International Conference "Water, Environment, Energy and Society" (WEES–2009)
New Delhi, 12–16 January 2009

An Assessment of Spatio-Temporal Representations in the Stochastic Generation of Daily Precipitation Sequences

R. Mehrotra and Ashish Sharma
The University of New South Wales
AUSTRALIA

ABSTRACT: Many hydrological and agricultural studies require simulation of weather variables that reflect the spatial and temporal dependence observed in point rainfall at multiple locations. This paper assesses three approaches for stochastic generation of multi-point daily rainfall that use different rationales for representing the spatio-temporal dependence that is observed. This assessment is based on an application of the three approaches to point rainfall occurrences at a network of 30 raingauge stations around Sydney, Australia, the rainfall amounts subsequently being generated on the wet days using a nonparametric amount model independent of the occurrence process. The approaches considered consist of a multisite modified Markov model (proposed by Mehrotra and Sharma, 2007b), a method for reconstructing space-time variability (proposed by Clark *et al.*, 2004), and the nonparametric *k*-nearest neighbour (KNN) model (as outlined in Lall and Sharma, 1996). The Modified Markov model simulates precipitation occurrences at individual locations considering low as well as high order Markovian dependence, the spatial dependence being simulated through the use of random innovations that exhibit a spatial dependence structure. In the reconstructing approach, the realisations for a given simulated day are ranked and matched with the rank of the days randomly selected from the similar dates in the historical record. The realisations are then re-ordered to correspond to the original order of the selected historical record thereby reflecting the observed spatio-temporal dependence in the generated series. The *k*-nearest neighbour approach reproduces spatial precipitation distribution structure by simulating precipitation occurrences jointly at multiple locations. Temporal persistence is preserved through Markovian assumptions on the rainfall occurrence process.

The three methods are evaluated for their ability to model various spatial and temporal rainfall attributes over the study area. Our results indicate that all the approaches are successful in reproducing the spatial pattern of the multi-site rainfall field. However, the different orders of assumed Markovian dependence in the observed data limit their ability in representing temporal dependence at time scales longer than a few days. While each approach comes with its own advantages and disadvantages, the alternative proposed by Mehrotra and Sharma (2007b) has an overall advantage in offering a mechanism for modelling varying orders of serial dependence at each point location, while still maintaining the observed spatial dependence with sufficient accuracy. The reordering method of Clark *et al.* (2004) is simple and intuitive, however, is primarily driven by the variability of the observed record, and may not be suited in applications where exogenous covariates can be of help in the simulation process. Implications of using these methods in a water resources management study are discussed.

INTRODUCTION

Stochastic models are commonly used to generate synthetic sequences of rainfall and other weather variables to enhance our understanding of hydrological system response, and in the design and operation of water resource systems. The single site weather generators are easy to formulate and are based on a relatively simple stochastic process. However, the single-site weather generators cannot satisfactorily reproduce the strong spatial correlations among weather variables, often necessary to evaluate the hydrological or agricultural behavior of a region. The spatial correlation of weather variables, specially the precipitation, may have essential effects on the discharge of a river and the formation of floods. A number of multi-site parametric stochastic models have been proposed in literature to address the crucial issue of spatial dependence among series of weather variables at multiple locations, both with and without conditioning on exogenous atmospheric predictors. These models simulate series of weather variables simultaneously at multiple locations, usually at daily time step (e.g., Bras and Rodriguez-Iturbe, 1976; Waymire *et al.,* 1984; Hay *et al.,* 1991; Bardossy and Plate, 1992; Wilson *et al.,* 1992; Hughes *et al.,* 1993; Hughes and Guttorp, 1994; Hughes *et al.,* 1999, Wilks, 1998). However, with the increase in the number of stations, the number of parameters of the model also grows almost exponentially.

It is also common, to incorporate synoptic information for generation of series of weather variables in an attempt to improve the spatial and temporal attributes of the generated series (Hughes and Guttorp, 1994, Hughes *et al.,* 1998, Mehrotra *et al.,* 2004, Sarthik and Bardossy, 2004). This is also found to improve the representation of the low frequency variability in the generated rainfall (Katz and Parlange, 1993). The low frequency variability is important for applications that are sensitive to proper representation of low-frequency persistence, or, the representation of sustained droughts and periods of above average rainfall (or above average wet days) in the generated record. The use of synoptic information while provides the opportunity to incorporate additional information in modeling of the series of weather variables, requires additional time series of atmospheric variables. Additionally, despite, bringing in simplifications for example assuming discrete weather states, with the increase in number of stations, the number of parameters also grows almost exponentially and makes it extremely difficult to estimate and statistically verify the parameters of the model. This has limited their applications for operational purposes.

To overcome the difficulty of parameter estimation of multi-site stochastic models, Wilks (1998) proposed a simple extension of commonly used single-site Markov model for precipitation to multi-sites by driving each of a collection of individual single-site models with temporally independent yet spatially correlated random numbers and successfully simulated daily precipitation while reproducing the observed spatial correlation and preserving the individual behaviours of the local models. Since then, this logic has been successfully applied for simultaneous simulations of rainfall and other weather variables at multiple locations (Wilks, 1999a, b; Qian *et al.*, 2002; Mehrotra *et al.*, 2006; Mehrotra and Sharma, 2007ab).

Recently, in an attempt to reproduce the spatial and temporal structure of observed multi-site series of weather variables without introducing additional complexity, Clark *et al.* (2003) and Clark *et al.* (2004) introduced a method to reconstruct the observed spatial (intersite) and temporal correlation statistics among multiple locations by reordering of the single site generated series of weather variables. Their approach involves re-shuffling an ensemble (realisation) of simulated observations to match the observed spatial and temporal characteristics. In essence, the procedure requires generating multiple realisations (ensembles) individually at each site and reshuffling the simulated series based on the observed information available at these sites. They claimed that the reordering procedure successfully reproduced the number of wet days, rainfall amounts and spell characteristics of the rain-fall series at number of stations, without requiring a complex model structure and tedious parameter estimation.

Another class of weather generators is based on nonparametric alternatives. These methods offer a different rational for generation of climate variables and have been used extensively for this purpose in recent years (Rajagopalan and Lall, 1999; Brandsma and Buishand, 1998; Buishand and Brandsma, 2001; Sharma and O'Neill, 2002; Harrold *et al.*, 2003ab; Mehrotra and Sharma, 2006). These methods offer the alternative of developing the temporal and spatial relationship among weather variables without a priori assumptions on the joint probability distribution associated with the these variables. For multi-site resampling, since the variables at these locations are simulated concurrently, dependence across space is accurately preserved. The *k*-nearest neighbour bootstrap (KNN) is a technique that conditionally resamples the values from the observed record based on the conditional relationship specified. The lack of any assumptions defining the joint distribution of the weather variables helps ensure an accurate representation of features such as nonlinearity, asymmetry or multimodility in the observed record of the variables being modelled.

This paper assesses the three modeling strategies for simultaneous simulation of daily rainfall at multiple locations. All the approaches follow a two step rainfall generation procedure, rainfall occurrences are modeled in the first step, amounts being generated subsequently on the days identified by the occurrence models as wet, using a logic, common to all occurrence models. These approaches therefore essentially differ in terms of rainfall occurrence generation procedure only. The approaches considered are (a) a multi-site Modified Markov model (Mehrotra and Sharma, 2007b); (b) an approach based on reordering of ensemble output in order to recover the space-time variability in the simulated series (Clark *et al.* (2003) and Clark *et al.* (2004)); and (c) a non-parametric *k*-nearest neighbour multi-site model (Buishand and Brandsma, 2001; Mehrotra *et al.*, 2006). Rainfall amounts for all these models are generated using Kernel Density Estimation (KDE) approach (Sharma and O'Neill, 2002; Harrold *et al.*, 2003b; Mehrotra and Sharma, 2006). These models (also including the rainfall amounts) are hereafter referred to as (a) modified Markov, (b) reordering and (c) KNN models. Results of these models are discussed and evaluated against each other and also with the results obtained using an order one Markov model (for rainfall occurrences) and KDE model (for rainfall amounts) applied to each site in isolation without any treatment for spatial dependence. This model is hereafter referred to as the independent model. A 43 year long record of daily rainfall at a network of 30 locations near Sydney, Australia, is used to compare the methods.

The paper is organised as follows. The methodological aspects of models used and data are presented in Section 2. Section 3 provides applications of different models and presents their comparison. In Section 4, finally, the results are summarized and conclusions are drawn.

METHODOLOGY AND DATA USED

As mentioned in the previous section, all approaches considered in the study, differ in terms of rainfall occurrence generation process only, the model of rainfall amounts being common to all approaches. Table 1 provides the details on the spatial and temporal dependence structures of the rainfall occurrence and amounts processes used in these approaches.

Modified Markov Model

We denote rainfall occurrence at a location k and time t as $R_t(k)$ and at the p^{th} time step before the current as $R_{t-p}(k)$. The modified Markov model as proposed in Mehrotra and Sharma [2007b] is based on the conditional simulation of $R_t(k)|Z_t(k)$ within the general framework of Markov process where $Z_t(k)$ represents a vector of conditioning variables at a location k and at time t that in addition to previous time steps values of rainfall imparting daily or short term persistence, can also include atmospheric variables, and/or other continuous variables explaining the higher time scale persistence (denoted as X_t (k), these variables being ascertained by aggregating the rainfall over multiple time steps to convey the slow varying temporal persistence that creates sustained low or high rainfall periods). For brevity, site notations are dropped in the subsequent discussions. The parameters (or the transition probabilities) of a model expressing the order one Markovian dependence (first order Markov model) are defined by $P(R_t|R_{t-1})$ with Z_t consisting of R_{t-1} only. Inclusion of additional predictors X_t in the conditioning vector Z_t would modify these transition probabilities as $P(R_t|R_{t-1}, X_t)$. The following parameterization is adopted to estimate $P(R_t|R_{t-1}, X_t)$,

$$P(R_t = j|R_{t-1} = i, X_t) = p_{ij} \frac{P(X_t|R_t = j, R_{t-1} = i)}{P(X_t|R_{t-1} = i)} \quad \ldots (1)$$

The first term of (1) defines the transition probabilities $P(R_t|R_{t-1})$ of a first order Markov model (representing order one dependence) while the second term signifies the effect of inclusion of predictor set X_t in the conditioning vector Z_t. If X_t consists of derived measures (typically linear combinations) of summation of number of wet days in pre-specified aggregation time periods, one could approximate the associated probability with a multivariate normal distribution, leading to the following simplification for $P(R_t|R_{t-1}, X_t)$,

$$P(R_t|R_{t-1}, X_t) = p_{ij} \frac{\frac{1}{\det(V_{ij})^{1/2}} \exp\left\{-\tfrac{1}{2}(X_t - \mu_{ij}) V_{ij}^{-1} (X_t - \mu_{ij})'\right\}}{\frac{1}{\det(V_i)^{1/2}} \exp\left\{-\tfrac{1}{2}(X_t - \mu_i) V_i^{-1} (X_t - \mu_i)'\right\}} \quad \ldots (2)$$

where μ_i represents the mean $E(X_t | R_{t-1} = i)$ and V_i is the corresponding variance-covariance matrix. Similarly, μ_{ij} and V_{ij} represent, respectively, the mean vector and the variance-covariance matrix of X when ($R_{t-1} = i$) and ($R_t = j$). The p_{ij} parameters represent

Table 1: Details of Temporal and Spatial Dependences Considered in the Models Used

Model	*Process*	*Dependence Modelled*	
		Temporal Dependence of Order	*Spatial Dependence by*
Independent	Occurrence	One	None
	Amount	One	None
Modified Markov	Occurrence	One and monthly and annual level	Spatially correlated random numbers
	Amount	One	Spatially correlated random numbers
Reordering	Occurrence	Using reshuffling	Reshuffling
	Amount	Using reshuffling	Reshuffling
KNN	Occurrence	Spatially averaged order one	Simultaneously picking up observations at all stations
	Amount	One	Spatially correlated random numbers

the baseline transition probabilities of the first order Markov model defined by $P(R_t|R_{t-1})$ and det() represents the determinant operation.

It may be noted that for some applications, the assumption of a multivariate normal may not be sufficient. In such situations, estimating conditional probabilities $P(X_t|R_t, R_{t-1})$ and $P(X_t|R_{t-1})$ in equation (1) either using more appropriate probability distributions or based on nonparametric alternatives (such as kernel density estimation) might be more appropriate. However, for the current application, assumption of normal distribution was found to provide good results.

The Reordering Method

The ensemble (realisation) reordering method is proposed by Clark *et al.* (2004). This method is fairly intuitive and simple, and involves a reordering of ensemble outputs for maintaining the space-time variability in the generated series. In brief, the procedure can be described as follows.

For a given day, the generated ensemble members (n) are ranked from lowest to highest. A short moving window of pre specified days (say fifteen days) is formed centred on the given day. A subset of n observed dates (number of observations being same as number of generated realisations) is randomly selected from the days of the historical record falling within this moving window (dates can be drawn from all years in the historical record except the year for which reordering is being performed). This observed subset is also ranked from lowest to highest separately for each variable and station. For each variable and station, each observation of the ranked observed subset is tagged with the corresponding ranked generated ensemble members, and is re-ordered to its original position with the tagged generated record. Further details on the reordering method are available in Clark *et al.* (2004).

This approach is intended to preserve both spatial and temporal correlations in the reshuffled generated series. Considering the spatial correlation at two stations, if these are highly correlated, then the observations at these stations on a given day are likely to have a similar rank. The rank of each simulated realisation at the two stations is matched with the rank of each randomly selected observation, meaning that, for all realisations, the rank will be similar at the two stations. When this process is repeated for all days, the ranks of a given realisation will on average be similar for the two stations, and the spatial correlation will be reconstructed once the randomly selected days are re-sorted to their original order.

In order to maintain the temporal persistence at each station, the random selections of dates that are used to construct the observed subset are only used for the first day. At a given station, historical observations following high temporal persistence, for subsequent days, on average, would have a similar rank. The ensemble output is assigned identical ranks to the randomly selected observations, and thus the temporal persistence is reconstructed once the ensemble output is re-sorted.

k-Nearest-neighbour Resampling

In the context of multi-site generation of rainfall occurrences, the *k*-nearest-neighbour approach considers simultaneously sampling with replacement of the rainfall occurrences at multiple locations, from the historical records of rainfall. To preserve lagged correlations, resampling is conditioned on the days in the historical record that have similar characteristics as those of the previously simulated days. The similarity is judged on the basis of previous day(s) values of rainfall occurrences at stations. The spatial rainfall distribution structure is maintained by simulating simultaneously at all the stations. Seasonal variations in occurrence generation processes are accommodated using a moving window approach. Further details on the method are available in Rajagopalan & Lall (1999), Buishand & Brandsma (2001), Beersma and Buishand (2003) and Mehrotra *et al.* (2004).

The Single-Site (independent) Rainfall Occurrence Model

The independent model used for comparison generates rainfall occurrence at individual site ignoring the spatial dependence across the stations. Similarly, the ensemble reordering method requires ensembles of generated rainfall occurrence at each site before introducing spatial and temporal dependence by reshuffling. In the present application, the single-site daily rainfall occurrence model used in both these approaches is a first-order Markov model.

The Single-Site Rainfall Amount Model

A nonzero rainfall amount must be generated for each wet day and location of the generated sequences of the occurrence models described in previous sub-sections. The model for rainfall amounts presented here is nonparametric, and is based on the kernel density procedure described in Sharma (2000); Sharma and

O'Niell (2002); Harrold *et al.* (2003b) and Mehrotra and Sharma (2007b). The model is mentioned hereafter as KDE model. It simulates rainfall amount at individual stations conditional on the previous day's rainfall, thereby imparting Markov order-one dependence in the generated series.

Modeling Observed Spatial Dependence in the Generated Rainfall Field

The Modified Markov model accommodates the spatial correlation structure in the generated multisite rainfall occurrences using spatially correlated random numbers. Similarly, KDE model used to generate rainfall amounts also accommodates spatial dependence (for modified Markov and KNN models only) in the generated rainfall amounts by making use of spatially correlated and serially independent random numbers during generation of rainfall at individual stations separately. The general logic of estimating the correlation matrices of random numbers for rainfall occurrence/amounts is available in Wilks (1998) and Mehrotra *et al.* (2006).

Data and Study Area

The study region is located around Sydney, eastern Australia spanning between 147°E–153°E longitude and 31°S–36°S latitude (Figure 1).

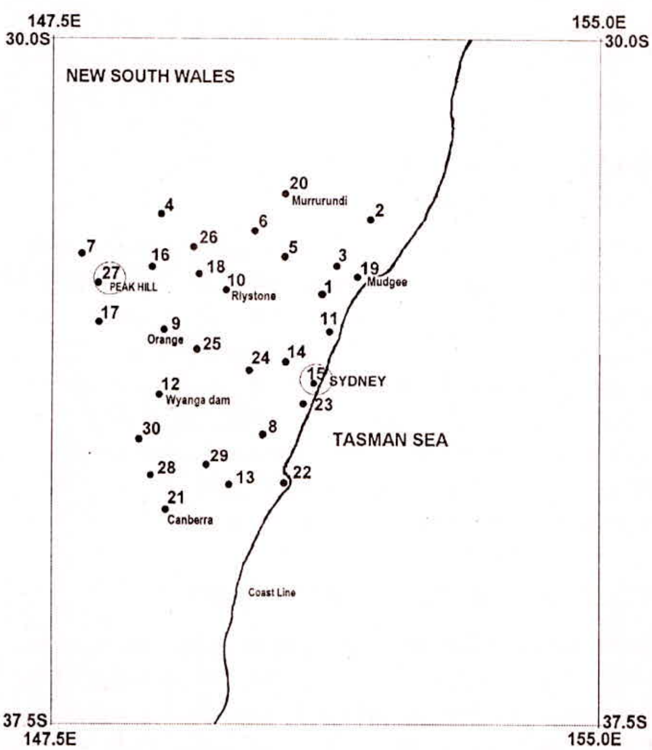

Fig. 1: Map of the study area showing the locations of rain-gauge stations

The most significant rainfall events in winter in this region involve air masses that have been brought over the region from the east-coast low-pressure systems. Orographic uplift of these air masses when they strike coastal ranges or the Great Dividing Range often produces heavy rain. For this study, a 43 year continuous record (from 1960 to 2002) of daily winter rainfall occurrences at 30 stations around Sydney, eastern Australia (see Figure 1) was used. The interstation distances between station pairs vary approximately from 20 to 340 km. Missing values at some stations (<0.5%), were estimated using the records of nearby stations. A day was considered as wet if the rainfall amount was greater than or equal to 0.3 mm.

MODELS APPLICATION AND RESULT DISCUSSIONS

For the KNN model, based on the results of a sensitivity analysis to different choices of width of moving window ℓ, a value of $\ell = 31$ days was chosen for use in our study. Similarly, an analysis was performed to find out the optimal value of k, the number of nearest neighbours, and consequently a value of $k = 10$, was adopted. For modified Markov model, following Mehrotra and Sharma (2007b), previous 30 and 365 days wetness state variables were considered appropriate in explaining the low frequency variability present in the observed record at all stations. The relationship between correlations of series of normally distributed random numbers and corresponding simulated rainfall series at each station pair was ascertained empirically on a daily basis considering the observations falling within a moving window of length 31 days centered on the current day. Similarly, for reordering method, for estimation of daily transition probabilities of the Markov model and, for KDE model, a moving window of length 31 days centred on the current day was found appropriate and considered in the study.

For all occurrence and amounts models except KNN used in the application, previous day value while for KNN model the previous day average rainfall state of the region was assumed to represent the short term dependence of the rainfall process.

The reordering method is applied to the generated series of rainfall amounts only, observed dependence structure of rainfall occurrences is automatically preserved once amounts are reordered.

In all the approaches, results were ascertained by generating 100 ensembles or realisations of the rainfall, based on which statistical performance measures were estimated. The comparison of results was based on the reproduction of various statistics of interest, representing spatial and temporal characteristics of rainfall including those of importance to water resources planning and management.

The graphical comparison of different models was performed on the basis of: (a) spatial dependence statistics, namely, log-odds ratio—a measure of the spatial correlation in the daily rainfall occurrence, and cross correlations—a measure of the spatial correlation in the daily rainfall amounts and aggregated wet days and rainfall amounts in a month and year; (b) average number of wet days and rainfall amounts in a month and year and their variability—a measure of the frequency at which wet days and amounts are simulated and their distribution within and across a year; (c) distributional attributes of wet days and rainfall amount in a year—again a measure of the year to year variability of rainfall occurrence and amounts; (e) spell length rainfall characteristics i.e. wet and dry spells—a measure of day-to-day dependence in the rainfall series and; (f) extreme rainfall characteristics i.e. maximum wet and dry spells and daily maximum rainfall amount—a measure of hydrologic extremes i.e. floods and droughts.

Spatial Dependence

Figure 2 presents observed and modelled log-odds ratios (for daily rainfall occurrences) and cross correlations (for monthly and annual wet days) for all models. For daily and yearly plots, points are shown for each station, while for monthly plots these are shown for each month and station. As the independent model ignores spatial dependence, the generated simulations from the model fail to preserve this characteristic. Use of spatially correlated random number in the modified Markov model helps reproducing the spatial dependence in the generated sequences quite successfully. Similarly, the reordering method by virtue of assigning equal ranked observations at all stations is able to maintain the ranked spatial dependence, albeit some bias for highly correlated stations. As the KNN approach considers precipitation occurrences concurrently at all the stations, the dependence between the stations is automatically preserved by the model. At higher time scales of month and year, the spatial correlations tend to exhibit large scatter for all the models with modified Markov model underestimating the simulated statistic.

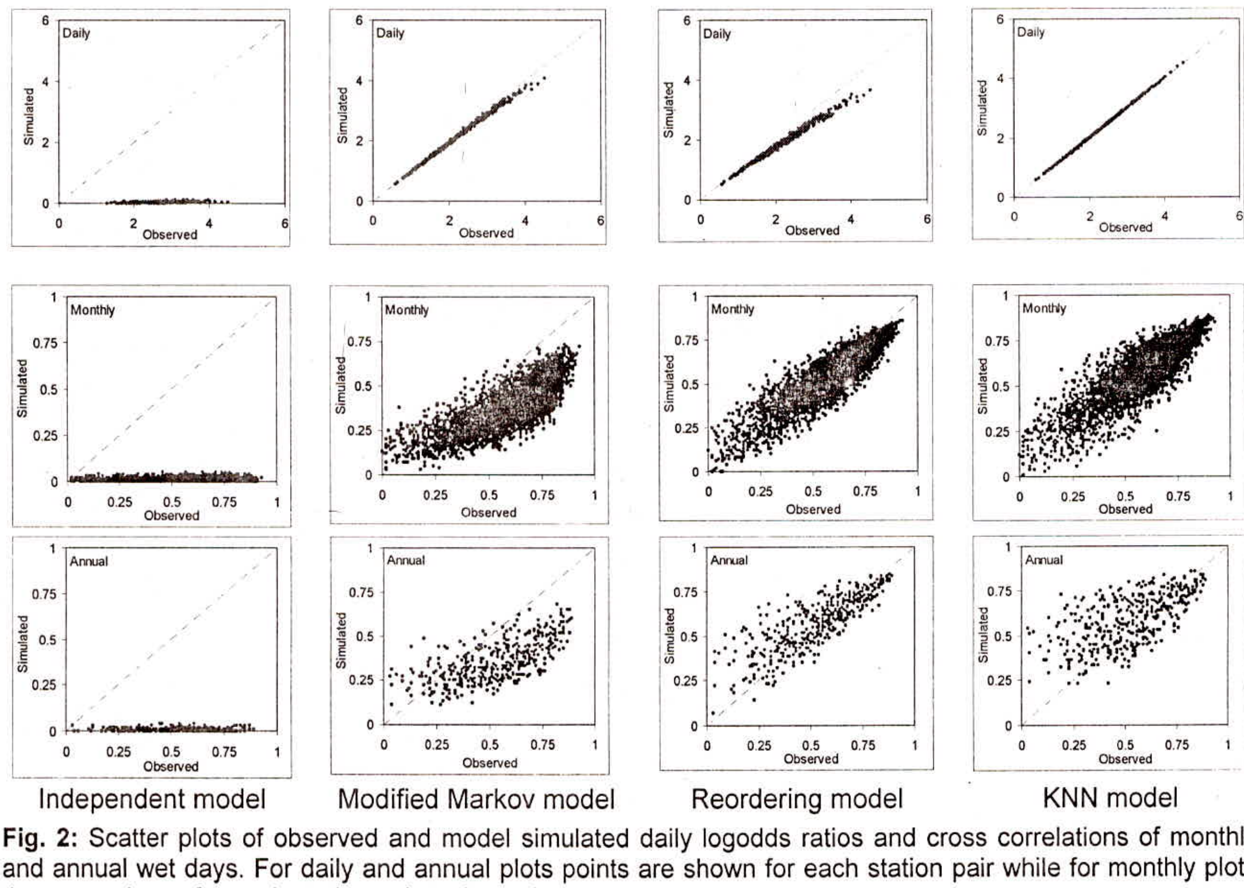

Fig. 2: Scatter plots of observed and model simulated daily logodds ratios and cross correlations of monthly and annual wet days. For daily and annual plots points are shown for each station pair while for monthly plots these are shown for each station pair and month

Similar to Figure 2, Figure 3 presents the cross correlations details of rainfall amounts at daily, monthly and annual time scales. As can be seen, this characteristic is also modelled well by all the three models with the exception of some nominal bias in daily cross correlations by the reordering model.

Having looked at spatial dependence, it would also be of interest to look at the performance of these models in reproducing the various at site temporal rainfall characteristics, of general importance, in water resources planning and management. These are discussed in the subsequent sections.

Aggregated Wet Days, Rainfall Amounts and Associated Variability

It is vital that average wet days and rainfall amount at raingauge network be reproduced accurately before using the simulated rainfall series as an input to any water balance modeling exercise. Figure 4 presents the scatter plots of observed and modelled monthly and annual wet days and rainfall totals at all stations for different models. As expected, all models, including the independent model, provide a good fit to the mean number of wet days and rainfall totals. It is easy to model the average monthly and annual rainfall totals by reproducing properly the probability of occurrence of wet days and average wet day rainfall amount. However, what is more desirable and often difficult, is the successful reproduction of within the year and over the years variability in the rainfall occurrence and amount processes, an effect termed as "overdispersion" (Katz and Parlange, 1998). It may be noted that the monthly and/or annual variance being not only directly related to the rainfall amount variance, probability of a wet day and the average wet day rainfall amount, but also to the monthly/annual variance of wet days, is difficult to represent accurately (Katz and Parlange, 1993, Katz *et al.,* 2003). It has been observed that these low-order Markov process based models, in general, undersimulate the variance at aggregated time scales of month, season and year (Buishand, 1978; Wilks, 1999b).

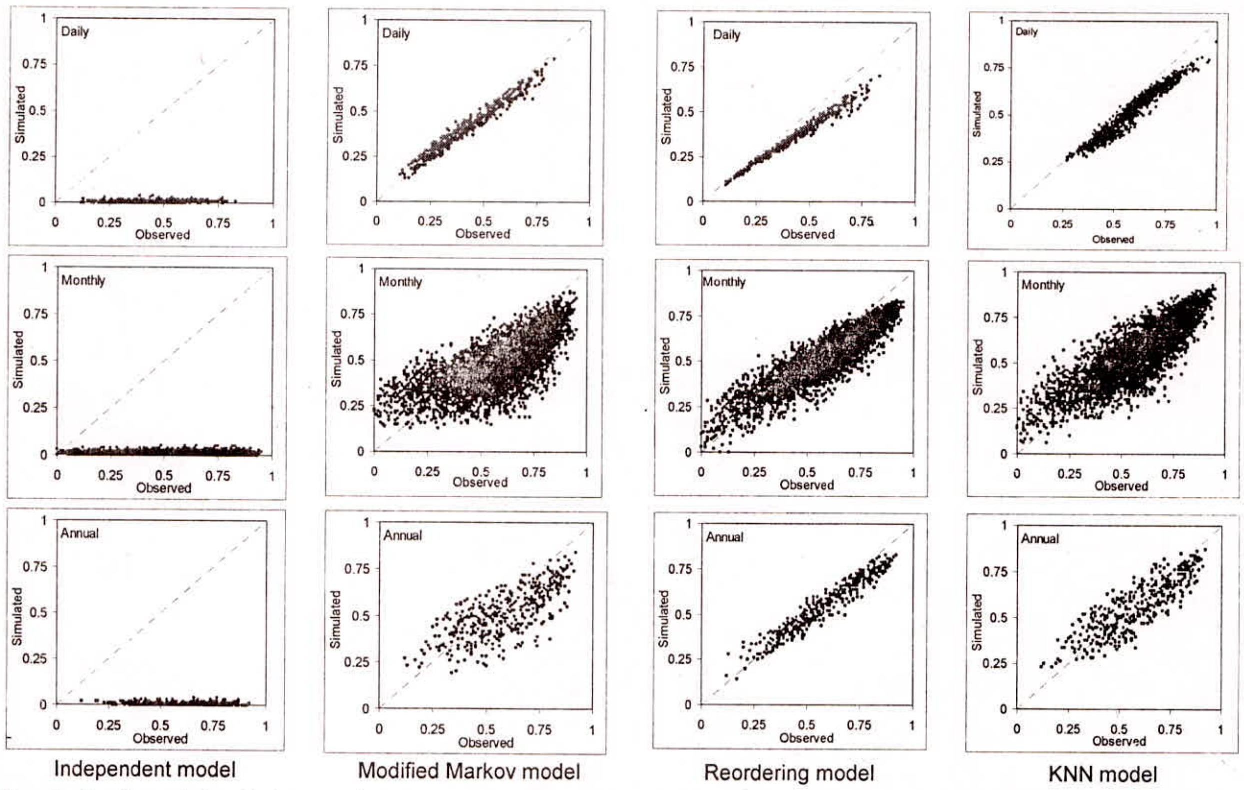

Fig. 3: Scatter plots of observed and models simulated cross correlations of daily, monthly and annual rainfall amounts. For daily and annual plots points are shown for each station pair while for monthly plots these are shown for each station pair and month

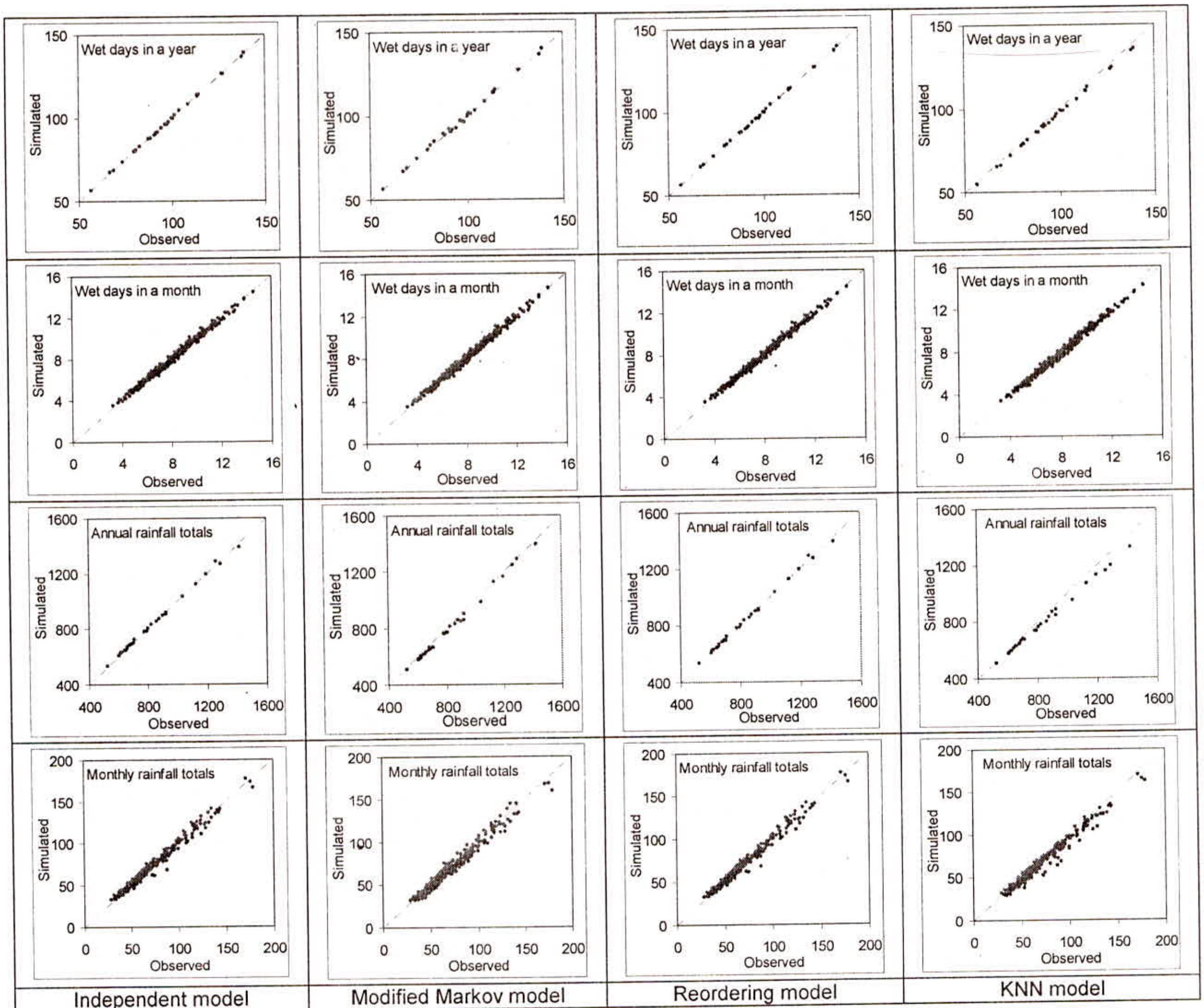

Fig. 4: Scatter plots of observed and models simulated monthly and annual wet days and rainfall totals. For annual plots points on the graph are shown for each station while for monthly plots these are shown for each station and month

However, this overdispersion characteristic assumes importance for applications that are sensitive to proper representation of low-frequency persistence, or, the representation of sustained droughts and periods of above average rainfall (or above average wet days) in the generated record. Figure 5 presents the scatter plots of observed and modelled standard deviations of monthly and annual wet days and rainfall totals. As can be depicted from these plots, independent and KNN models grossly underestimate the variability whereas modified Markov model and reordering models are successful in reproducing the variability of aggregated wet days at monthly and annual time scales. Similarly, all the models except the KNN are successful in reproducing the variance of monthly rainfall totals whereas modified Markov and reordering models are more successful in preserving the variance of rainfall amounts at annual time scale. Obviously, the order one dependence of independent and KNN models is insufficient in explaining the higher time scale or low frequency variability, specifically at annual level, as exhibited by the observed record. The modified Markov model that considers conditioning on previous 30 and 365 days wetness state is able to recognise this variability quite well. Similarly, the reordering model that follows the calendar dates in the selection of observed samples is also able to reproduce this variability quite successfully.

Another important rainfall statistic related to the annual variance is the distribution of over the year rainfall variability. Figures 6 and 7 present the probability plots of the distribution of wet days and total rainfall per year for the two selected stations 27 (Figure 6) and 15 (Figure 7) representing, respectively, dry and wet regions. These plots show that the driest year on record for station 27 has only 35 wet days, but the wettest year has approximately 105 wet days.

Likewise, station 15 has 90 and 170 days for driest and wettest years, respectively. The generated sequences from independent and KNN models do not reproduce these distributions whereas modified Markov and reordering models successfully capture these highs and lows (top two rows of Figures). The standard deviation of the number of wet days per year is directly related to this distribution, and thus generated sequences from independent and KNN models also under-represent the historical annual-level standard deviation. For total rainfall (bottom two rows of Figures) these differences are not so pronounced amongst the models.

Extreme Rainfall Characteristics

It has been observed that assuming a low-order Markov dependence, in general, undersimulates long dry spells (runs of consecutive dry days) (Buishand, 1978; Guttorp, 1995; Racsko *et al.*, 1991; Semenov and Porter, 1995; Wilks, 1999b). Figure 8 presents the scatter plots of observed and models generated highest wet and dry spell lengths and, maximum rainfall amount at all stations. All the models are successful in reproducing the maximum wet spell length (top row) at all stations with the exception of modified Markov model which somewhat overestimates this characteristic at a few stations. Similarly, generated sequences from independent and KNN models underestimate the maximum dry spell lengths at majority of stations whereas modified Markov and reordering models adequately reproduce this characteristic at all stations (middle row). Daily maximum rainfall is adequately reproduced by all the models at majority of stations (bottom row) except for KNN model which shows minor underestimation of this characteristic at a few stations.

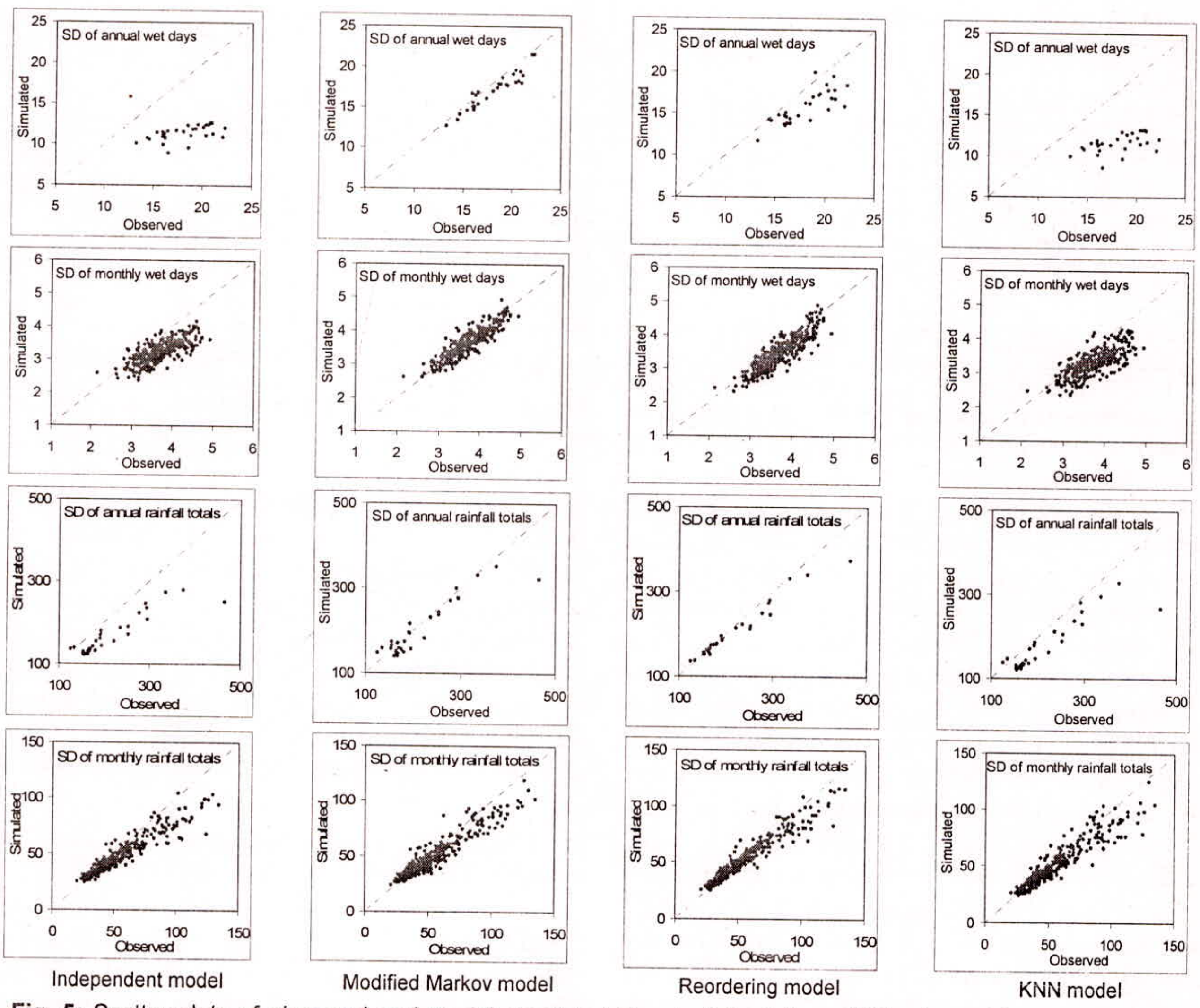

Fig. 5: Scatter plots of observed and model simulated Standard Deviations (SD) of monthly and annual wet days and rainfall totals. For annual plots points on the graph are shown for each station while for monthly plots these are shown for each station and month

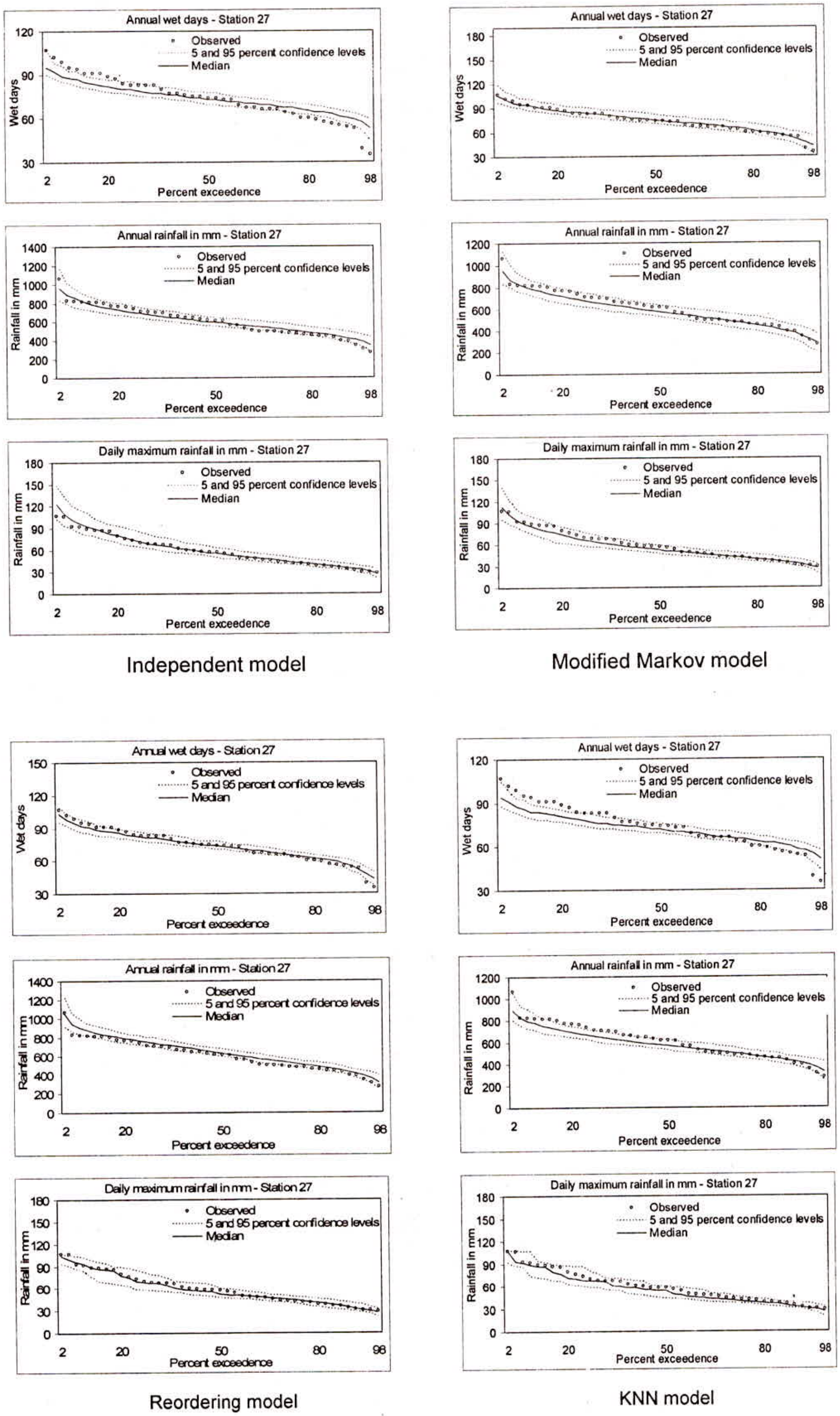

Fig. 6: Distribution plots of observed and model simulated annual wet days and rainfall amounts and, daily maximum rainfall amount for a selective station 27 (dry region). Historical observations are shown as circles while 5 and 95 percentile simulated values are shown as dotted lines with simulated median values as solid lines

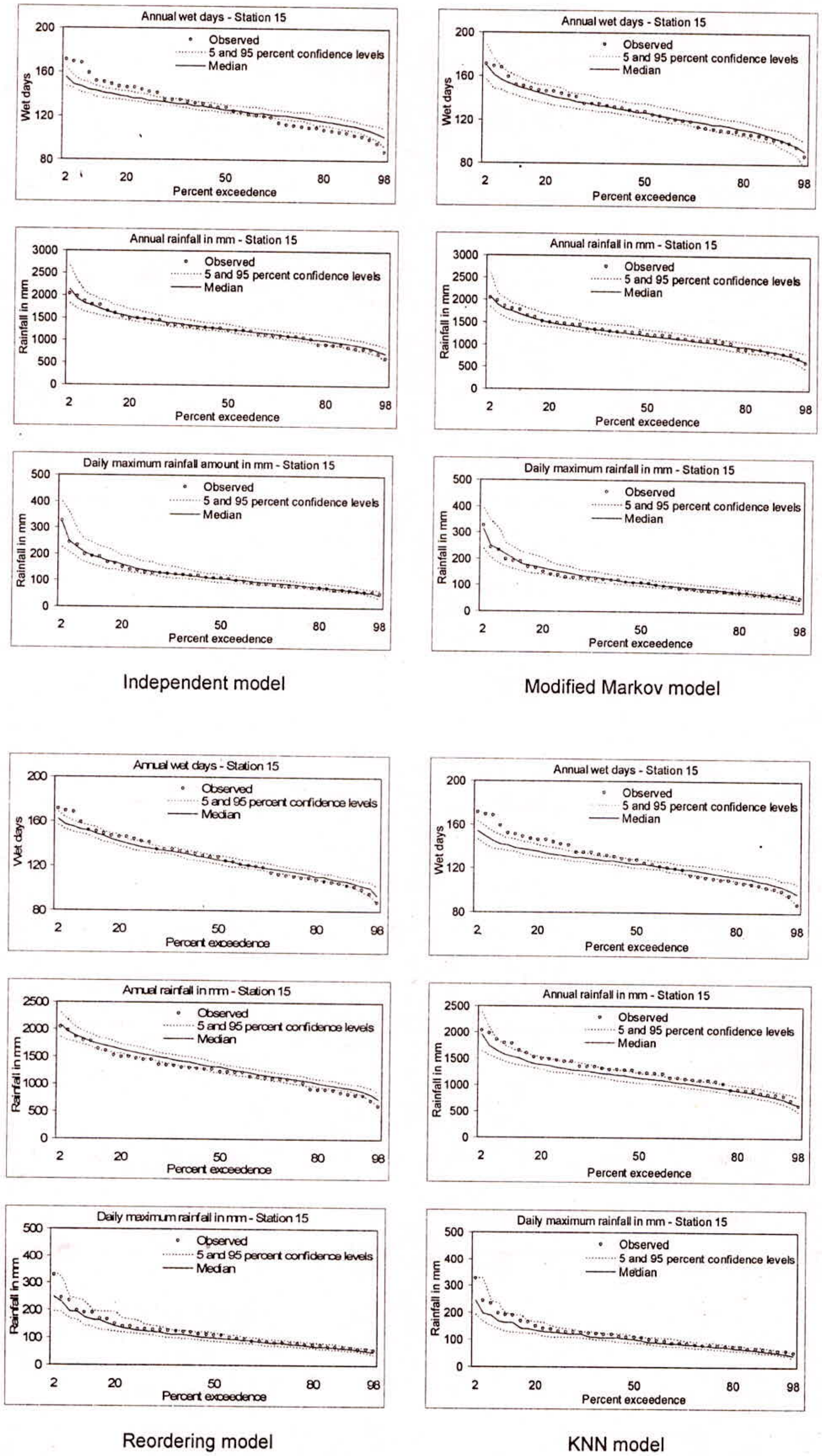

Fig. 7: Distribution plots of observed and model simulated annual wet days and rainfall amounts and, daily maximum rainfall amount for a selective station 15 (wet region). Historical observations are shown as circles while 5 and 95 percentile simulated values are shown as dotted lines with simulated median values as solid lines

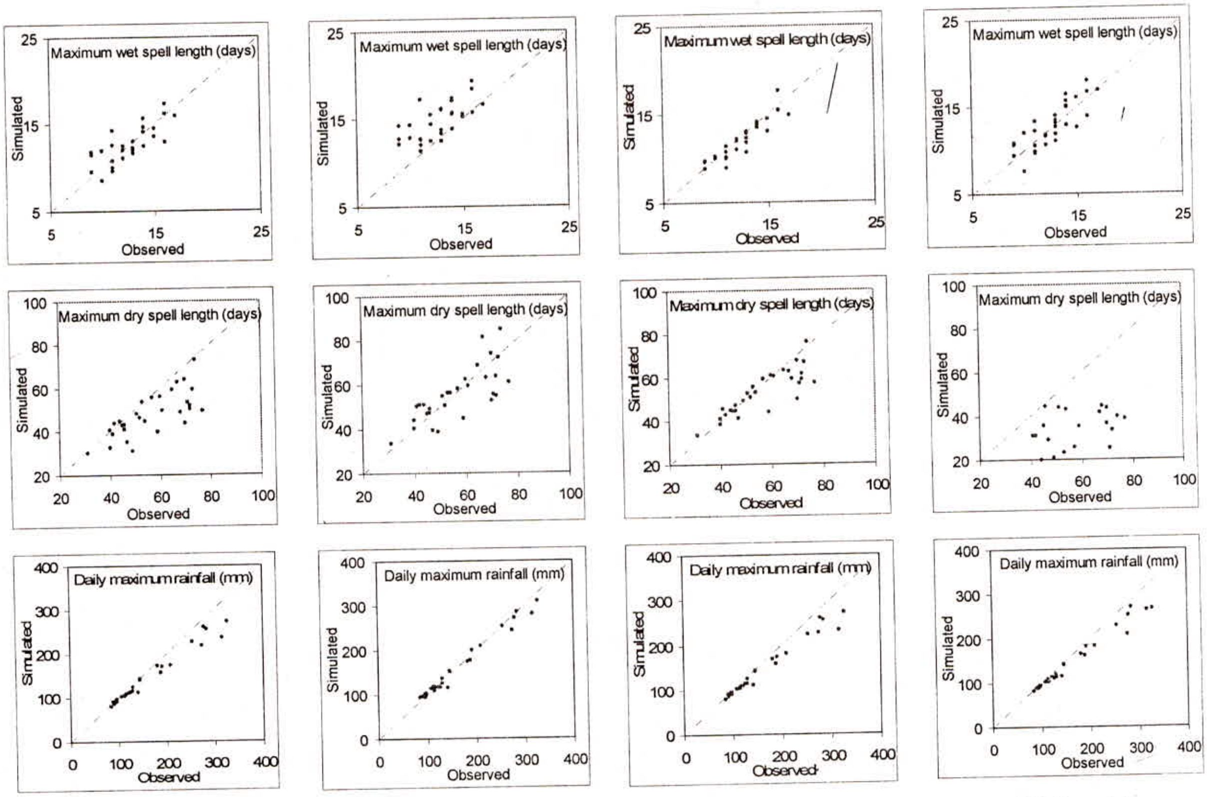

Fig. 8: Scatter plots of observed and model simulated maximum wet and dry spells and daily maximum rainfall amount. Points on the graph are shown for each station

Combined Spatio-temporal Rainfall Characteristics

Another important statistic, indicative of proper reproduction of spatial as well as temporal dependence is the marginal and conditional probabilities of area averaged wetness state (for rainfall occurrence) and rainfall amounts. Figure 9 presents the probability plots of observed and modelled daily area averaged wetness state. Plots of rainfall amounts being similar for all models are not presented here. On these plots observed statistic is shown as circles while simulated values are shown as lines. Modified Markov, reordering and KNN models successfully reproduce these probabilities at daily time scale. As KNN method considers conditioning on area averaged wetness state, it is structured to reproduce this statistic appropriately. Reordering method tends to match the observed ranks in the generated simulations at daily time scale and therefore also matches the area averaged statistic quite well. In spite of the fact that modified Markov model successfulty reproduces the spatial dependence and is structured to reproduce the order one temporal dependence at individual station, it somewhat under simulates the area averaged statistic. It appears that for proper reproduction of temporal dependence of area averaged statistics, modelling of the spatial and at-site lagged temporal correlations might not suffice. Consideration of lagged cross correlations or conditioning on a variable representing area averaged conditions might help improving the results.

SUMMARY AND CONCLUSIONS

This paper has presented an assessment of three multisite models of daily rainfall, which are based on different concepts of reproducing the spatial and temporal dependence in the generated rainfall sequences. While short term temporal dependence in all the models is formulated assuming an order one Markovian dependence, these differ in terms of the modeling of higher time scale temporal and spatial dependences. These schemes do not pretend to model directly the physical processes underlying the spatial and higher order temporal distributions of weather variations, but rather are based on relatively simple stochastic simulation of weather derived from the behavior of the observed rainfall.

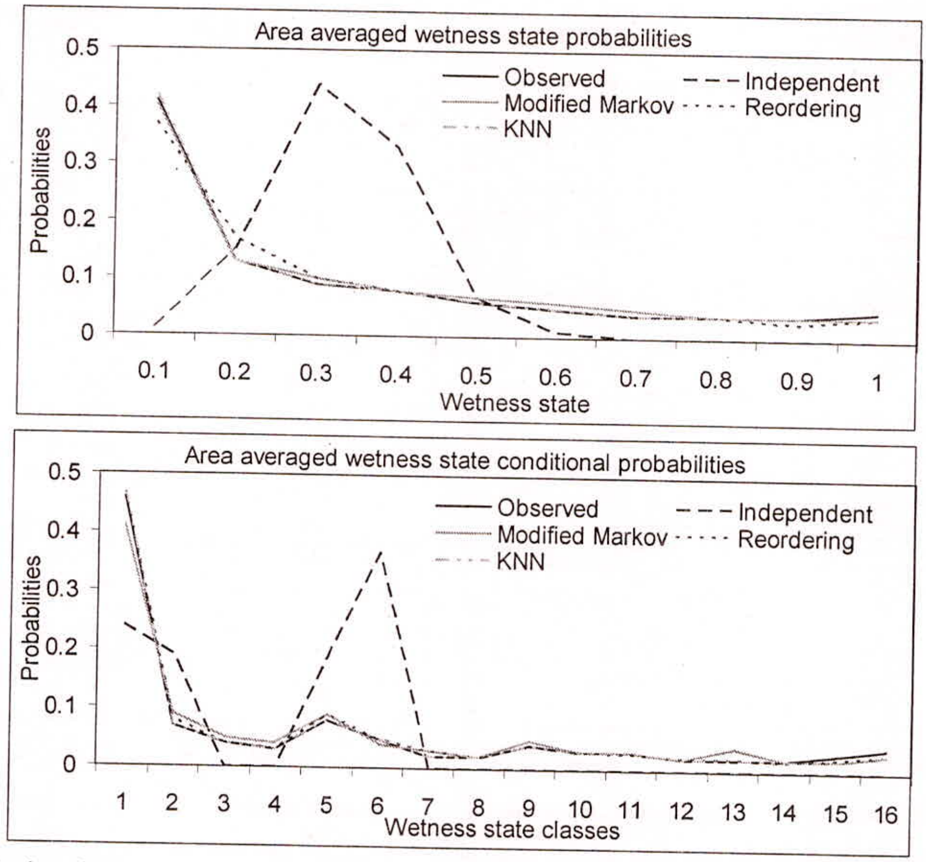

Fig. 9: Marginal and conditional probabilities under discrete area averaged wetness states. For marginal probabilities wetness state classes varies from 0 to 1 with a bin width of 0.1 while for conditional probabilities bin width is 0.25 constituting 4 classes and 16 possible transitions of going from one weather state (of previous day) to another (of current day) on a day

Representing spatial dependence has always been a topic of concern among the researchers when it comes to stochastic simulation of rainfall. The three methods presented here provide a simple yet effective logic of reproducing this characteristic at a network of stations.

We found that the models with low-order Markovian dependence (independent and KNN models) are inadequate in producing the extended dry spells with proper frequency. This is in agreement with the findings of other researchers (e.g., Buishand, 1978; Racsko *et al.,* 1991; Lettenmaier, 1995). The use of higher time scale wetness state (modified Markov model) or following the dependence of the observed record (reordering model) facilitates retaining the higher order dependence also in the generated simulations.

The KNN model is structured to retain the spatial dependence, however, does not support the reproduction of the higher time scale characteristics of the rainfall. Conditioning on higher time scale variables or following the procedure as mentioned in Mehrotra and Sharma (2007a) is expected to further improve the performance of the model.

The reordering model is based on simple logic of reproducing the ranked statistics at individual stations which in turn helps reproducing the desired spatial and temporal statistics in the shuffled generated sequences. The approach equally holds good for other weather variables as well. However, the approach has limitations if there are many similar records in the observed data such as is the case with rainfall which contains many zeros. The poor simulation of daily spatial correlations is the result of having many zeros at few stations. Additionally, were the approach to use a small moving window and a limited historical record size, it will simulate realizations that have an identical rank structure (in space and time) to what is observed. While the spatio-temporal dependence statistics will be represented well (perfectly in the rank space), the stochasticity of the simulated sequences will be compromised. For the success of the method, the observed record should have enough length and variability with not many repeated observations. It should be noted that the good simulation of difficult to represent statistics such as long term persistence is because of the observed rank structure being replicated without significant alteration in formulating the ensembles. Also, the approach is most suited for applications where aim is to have observed statistics being reproduced in the simulated sequences. However, the approach is not suited for downscaling or climate change related studies where observed rainfall spatio-temporal structure is expected to behave in a different manner in the changed climate conditions.

REFERENCES

Bardossy, A. and Plate, E.J. (1992). "Space-time model for daily rainfall using atmospheric circulation patterns." *Water Resour. Res.* 28, 1247–1259.

Brandsma, T. and Buishand, T.A. (1998). "Simulation of extreme precipitation in the Rhine basin by nearest neighbour resampling." *Hydrol. Earth Syst.* Sci. 2, 195–209.

Bras, R., Rodriguez-Iturbe, I. (1976). "Rainfall generation: a nonstationary time varying multidimensional model." *Water Resour. Res.* 12, 450–456.

Buishand, T.A. (1978). "Some remarks on the use of daily rainfall models." *J. Hydrol.* 36, 295–308.

Buishand, T.A. and Brandsma, T. (2001). "Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling." *Water. Resour. Res.* 37, 2761-2776.

Clark, M.P., Gangopadhyay, S., Hay, L.E., Rajagopalan, B. and Wilby, R.L. (2004). The Schaake shuffle: "A method for econstructing space-time variability in forecasted precipitation and temperature fields." *J. Hydrometeorol.* 5, 243– 262.

Clark, M.P., Gangopadhyay, S., Brandon, D., Werner, K., Hay, L., Rajagopalan, B. and Yates, D. (2004). "A resampling procedure for generating conditioned daily weather sequences." *Water Resour. Res.*, 40, W04304, doi:10.1029/2003WR002747.

Guttorp, P. (1995). "Stochastic Modeling of Scientific Data." Chapman and Hall, London (Chapter 2).

Harrold, T.I., Sharma, A. and Sheather, S.J. (2003a). "A nonparametric model for stochastic generation of daily rainfall occurrence." *Water Resour. Res.* 39(12), 1343, doi:10.1029/2003WR002570.

Harrold, T.I., Sharma, A. and Sheather, S.J. (2003b). "A nonparametric model for stochastic generation of daily rainfall amounts." *Water Resour. Res.* 39(12), 1300, doi:10.1029/2003WR002570.

Hay, L.E., McCabe, G.J. Jr., Wolock, D.M. and Ayres, M.A. (1991). "Simulation of precipitation by weather type analysis." *Water Resour. Res.* 27, 493–501.

Hughes, J.P., Lettenmaier, D.P. and Guttorp, P. (1993). "A stochastic approach for assessing the effect of changes in synoptic circulation patterns on gauge precipitation." *Water Resour. Res.* 29, 3303–3315.

Hughes, J.P. and Guttorp, P. (1994). "A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena." *Water Resour. Res.* 30, 1535–1546.

Hughes, J.P., Guttorp, P. and Charles, S.P. (1999). "A non-homogeneous hidden Markov model for precipitation occurrence." *Appl. Stat.* 48(1), 15–30.

Katz, R.W. and Parlange, M.B. (1993). "Effects of an index of atmospheric circulation on stochastic properties of precipitation." *Water Resour. Res.* 29, 2335–2344.

Katz, R.W. and Parlange, M.B. (1998). "Overdispersion phenomenon in stochastic modeling of precipitation." *J. Climate*, 11, 591–601.

Katz, R.W., Parlange, M.B. and Tebaldi, C. (2003). "Stochastic modeling of the effects of large-scale circulation on daily weather in the southeastern U.S." *Clim. Change.* 60, 189-216.

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

Lettenmaier, D. (1995). "Stochastic modeling of precipitation with applications to climate model downscaling." In: von Storch, H., Navarra, A. (Eds.), *Analysis of Climate Variability, Applications of Statistical Techniques*, Springer-Verlag, Berlin, pp. 197–212.

Mehrotra, R., Sharma, A. and Cordery, I. (2004). "Comparison of two approaches for downscaling synoptic atmospheric patterns to multisite precipitation occurrence." *J. Geophys. Res.* 109, D14107, doi:10.1029/2004JD004823.

Mehrotra, R., Srikanthan, R. and Sharma, A. (2006). "A comparison of three stochastic multi-site precipitation occurrence generators." *J. Hydrol.* 331, 280–292.

Mehrotra, R. and Sharma, A. (2006). "A nonparametric stochastic downscaling framework for daily rainfall at multiple locations." *J. Geophys. Res.* 111, D15101, doi:10.1029/2005JD006637.

Mehrotra, R. and Sharma, A. (2007a). "A semi-parametric model for stochastic generation of multi-site daily rainfall exhibiting low-frequency variability." *J. Hydrol.* 335, 180–193.

Mehrotra, R. and Sharma, A. (2007b). "Preserving low-frequency variability in generated daily rainfall sequences." Submitted to *J. Hydrol.* for possible publication.

Qian, B., Corte-Real, J.A. and Xu, H. (2002). "Multisite stochastic weather models for impact studies." *Int. J. Climatol.* 22, 1377–1397.

Racsko, P., Szeidl, L. and Semenov, M. (1991). "A serial approach to local stochastic weather models", *Ecological Modeling*, 57, 27–41.

Rajagopalan, B. and Lall, U. (1999). "A k-nearest neighbour simulator for daily precipitation and other weather variables." *Water Resour. Res.* 35, 3089–3101.

Semenov, M.A. and Porter, J.R. (1995). "Climatic variability and the modelling of crop yields." *Agri. Forest Meteorol.* 73, 265–283.

Sharma, A. and O'Neill, R. (2002). "A nonparametric approach for representing interannual dependence in monthly streamflow sequences." *Water Resour. Res.* 38(7), 5.1–5.10.

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

Stehlik, J. and Bardossy, A. (2002). "Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation." *J. Hydrol.* 256, 120–141.

Waymire, E., Gupta, V.K. and Rodriguez-Iturbe, I. (1984). "A spectral theory of rainfall at the meso-b scale." *Water Resour. Res.* 20, 1453–1465.

Wilks, D.S. (1998). "Multisite generalization of a daily stochastic precipitation model." *Jour. Hydrol.* 210, 178–191

Wilks D.S. (1999a). "Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain." *Agri. Forest Meteorol.* 96, 85–101.

Wilks, D.S. (1999b). "Multisite downscaling of daily precipitation with a stochastic weather generator." *Clim. Res.* 11, 125–136.

Wilson, L.L., Lettenmaier, D.P. and Skyllingstad, E. (1992). "A hierarchical stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation." *J. Geophys. Res.* D3, 2791–2809.