(-\frac{1}{2\lambda^2}\left[\mathrm{x}-\mathrm{x}_j\right]^T\left[\mathrm{S}_{xx}\right]^{-1}\left[\mathrm{x}-\mathrm{x}_j\right]\right)}$$

b_i is the conditional mean associated with each kernel:

$$b_i = y_i - \left[\mathrm{S}_{xy}\right]^T\left[\mathrm{S}_{xx}\right]^{-1}\left[\mathrm{x}-\mathrm{x}_i\right]$$

and n and λ are the sample size and the bandwidth respectively.

Note that the above procedure is very similar to the logic used in developing the partial mutual information criterion described in section 2 of this paper. The main difference in the two approaches is that the PMI criterion requires the estimation of the conditional expectation or conditional mean of a variable given a set of predictors, whereas here the outcome is the conditional probability density function whose mean is equal to the conditional expectation described in section 2.

Estimation of the conditional probability density provides the basis for the hydrologic prediction model. Rainfall can be predicted either by resampling from the conditional probability density estimate in (10), or simply chosen as an appropriate design quantile of the conditional probability density or cumulative distribution function. Resampling does not involve estimation of the PDF along a grid, but simply the selection of one of the kernel's that constitute the conditional probability density, and subsequent resampling from the kernel which is itself a legitimate Gaussian probability density function. The resampled rainfall values can then be used as inputs for Monte Carlo simulation studies for evaluating the performance of the reservoir under current or proposed operating rules.

Application to Warragamba seasonal rainfall

The predictor identification and seasonal prediction methods described in the previous sections were applied to 127 years (1871-1997) of quarterly rainfall data from the Warragamba dam catchment near Sydney, Australia, the four quarters being Autumn (March, April, May), Winter (June, July, August), Spring (September, October, November) and Summer (December, January, February). Two sets of plausible predictors were considered. In the first case, three commonly used indices formulated to measure the strength of an ongoing ENSO anomaly, were used. These were, the Southern Oscillation Index (SOI), the NINO3 index and the NINO3.4 index. While the first index is a standardised measure of sea level pressure anomaly differences between Tahiti (in the mid-Pacific) and Darwin (in northern Australia), the other two respectively measure the average sea surface temperature anomalies within 150°W - 90°W (an equatorial band extending from the mid-Pacific to the western coast of South America), and 170°W - 120°W (an equatorial band more towards the central Pacific than the eastern Pacific region used in deriving the NINO3 series) (see [*Allen et al.*, 1996] for details). The second set of predictors were chosen from a grid of 5° x 5° latitude-longitude averaged sea surface temperature anomalies, reconstructed from observations as described by Kaplan et al. [1998; 1997]. The efficiency of the selected predictors in defining rainfall characteristics was ascertained by fitting a Generalised Additive Model (GAM) [*Hastie and Tibshirani*, 1990] to the data.

The GAM serves as a nonparametric regression model to forecast the expected value of rainfall. This serves as a basis for estimating the R-square of the regression fit, which serves as a means of evaluating the performance of the GAM and the identified predictors in an ensemble forecast context. The code “gam.fit” in the statistical software package S-Plus [*Mathsoft*, 1999] was used with default options in attaining the results shown here.

Table 1. Rainfall prediction using SOI, NINO3 and NINO3.4.

Season	Lead time (quarters)	Predictor Set	PMI	95^{th} percentile PMI	R-Square
Autumn	Lead 1	SOI_{t-1}	0.148	0.120	0.234
		SOI_{t-7}	0.142	0.112	
	Lead 2	$NINO3_{t-2}$	0.148	0.117	0.203
	Lead 3-4	SOI_{t-7}	0.127	0.124	0.092
Winter	Lead 1	SOI_{t-1}	0.202	0.159	0.190
		SOI_{t-9}	0.154	0.150	
Spring	Lead 1	$NINO3_{t-1}$	0.175	0.145	0.106
	Lead 2-4	$NINO3_{t-10}$	0.162	0.152	0.080
Summer	Lead 1	SOI_{t-1}	0.194	0.143	0.287
		SOI_{t-4}	0.182	0.156	
	Lead 2-4	SOI_{t-4}	0.157	0.145	0.093

Table 2. Lead 1-4 predictors of Warragamba quarterly rainfall using sea surface temperature anomaly data. The locations indicated refer to the center of the 5° x 5° latitude-longitude grid cell that was identified as a predictor.

Quarter	Lead	1^{st} Predictor	2^{nd} Predictor	3^{rd} Predictor

Autumn	1	27.5ºS×142.5ºW Lag 1 R^2=0.229	47.5ºN×172.5ºE Lag 7 R^2=0.355	37.5ºN×172.5ºE Lag 2 R^2=0.397
	2-3	32.5ºN×72.5ºW Lag 10 R^2=0.197	32.5ºN×132.5ºW Lag 7 R^2=0.257	62.5ºN×177.5ºE Lag 7 R^2=0.368
	4-6	32.5ºN×72.5ºW Lag 10 R^2=0.197	32.5ºN×132.5ºW Lag 7 R^2=0.257	62.5ºN×177.5ºE Lag 7 R^2=0.368
Winter	1-4	27.5ºS×152.5ºE Lag 8 R^2=0.119	22.5ºS×102.5ºE Lag 8 R^2=0.257	12.5ºN×92.5ºE Lag 4 R^2=0.413
Spring	1	62.5ºN×27.5ºW Lag 4 R^2=0.264	2.5ºS×57.5ºE Lag 9 R^2=0.349	27.5ºS×157.5ºW Lag 5 R^2=0.436
	2-4	62.5ºN×27.5ºW Lag 4 R^2=0.264	2.5ºS×57.5ºE Lag 9 R^2=0.349	27.5ºS×157.5ºW Lag 5 R^2=0.436
Summer	1	2.5ºS×172.5ºE Lag 1 R^2=0.171	47.5ºN×2.5ºW Lag 8 R^2=0.300	67.5ºN×27.5ºW Lag 4 R^2=0.385
	2-6	22.5ºS×107.5ºE Lag 10 R^2=0.142	32.5ºS×57.5ºE Lag 6 R^2=0.233	22.5ºN×162.5ºE Lag 7 R^2=0.350

Tables 1 and 2 present the results from the predictor identification study. The table gives a list of predictors found to be "useful" according to a significance criterion that represents the 95th percentile PMI score that exists when there is a forced independence between the variables, achieved by randomising one of the two variables considered. The table also presents the model R^2 one would achieve when using the GAM regression model described earlier, using the predictors listed. A "lead 1" prediction of Autumn rainfall represents the use of climatic information older than 1 quarter prior to the Autumn rainfall being predicted. The superior performance achieved when using sea surface temperature anomalies as the predictors of the Warragamba rainfall is apparent from the results. The number of predictors identified as significant using the 95th percentile randomised PMI score is greater than the number of predictors (3) listed in table 2. The first three predictors were used as the predictor set for the rainfall prediction results presented next. This number of predictors (3) was chosen as an appropriate number given the length of the record that was available to formulate the rainfall prediction model.

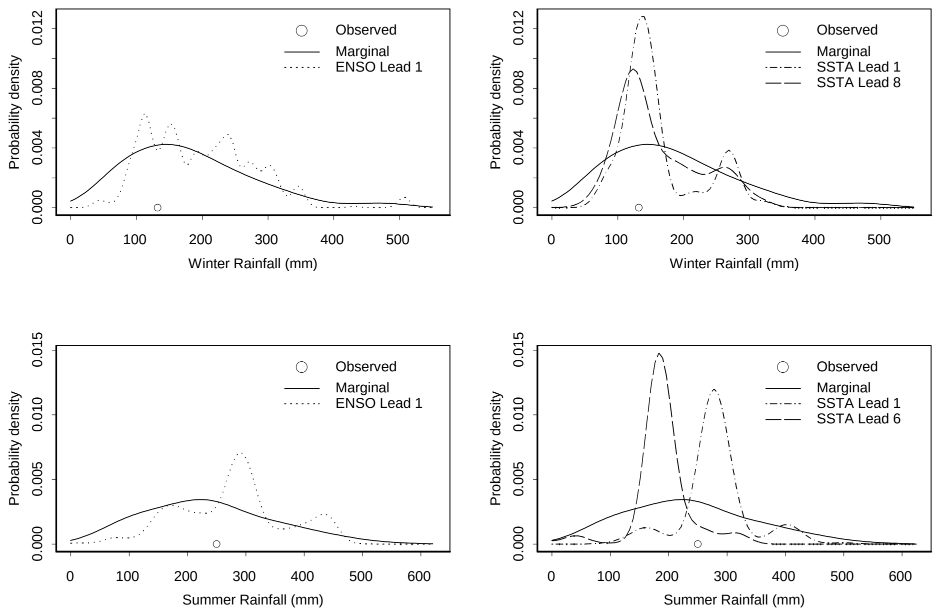

Figure 2. Application of the nonparametric ensemble forecast procedure to the 1996-97 Warragamba seasonal rainfall.

Figure 2 presents some of the predictions from the nonparametric hydrologic prediction model described earlier. The results shown pertain to a few select years of the Warragamba rainfall record. It should be noted that the data used in formulating the prediction model did not include the seasons and years which have been used in the prediction results shown. Of interest in these figures is how better or worse are the predictions in comparison to what would be obtained had no prediction been made. The conditional probability density function represents the range of values the predicted rainfall is likely to take when current climatological conditions are taken into account. The unconditional or marginal PDF represents the range of values the rainfall has assumed based on the available historical record. For the prediction model to be of any good, the variation in the conditional PDF should be lesser than what is present in the historical record. At the same time, the conditional PDF should not have a high bias as regards the observed rainfall value. Both these conditions appear to be satisfied with regards to the predictions based on the SSTA data. The ENSO index based predictions, in most cases, are not significantly better than what is offered by the unconditional or marginal probability density function. This is to be expected as was inferred by the results tabled in tables 1. While the results shown here pertain to only a few selected seasons and years, similar results, sometimes better and sometimes worse, were obtained for the other seasons studied. These results have not been presented here due to space limitations.

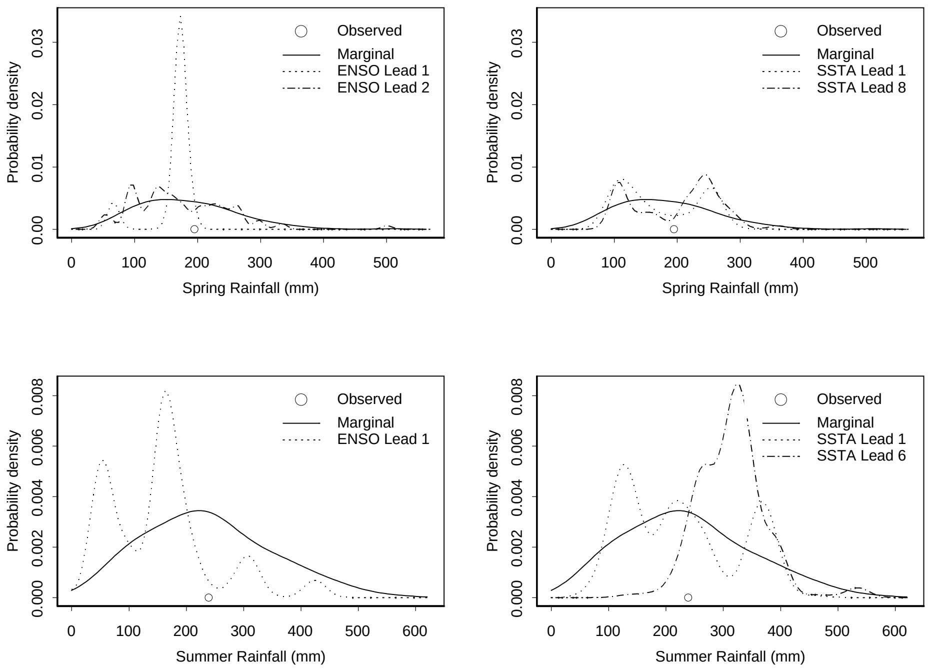

Figure 3. Application of the nonparametric ensemble forecast procedure to the 1997-98 Warragamba seasonal rainfall.

SUMMARY AND CONCLUSIONS

This paper presented a summary of some of the approaches developed and results obtained in course of developing a probabilistic prediction model for seasonal rainfall. The two main approaches that were developed in course of this project were: (1) the Partial Mutual Information (PMI) criterion, a measure of linear or nonlinear partial dependence that may exist between a set of predictors and a dependent variable; and (2) a nonparametric hydrologic prediction model that enables resampling of the predicted rainfall for Monte Carlo simulation studies, or simply produces selected quantiles the rainfall is likely to assume in the coming season. These approaches are of great use in a water resources management context and are being put to use by various water agencies in Australia and elsewhere. More details on this study can be obtained from [*Sharma*, 2000a; *Sharma*, 2000b; *Sharma et al.*, 2000].

Acknowledgments

The author is thankful to the Sydney Water Corporation for funding this research, and to Kin Choi Luk, Ian Cordery, Upmanu Lall and Amir Deen for their comments and suggestions at several stages of the project.

References

Allen, R., J. Lindesay, and D. Parker, *El Nino Southern Oscillation and Climatic Variability*, 405 pp., CSIRO Publishing, Collingwood, VIC, Australia, 1996.

Fraser, A.M., and H.L. Swinney, Independent coordinates for strange attractors from mutual information, *Phys. Rev. A*, *33* (2), 1134-1140, 1986.

Hastie, T., and R. Tibshirani, *Generalized Additive Models*, Chapman and Hall, London, 1990.

Kaplan, A., M.A. Cane, Y. Kushnir, A.C. Clement, M.B. Blumenthal, and B. Rajagopalan, Analyses of Global Sea Surface Temperature 1856-1991, *Journal of Geophysical Research Oceans*, *103* (C9), 18567-18589, 1998.

Kaplan, A., Y. Kushnir, M.A. Cane, and M.B. Blumenthal, Reduced Space Optimal Analysis For Historical Data Sets - 136 Years of Atlantic Sea Surface Temperatures, *Journal of Geophysical Research Oceans*, *102* (C13), 27835-27860, 1997.

Mathsoft, *S-PLUS 2000 User's Guide*, Data Analysis Products Division, Mathsoft, Seattle, WA, USA, 1999.

Scott, D.W., *Multivariate Density Estimation: Theory, Practice and Visualisation*, 317 pp., John Wiley & Sons Inc., New York, 1992.

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

Sharma, A., Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: 3 - A nonparametric probabilistic forecast model, *Journal of Hydrology (In Press)*, 2000b.

Sharma, A., K.C. Luk, I. Cordery, and U. Lall, Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: 2 - Predictor Identification of quarterly rainfall using ocean-atmosphere information, *Journal of Hydrology (In Press)*, 2000.

Silverman, B.W., *Density estimation for statistics and data analysis*, 175 pp., Chapman and Hall, New York, 1986.