Stochastic Hydrology

a) Data Mining for Evolution of Association Rules for Droughts and Floods in India using Climate Inputs

An accurate prediction of extreme rainfall events can significantly aid in policy making and also in designing an effective risk management system. Frequent occurrences of droughts and floods in the past have severely affected the Indian economy, which depends primarily on agriculture. Data mining is a powerful new technology which helps in extracting hidden predictive information (future trends and behaviours) from large databases and thus allowing decision makers to make proactive, knowledge-driven decisions. In this study, a data mining algorithm making use of the concepts of minimal occurrences with constraints and time lags is used to discover association rules between extreme rainfall events and climatic indices. The algorithm considers only the extreme events as the target episodes (consequents) by separating these from the normal episodes, which are quite frequent and finds the time-lagged relationships with the climatic indices, which are treated as the antecedents. Association rules are generated for all the five homogenous regions of India and also for All-India by making use of the data from 1960-1982. The analysis of the rules show that strong relationships exist between the climatic indices chosen, i.e., Darwin Sea Level Pressure (DSLP), North Atlantic Oscillation (NAO), Nino 3.4 and SST values and the extreme rainfall events. Validation of the rules using data for the period 1983-2005, clearly shows that most of the rules are repeating and for some rules, even if, they are not exactly the same, the combinations of the indices mentioned in these rules are the same during validation period with slight variations in the class intervals taken by the indices.

b) Nonlinear Ensemble Prediction of Chaotic Daily Rainfall

The significance of treating rainfall as a chaotic system instead of a stochastic system for a better understanding of the underlying dynamics has been taken up by various studies recently. However, an important limitation of all these approaches is the dependence on a single method for identifying the chaotic nature and the parameters involved. Many of these approaches aim at only analyzing the chaotic nature and not its prediction. In the present study, an attempt is made to identify chaos using various techniques and prediction is also done by generating ensembles in order to quantify the uncertainty involved. Daily rainfall data of Malaprabha basin, India for the period 1955 to 2000 is used for the study. Auto-correlation and mutual information methods are used for preliminary investigation to identify chaos and also to determine the delay time for the phase space reconstruction. Optimum embedding dimension is determined using correlation dimension, false nearest neighbour algorithm and also nonlinear prediction methods. The low embedding dimensions of 5 – 7 obtained from these methods indicate the existence of low dimensional chaos in the rainfall series. A range of plausible parameters is used for generating an ensemble of predictions of rainfall for each year separately for the period 1996 to 2000 using the data till the preceding year (Figure 62). For analyzing the sensitiveness to initial conditions, predictions are done from two different months in a year viz., from the beginning of January and June.

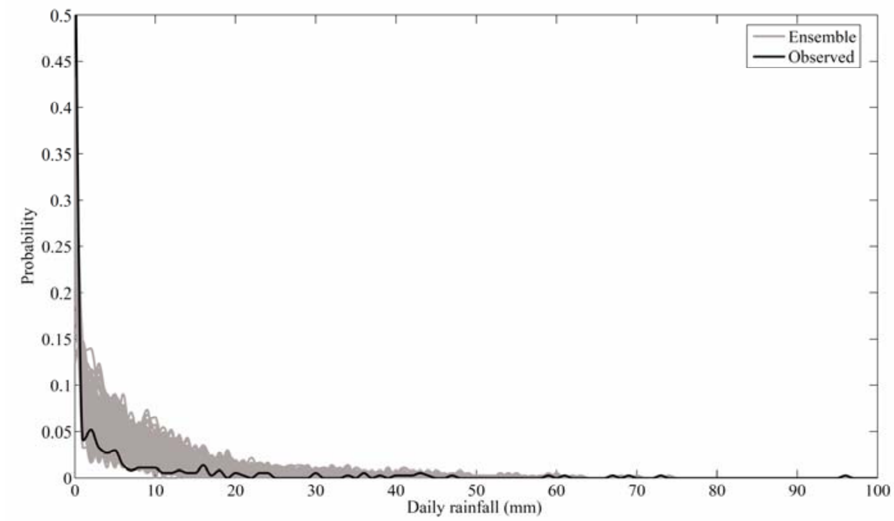

Figure 62: Probability density function of predictions

The reasonably good predictions obtained indicate the efficiency of the nonlinear prediction method for predicting the rainfall series. Also, the rank probability skill score and the rank histograms show that the ensembles generated are reliable with a good spread and skill. A comparison of results of predictions beginning from January and June indicates that June prediction is able to give better results, due to its closeness to the summer monsoon months.

c) Bayesian Networks for Forecasting Nonstationary Hydroclimatic Time Series

Forecasting hydrologic time series is an important tool for adaptive water resources management. Nonstationarity, caused by climate forcing and such other factors, makes the forecasting task too difficult to be modeled by traditional Box-Jenkins approaches. However, potential of Bayesian dynamic modeling approach is investigated through an application to forecast nonstationary hydroclimatic time series. The target is the time series of the volume of Devils' Lake, located in North Dakota, USA, which was proved difficult to forecast by traditional methods. Two different Bayesian dynamic modeling approaches have been discussed, namely, constant model and dynamic regression model (DRM). Constant model uses the information of past observed values of the same time series, whereas DRM utilizes the information of a causal time series as an exogenous input. Noting that the North Atlantic Oscillation (NAO) index appears to co-vary with the time series of the Devil's Lake annual volume, its use as an exogenous predictor is explored in the case study. The results of both the Bayesian dynamic models are compared with those from the traditional Box-Jenkins time series modeling approach. Although, in this particular case study, it is observed that DRM performs marginally better than traditional models, major strength of Bayesian dynamic models lies in quantification of prediction uncertainty, which is of significant importance in hydrology.

d) Non Parametric Approach for Simulating Hydrologic and Meteorologic Time Series

Characterizing the uncertainty in rainfall using stochastic models has been a challenging area of research in the field of operational hydrology for about four decades. Simulated sequences drawn from such models are useful in applications such as agricultural planning, reservoir management, and erosion prediction. Research has been undertaken to develop an efficient block bootstrap model to simulate synthetic rainfall sequences that are statistically indistinguishable and consistent with the historical record of rainfall. Application of the proposed model to rainfall data sets from India, Australia, and USA showed that it is a viable alternative to recently proposed nonparametric *k*-nearest neighbor bootstrap in vogue in hydrologic literature. The developed model overcomes shortcomings of conventional parametric models associated with parsimony and normalizing transformations. It is data driven and can reproduce nonlinearities.

Further, a new Hybrid stochastic approach is developed for multi-site multi-season simulation of streamflows by effectively blending parsimonious linear parametric model and nonparametric bootstrap technique. The proposed approach provides promising alternative to conventional linear and non-linear parametric models as well as nonparametric models in vogue in literature for predicting the uncertainty in streamflows at both spatial and temporal scales in a river basin. Simulated sequences drawn from the proposed model find application in integrated river basin planning, design and operation/management of water resources systems.

e) Regionalization Methods for Predicting Floods

The main challenge in predicting flood magnitudes (quantiles) at target sites is to deal with paucity of at-site data to calibrate parameters of models chosen for the task. An appealing approach to overcome this impediment is to augment information at target site with data or model parameters transferred from sites depicting similar hydrological behavior. To address this problem, new procedures were developed (in collaboration with researchers at Purdue University, USA) based on the concepts of *L*-moments, fuzzy clustering theory, self-organizing feature maps and neural gas networks to: (i) decide plausible groups of similar watersheds in the area of interest, and (ii) systematically adjust the groups to arrive at homogeneous regions for flood frequency analysis. The proposed methods differ from conventional methods in the strategies used to form fuzzy clusters and to arrive at fuzzy regions. The developed procedures reduce subjectivity and effort associated with the task of adjusting the groups of watersheds to arrive at homogeneous regions. Flood quantiles estimated based on the developed methods find use in design and risk assessment of water control structures (e.g., highway culverts, railways and road bridges, urban storm sewers, dams, levees, flood walls and floodway channels), economic evaluation of flood protection projects, land-use planning and management, and flood insurance assessment. Currently, the developed procedures are being used for predicting floods in Indian watersheds. Figure 63 shows homogeneous regions delineated in Godavari basin, India.

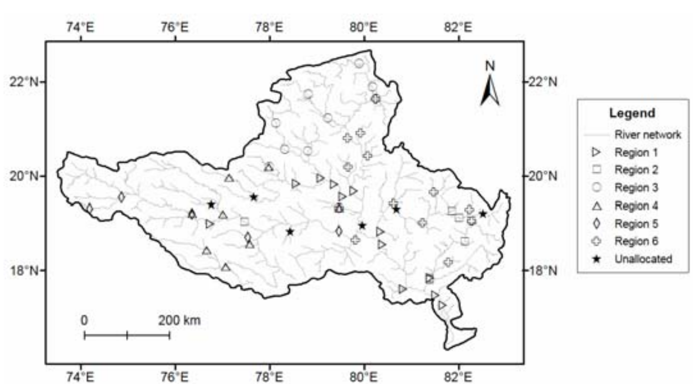

Figure 63: Homogeneous regions formed in Godavari river basin.

f) New Approach for Prediction of Precipitation in Data Sparse Regions

Effective prediction of precipitation in a region is necessary for applications in agricultural planning, and for investigating frequency and spatial distribution of meteorological droughts. Precipitation quantiles at target site may be predicted using at-site frequency analysis procedures. However, the predictions may not be reliable when local precipitation data are sparse. In case of ungauged sites and locations having sparse data, quantiles can be better estimated using regional frequency analysis (RFA). This involves delineation of hydrologically homogeneous region (pooling group) for target site using a regionalization approach, and pooling information at target site with that transferred from other sites within the region for use in quantile estimation. The conventional methods of RFA use statistics of rainfall as attributes to form regions. Owing to this, sufficient number of sites having adequate contemporaneous data is required to form meaningful regions. These methods may not be useful to form meaningful regions in areas having sparse data. Further, identification of pooling group(s) for ungauged site is not possible, because attributes (precipitation statistics that are necessary to identify pooling group) are unknown. Furthermore, validation of the delineated regions is not possible, since the use of same statistics to form regions and subsequently to test their homogeneity is meaningless. To alleviate these problems, a novel approach is proposed. It allows prediction of precipitation in a data sparse region using regionalization based on large scale atmospheric variables (LSAV) affecting hydrometeorology in the region. The LSAV, location parameters (latitude, longitude and altitude) and seasonality of rainfall were suggested as features for regionalization using hard/fuzzy cluster analysis. The proposed approach allows independent validation of the identified regions for homogeneity by using statistics computed from observed precipitation data, and it has the ability to form regions even in areas where the rain gauge density is sparse. Effectiveness of the proposed approach is illustrated through application to India for RFA of annual and summer monsoon rainfall (Figure 64).

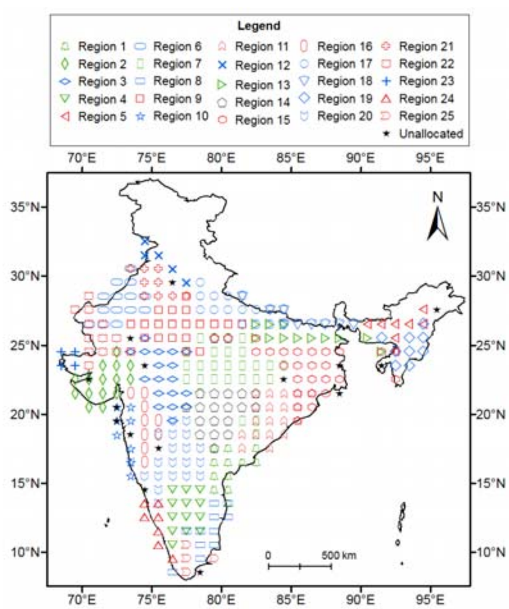

Figure 64: Homogeneous summer monsoon rainfall regions formed based on proposed approach.

Comparisons of the newly formed summer monsoon rainfall (SMR) regions were made with those delineated in previous studies, to bring out the potential of the proposed approach. Results indicated that the SMR regions currently in use by India Meteorological Department (IMD) are statistically heterogeneous. The homogeneous annual rainfall regions newly formed from the proposed analysis are being used to construct curves depicting spatiotemporal characteristics of meteorological droughts in entire India. They find use in planning and management of water resources systems, and in developing drought preparedness plans.