University of Waterloo

Development and Application of a K-NN Weather Generating Model

CFCAS project:
Assessment of Water Resources Risk and Vulnerability to Changing Climatic Conditions

Project Report III

March 2004

Prepared by

Mohammed Sharif

and

Donald H. Burn

CFCAS Project Team:

University of Western Ontario
Slobodan P. Simonovic
Gordon McBean
Juraj M. Cunderlik

University of Waterloo
Donald H. Burn
Linda Mortsch
Mohd Sharif

Upper Thames River Conservation Authority
Rick Goldt
Mark Helsten

Contents

List of figures

List of tables

1. Development and Application of a K-NN Weather Generating Model

1.1 Introduction

The main objective of this study is to develop a generic weather generating model that can be used to hypothesize plausible climate change scenarios for any given basin. Development of weather generator forms a part of a larger Canadian Foundation for Climatic and Atmospheric Sciences (CFCAS) funded project "Assessment of Water Resources Risk and Vulnerability to Changing Climatic Conditions" that aims at improving the prediction of hydrologic extremes, including both flood and drought events in the Upper Thames River basin (UTRb) in the Canadian province of Ontario. This report describes the development and application of a generic weather generator, based on the K-nearest neighbour (K-NN) algorithm (Buishand and Bradsma, 2001; Rajagopalan and Lall, 1999; Yates et al., 2003).

Global climate is expected to change significantly due to the continuously increasing levels of carbon dioxide and other greenhouse gases. By the year 2056 the CO_2 concentration in the atmosphere is likely to double (Jenkins and Derwint, 1990). Globally, the 10 warmest years on record have occurred since 1990s (1989-98) with 1998 the warmest year in atleast a millennium (Mann et al., 1999). Mean annual global increases of air temperature of 0.5 C in the last century are evident, and the precipitation has increased over land in high latitudes in the Northern Hemisphere especially in the cold season (IPCC, 1996). The future projections of climate change indicate a global average warming of between 1.5° to 4.5° C, greater surface warming at high latitudes in winter, but less during the summer. An increase of 3 to 15 % in global precipitation is expected mainly due to globally increasing temperatures which causes greater evaporation from sea surface water. A year-round increase in precipitation in high-latitude regions is expected, whilst some tropical areas may experience small decreases. Climate change in Canada has followed a trend somewhat similar to the global trend. Mean temperature in Canada has risen 1.1 C nationally and 1.7 C regionally in the past century (Gullet and

Skinner, 1992) and temperature increases are expected to be greater in winter than in summer.

Global climate changes are anticipated to have potentially serious impacts on many aspects of the natural environment including the earth's water resources. Spatial and temporal distribution of water would change impacting the availability of water to meet human and other needs. On a regional scale, the climatic impacts would be influenced by a complex combination of temperature, precipitation, evaporation, and changes in runoff production. Relatively small changes in the precipitation and temperature characteristics of a region, which are the principal driving forces behind the hydrological cycle, can significantly alter water resource systems of a region. Due to increased vigour of the hydrological cycle, the magnitude and timings of occurrence of extreme events such as flood and drought are likely to change considerably thereby introducing additional uncertainty in the management of existing water resource systems (IPPC, 2001). A major concern arising out of changed climate scenarios is that of habitat loss as there may be no alternatives to some species. The timing and magnitude of specific hydrologic events such as freeze-up/break-up, the severity of the spring freshet, or the duration of the low flow period is vital to the life cycle of many species. On the human front, the economic effects of extreme events can be devastating. For example, income and employment losses due to business disruptions are a common feature associated with a flood event. Other direct effects include loss to human life and property, loss of livestock, and crop damages.

Reservoir operations, crop production, erosion processes, runoff production and many other hydrological processes are likely to be impacted by climate change. Revelle and Waggoner, (1983) and Gleick (1987) have indicted that climate change can adversely affect the availability of water supply. Patterns of water demand will need to alter in response to changes in water supply. Some other impacts of climate change that have been identified includes changes in the quantity of runoff produced (Gleick, 1986; Lattenmaier and Gan, 1990), and changes in the timings of the hydrologic events (Lattenmaier and Gan, 1990; Kite, 1993; Burn, 1994; Simonovic, 2001). Any change in the hydrological processes will ultimately affect the water resource systems. Burn and Simonovic (1996) investigated the potential impacts of climate change on the performance of reservoir operations. Hydrologic scenarios representing different sets of climatic conditions were generated and used as an input to a reservoir operation model. It was concluded that the reservoir performance is sensitive to climate change. Westmacott and Burn (1997) evaluated the effects of climatic changes on hydrological variables pertaining to the magnitude and timing of hydrological events in the Churchill-Nelson River Basin in west-central Canada. The magnitude of hydrologic events was found to decrease over time while snowmelt runoff events occurred earlier. Mortsch et al. (2000) studied the impact of changing climatic conditions on the Great Lakes region under the scenario of doubling of CO_2 concentration. Climate change scenarios considered by Mortsch et al. (2000) indicated declines in runoff and lake levels that could lead to potential water allocation problems in the region. Southam et al. (1999) evaluated the impact of climate change in Ontario's Grand River basin under 21 scenarios of future surface water supplies, streamflow regulation, population and water use. It was concluded that climate change may have serious impacts on the capability of Grand River to assimilate wastewater and yield a reliable supply of water for municipal purposes while maintaining existing water quality standards.

At present, there is no ideal method for generating future climate scenarios (Gleick, 1989; Simonovic, 2001). Existing methods can be classified into three categories: empirically-based, process-based and linked methods that combine empirically-based and process-based concepts. Empirically based models make use of the historical observations to identify trends in important climatic variables such as temperature and rainfall and changes in important weather patterns such as El Nini-Southern Oscillation (ENSO) cycle. Process based models use mathematical representations of the processes that govern atmospheric and oceanic circulation to estimate future climate variables and seasonal changes in climate. Global circulation models (GCMs) are the most sophisticated process-based models that simulate the climate system. During the last decade, a number of complex GCMs have attempted to simulate future anthropogenic climate change scenarios. GCMs have predicted considerable warming and changes in precipitation pattern under the well know scenario of doubling of CO_2 emissions. If these changes materialize, it is suggested that the ice regime may be modified in different ways in different regions. For example, in temperate regions such as the southwestern Ontario, the brief and capricious river ice cover may disappear completely, or become more intermittent (Clair et al., 1996). This could be beneficial to the socio–economic sectors but would be harmful to the aquatic life that depends on the ice cover for winter survival. The projections made by GCMs are usually broad and it is not possible to identify specific areas and communities that may be vulnerable to climate change, or to anticipate the magnitude of economic and ecological impacts. These limitations are caused mainly due to the relatively large spatial resolution of GCMs, which is of the order of 2°×2.5° in the horizontal (latitude × longitude). Such a resolution is unsatisfactory for catchment level hydrologic processes and gives rise to uncertainties when downscaling is carried out using the output from a GCM. Promising work addressing the issue of spatial resolution has been carried out by Hughes and Guttorp (1994), and Wilby (1994). But still there is a great deal of uncertainty regarding the regional GCM output under future scenarios of increasing CO_2 and aerosol changes. Further, a climate change scenario based on the output from a GCM represents only one of the many future climate change scenarios whereas exploring several alternative climate scenarios would be more useful for effective management of water resource systems. Estimates of weather variables particularly precipitation on finer geographic and temporal scales are needed to predict the potential effects of climate change on a regional scale. Development of local weather generators to model hydrological impact of climate change has largely been motivated by the acknowledged limitations of GCMs in evaluating the regional climatic impacts.

This report presents a generic weather generator, based on the K-nearest neighbour (K-NN) algorithm (Buishand and Bradsma, 2001; Rajagopalan and Lall, 1999; Yates et al., 2003) for producing a variety of synthetic weather sequences that can be used as an input into hydrological models. Particular emphasis is laid on the generation of weather sequences that model unprecedented precipitation events in the basin. The rest of the report is organized as follows. A brief background on stochastic weather generators is presented in the next section. Section 1.3 outlines the methodology used to adapt K-NN algorithm for simulating daily weather sequences conditioned upon alternative climate change scenario. Procedure for strategic resampling is described in section 1.4. Application of the algorithm to

Upper Thames River basin in Canada is described in section 1.5. Model results are presented in the section 1.6. The report concludes with the summary of results and a discussion of the findings.

1.2 Background

Some traditional weather generating approaches are discussed in this section. Stochastic generation of weather variables especially precipitation has been an extensive research topic. A number of weather generators based on parametric statistical techniques have been effectively used to generate plausible climate scenarios, and have themselves been used as downscaling techniques in global climate change studies (Wilks, 1992). Daily weather generators are most common due to the wide availability of meteorological data on this time scale, and due to the fact that most impact assessment models are driven by daily data. The traditional weather generating approach (Nicks and Harp, 1980) focuses on independent generation of precipitation first while the remaining variables are modelled conditioned upon precipitation occurrence (i.e. precipitation or no precipitation). Daily precipitation amounts are generated using a two-state first order Markov model from an assumed probability distribution fitted to the observed values. Different model parameters are fitted to each period in order to capture the seasonality in the values of the variables themselves and in their cross-correlations. More complex models describe more than one precipitation state (e.g. low, medium, and high precipitation amounts). Todorovic and Woolhiser (1975) combined the first order Markov model for daily precipitation occurrence with a statistical model for daily non-zero precipitation amounts. Exponential distribution was used to describe the precipitation amounts. More elaborate models have been proposed for the distribution of precipitation amounts given the occurrence of a wet day. Katz (1977), Buishand (1978), and Stern and Coe (1984) used two-parameter gamma distribution to describe the occurrence of precipitation amount on wet days. Smith and Schreiber (1974), Woolhiser and Roldan (1982), and Wilks (1999) fitted three-parameter mixed exponential distribution to describe precipitation amounts on wet days. An excellent review of stochastic weather models has been presented by Wilks and Wilby (1999).

Although a large number of precipitation models have been developed, many practical applications require that weather generators produce other meteorological variables in addition to precipitation. To generate other variables probability distributions are fitted independently for each variable for each period and for each precipitation state. Assumption is made that each variable is conditionally independent and identically distributed. Richardson (1981) describes a Markov-chain exponential model in which the precipitation is generated independently of the other variables. The other variables are generated by using a multivariate model with the parameters of the variables conditioned on the wet or dry status of the day as determined by the precipitation model. Stochastic weather generators of the type proposed by Richardson (1981) are commonly referred to as WGEN (for 'weather generator' as in Richardson and Wright, 1984). WGEN is a multivariate time series model that can be used to generate stochastically the daily values of maximum and minimum temperatures, precipitation and solar radiation for any required length of time. Daily precipitation in WGEN is modelled by a two parameter gamma distribution which tends to match the observed data significantly better than the simple exponential distribution used by Richardson, 1981.

Crop production and natural ecosystem models often require additional weather variables such as wind speed and relative humidity. Nicks et al. (1990) describe an extended version of WGEN called WXGEN that takes into account the non-normal distribution of wind speed and relative humidity. However, wind speed is not linked to any of the other variables and relative humidity is linked only to precipitation occurrence. Wind speed and dewpoint (from which relative humidity can be derived) are included in the weather generator GEM (Generation of weather elements for Multiple Applications) developed by Hanson and Johnson (1998). Their model permitted dependence among all the weather variables assuming normal distribution for these two variables. Parlange and Katz (2000) further extended WGEN to include daily mean wind speed and dewpoint in the model. The key to the extension is the transformation of the variables that are not normally distributed, for example power transformation was used to take into account the positively skewed distribution of wind speed. Their model is effectively a hybrid of WXGEN and GEM, combining the individual improvements of these stochastic weather generators. Application of the model to the data in the Pacific Northwest was presented.

A major drawback associated with the 'Richardson type' weather generators is that the persistent events such as drought or prolonged rainfall are not very well reproduced by them. To overcome this problem, serial approach to weather generation has been presented by Rackso et al. (1991) and Semenov et al. (1998) among others. In this approach, the sequence of dry and wet series of days is modelled first and the precipitation amounts and other variables are generated conditioned on the wet or dry series. Rackso et al. (1991) used predefined distributions for modelling of wet and dry series whereas semi-empirical distributions are used in LARS-WG (Semenov and Barrow 1997; Semenov et al., 1998). Since LARS-WG uses every single observation in the modelling process, it is expected to perform better than the models such as those of Rackso et al. (1991) that are based on the fitting of a predefined distribution to the observed data. Performance evaluation of WGEN and LARS-WG at 18 sites chosen from different parts of the world carried out by Semenov et al. (1998) confirmed the superiority of LARS-WG. It matched the observed data more closely than the WGEN which may be attributed to the use of more complex distributions in LARS-WG. Both generators, however, had difficulty in reproducing the annual variability in monthly means of the variables.

A number of applications of weather generators for multisite simulation of variables have been reported in the literature. Smith (1994) presented an extension of the Markov model to bivariate time series of daily precipitation at two different stations. Further extension to more stations was limited by the number of parameters needed to be incorporated in the model. Wilby (1994) developed a stochastic model for the synthesis of daily precipitation data by weather type analysis. The model was applied to generate daily rainfall at two sites in southern England. Wilks (1998) developed a multisite version of first order Markov model with mixed exponential distribution for wet day precipitation amounts. Use of mixed exponential distribution rather than gamma distributions produced synthetic sequences with interannual variability similar to the observed data. Means, variances, and interstation correlations of monthly precipitation totals were well preserved in the model proposed by Wilks (1998).

A number of parametric weather generators have been developed but they have several drawbacks. First and most importantly, they do not reproduce various aspects of spatial and temporal dependence of variables adequately. Second, an assumption has to be made regarding the form of probability distribution of the variables, which is often subjective. Third, non Gaussian features in the data cannot be adequately captured as multivariate autoregressive models (MAR) models implicitly assume a normal distribution which is difficult to satisfy. Fourth, a large number of parameters are separately fitted to each period and the number further increases if the simulations are to be conditioned. Fifth, the models are not easily transportable to other sites due to the site-specific assumptions made regarding the probability distributions of the variables.

Nonparametric methods can circumvent most problems associated with the parametric methods. Simple nonparametric techniques essentially involve random resampling from the historical data to generate synthetic sequences of required duration. Such sequences often fail to capture time correlation of the data series. Complex procedures have been developed that can capture to a large extent the prominent time correlation between the weather data. The most promising non parametric techniques for generating weather data is the K-NN resampling approach. The works of Young (1994), Lall and Sharma (1996), Lall et al. (1996), Rajagopalan and Lall (1999), Buishand and Brasma (2001), and Yates et al. (2003) describe various forms of K-NN resampling scheme. A K-NN algorithm typically involves selecting a specified number of days similar in characteristics to the day of interest. One of these days is randomly resampled to represent the current day's weather. Young (1994) employed a K-NN strategy to select a day randomly from amongst the 3 to 5 nearest neighbors. A discriminant function was used to identify the days having weather closest to the current day's weather. Young's model mostly preserves the correlation between the temperature and the precipitation and the wet or dry spell statistics. Simulated sequences, however, showed reduced persistence and underestimation of the fraction of dry months.

Lall and Sharma (1996) presented a nearest neighbour bootstrap for resampling hydrologic time series. Multivariate nearest neighbour probability density estimation provided the basis for the resampling scheme developed. Resampling is done from k nearest neighbors in terms of a weighted Euclidean distance. Prior assumptions regarding the distribution of the precipitation amounts were not necessary. To preserve temporal dependence, a new day is resampled from the historical data set by conditioning on the simulated values for previous days. Rajagopalan and Lall (1999) compared nearest neighbour resampling with a parametric time series model. Six daily weather variables were simultaneously generated for Salt Lake City, Utah. Comparison of the results with those obtained using a parametric time series model clearly demonstrated the superiority of the nonparametric approach. Unlike the method of Young (1994) who used only 3-5 nearest neighbors for resampling, Rajagopalan and Lall (1999) used relatively large number of nearest neighbors for resampling. Lall et al. (1996) used K-NN resampling scheme with kernel density estimators to represent the probability distributions of dry spell lengths, wet spell lengths, and wet day precipitation amounts.

Brandsma and Buishand (1998) describe the application of nearest neighbour resampling procedure to single site simulation of daily precipitation and temperature for multiple stations in the Rhine Basin. Conditional simulation of weather variables on atmospheric flow was also considered by them. Bardosy and Plate (1992) also describes a space-time model for generating daily precipitation data using atmospheric circulation patterns. Buishand and Brandsma (2001) extended the nearest-neighbor resampling to simultaneous simulation of daily precipitation and temperature at multiple stations in the German part of the Rhine Basin. A moving window of 61 days was used. For a historical record of N years the nearest neighbors are selected from 61N days. Since the resampling is done from the historical data, the correlation between precipitation at different stations and that between daily precipitation and temperature is automatically preserved. Yates et al. (2003) describe a K-NN resampling scheme that largely preserves important cross correlations and autocorrelations. Mahalanobis distance metric (Davis, 1986), was used to determine the closeness of any given neighbor to the current vector. Use of Mahalanobis distance is considered superior than the Euclidean distance approach as it does not require explicit weighing and standardization of the variables. A resampling strategy is introduced that can be used to generate desired climate scenarios. A simulated annealing (SA) based non parametric approach has been presented by Bardosy (1998) in which the reshuffling of data is carried out in a manner that retains important statistical properties of the observed data series. The SA approach is quite intensive computationally but may prove to be particularly useful for simulation of series with short time steps.

1.3 The K-NN Algorithm

This section describes the K-NN algorithm used in this study. Consider that the daily historic weather vector consists of p variables. Here p = 3 including maximum temperature (TMX), minimum temperature (TMN), and precipitation (PPT)). Suppose the number of stations considered in the model is q and the data is available for N years. Let X_t^j denotes the vector of weather variables for day t and station *j,* where $t = 1,...,J$, and $j = 1,...,q$; J being the total number of days in the time series. The vector consisting of the current day weather variables is called the feature vector and can be expressed, in expanded form, as $X_t^j = [x_{1,t}^j, x_{2,t}^j,..., x_{p,t}^j]$ where $x_{i,t}^j$ represents the value of the weather variable i at time step t and for station j. Suppose that the simulation with K-NN model begins on day t corresponding to January 1. The algorithm cycles through various steps to select a day having weather closest in characteristic to the current day's weather from amongst the predetermined number of nearest neighbors. The weather of the selected day is adopted to represent the weather for given day in the simulation period. The closeness of the current weather vector, X_t^j to various potential neighbors is determined using the Mahalanobis distance metric which does not require explicit weighing and standardization of the variables (Yates et al., 2003). Buishand and Brandsma (2001) selected k-nearest neighbors in terms of a weighted Euclidean distance which requires that the weather variables be standardized. Different steps of the algorithm are described next.

1. Compute regional means of the p variable across the q stations for each day of the historical record

$$\overline{X}_t = [\overline{x}_{1,t}, \overline{x}_{2,t},..., \overline{x}_{p,t}]$$

where $\bar{x}_{i,t} = \frac{1}{q}\sum_{j=1}^{q} x_{i,t}^{j}$, $i = 1,\ldots,p$

2. Determine the size, L of data block that includes all potential neighbors to the current feature vector from which resampling is to be done. A temporal window of width w is chosen and all days within window are considered as potential candidates to the current feature vector. Yates et al. (2003) used a temporal window of 14 days which implied if that if the current day is January 20 then the window of days consist of all days between 13 January and January 27 for all N years but excluding January 20. Thus, the data block of potential neighbors from which to resample consist of $L = (w+1)*N-1$ days. A fixed length, 14-day temporal window was used in this study. For $w = 14$ and $N = 38$, $L = 569$.

3. Compute mean vectors across q stations for each day in the data block consisting of potential neighbors using the expressions given in step 1.

4. A covariance matrix C_t is computed for current day t using the using the data block of size $L \times p$.

5. The weather on the first day t (e.g., 1 January) comprising all p variables at q stations is randomly chosen from the set of all days of the historic record of N years (e. g., all January 1 days have equal probability of selection) and includes all p variables. The algorithm proceeds to select one of the nearest neighbors to represent the weather of the given day in the simulation period.

6. Compute Mahalanobis distances between the mean vector of the current day's weather, $\overline{X}_t$ and the mean vector for day i, $\overline{X}_i$ where $i = 1,\ldots,L$.

$$d_i = \sqrt{(\overline{X}_t - \overline{X}_i)C_t^{-1}(\overline{X}_t - \overline{X}_i)^T}$$

where T represents the transpose operation, and C_t^{-1} is the inverse of covariance matrix

7. Determine the number of first K nearest neighbors to be retained for resampling out of the total of L neighbors.. The choice of K is vital for best reproduction of the desired statistics in the simulated sequences. Resampling with small number of nearest neighbors is unlikely to maintain diversity in the simulated sequences. Similarly, a relatively large value of K might not reproduce the required statistics in the simulated data. Lall and Sharma (1996) suggested the use of the generalized cross validation score (GCV) which is similar to Akaike information criteria (AIC) in the traditional AR models for choosing K. Rajagoplan and Lall (1999) and Yates et al. (2003) have recommended the use of heuristic method to choosing K according to which $K = \sqrt{L}$. The performance of algorithm with this value of K was found to be good. In this study $L = 569$, and hence a value of K equal to 24 has been adopted.

8. Sort the Mahalabonis distances in ascending order and retain the first K nearest neighbors. Assign weights to each of these j neighbors according to the probability metric defined as $p_j = \frac{1/j}{\sum_{i=1}^{K} 1/i}$. The neighbor with the shortest distance is assigned the highest weight, where the neighbor with the longest distance (i.e. the Kth neighbour) gets the least weight. Lall and Sharma (1996) developed this function through a local Poisson approximation of the probability density function of state space neighbors.

9. The weather on the given day in the simulation period is represented by the day $t+1$, which is selected from amongst the K-nearest neighbors. To obtain the day $t+1$, a random number , $r \subset (0,1)$ is first generated and if $p_1 < r < p_K$, then a day j for which r is closer to p_j is selected. If $r \geq p_1$, the day corresponding to d_1 is selected and if $r \leq p_K$, then the $t+1$ day corresponding to d_K is selected.

Steps 6 to 9 are repeated to generate as many years of synthetic data as required. If multiple sequences of data are required, then the algorithm starts at step 5.

With K-NN algorithm, the spatial dependence is preserved by resampling simultaneously the same day weather as the weather for all the stations. To preserve temporal dependence, a new day is resampled from the historical data by conditioning on the simulated values for previous days. Due to the manner in which the given day's weather is simulated, autocorrelations for different variables, cross correlations between the variables and interstation correlations are most likely to be preserved. Further, no assumptions need to be made regarding the probability distributions of various variables.

1.4 Strategic Resampling

This section describes how strategic resampling could be carried out to generate synthetic weather sequences with required attributes. In hydrological studies, it is often desired to test various models with the weather data that consists of divergent pattern relative to the historically observed climate, such as a gradual warming trend over a certain period of time and region. Exploring the response of hydrological models to a single sequence of weather data would lead to a solution with limited practical significance. Ideally, the response of the model should be evaluated with an ensemble of weather sequences so that a better understanding of the complex phenomenon that drives the model could be obtained. With strategic resampling, a large variety of weather sequences may be generated and subsequently used as an input into hydrological models. Strategic resampling simply implies that a new set of years from the historical record based on some prescribed conditioning criteria shall be used in the K-NN algorithm to derive new weather sequences with required attributes. With strategic resampling, it would also be possible to generate synthetic weather sequences that adequately model the drought and flood conditions.

Strategic resampling can be carried out in the following manner. Assume that a year has l periods of d days each. The regional periodical mean for variable $y^i_{j,t}$ where i is the year, j is the station and t is the day, may be computed as follows

$$M^i_l = \frac{1}{d \times q} \sum_{j=1}^{q} \sum_{t=1}^{d} y^i_{j,t}$$

where M_l^i is the regional periodical mean corresponding to period l and year i.

The regional periodical mean for period l of the entire record may be computed as

$$\overline{M}_l = \frac{1}{N}\sum_{i=1}^{N} M_l^i$$

The regional periodical deviations for each year and for each period are computed as $D_l^i = M_l^i - \overline{M}_l$

Once the deviations are computed for different years and different periods, a ranked list of years for a particular period can be generated by sorting the years according to the magnitude of deviations of that period. Suppose a ranked list for the month of January is required, then different years are ranked according to the deviation computed for January. The deviations for each year are computed as the difference between the mean value of the variable in January for that year and the overall historical mean of the variable as described above. An index is then assigned to each year in the ranked list based on the relative position of that year in the sorted list. A general integer function of the following form can be used to select different years from the ranked list (Yates et al., 2003).

$$I_w^i = INT[N(1 - r^{S_w^i})] + 1$$

where I_w^i is the index corresponding to year i and period w, r is a normally distributed random number between 0 and 1, S_w^i is the shape parameter that can be suitably adjusted to bias certain years over others. Suppose that the years in the ranked list are arranged such that the coldest year has an index of 1 and the warmest year has an index of N. If it is required to create bias towards the selection of warmer years, then S_w^i should be assigned a value greater than 1. Similarly, values S_w^i less than 1 would create bias towards selection of colder years. The exact value of S_w^i would, however, depend upon the amount of biasing required. If no biasing is required, value of shape parameter S_w^i is set to 1. To produce a subset of years that could be used to generate a weather sequence with attributes similar to that of historical data series, a simpler function that returns random integers between the specified upper and lower bounds may be used.

1.5 The Upper Thames River Basin

The Thames River watershed is nestled in the agricultural heartland of southwestern region located in the Canadian province of Ontario. The southwestern Ontario is a highly developed region and as such the basin faces pressure from urban and rural land uses. The Thames River is the major river of the basin. It is 273 km long and has a catchment area of around 5,825 km^2, making it the second largest watershed in southwestern Ontario. Despite these pressures, The Thames remains one of the most biologically diverse rivers in Canada. The water quality of the Thames River is impacted by drainage practices, runoff, spills and bank alterations among others. Most of the precipitation comes in the form of winter snow. Rainfall occurs mainly in spring, with some in fall.

1.5.1 Data Description

Daily maximum temperature, minimum temperature and precipitation data from 9 nine stations in the basin was used for the period 1964-2001. The geographical location of stations as determined from their latitudes and longitudes is shown in Figure 1. The data used in this study is Environment Canada corrected. The mean annual values of different weather variables and the latitude and longitude of each meteorological station are presented in Table 1. The meteorological stations in the basin are distributed across an area of approximate dimension 60 km (east – west) by 50 km (north-south). The interstation distances range from approximately 10 km to 60 km.

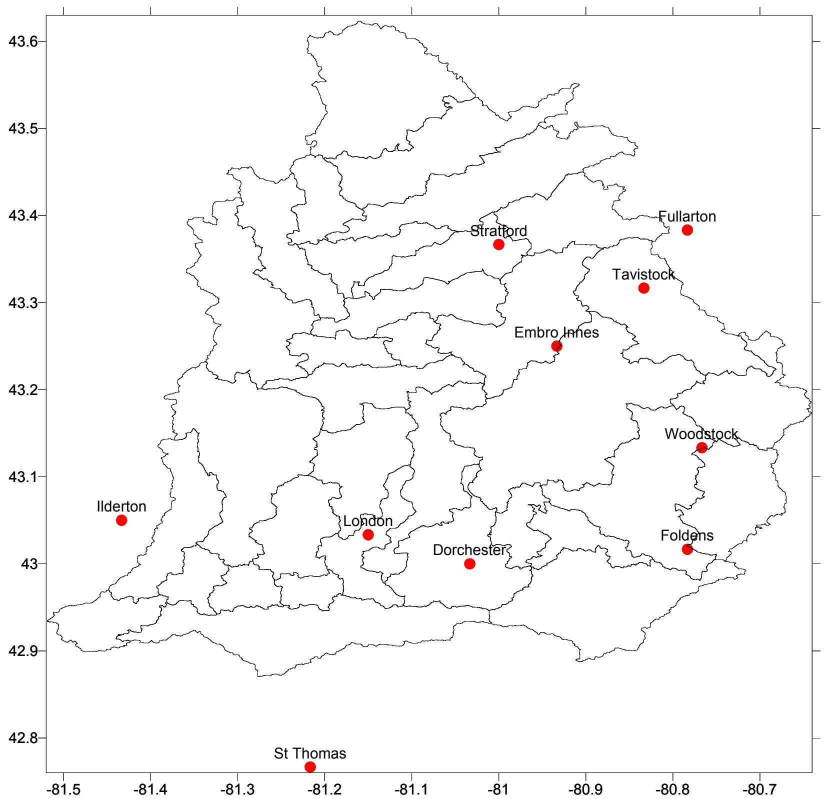


Figure 1 Geographical location of different stations in the basin

There were a large number of missing records in the available data which were filled-in using a two-step procedure. In the first step, missing records for London, Ilderton, Foldens, Stratford and Woodstock were filled in with the mean values. Once the data set for these five stations was complete, the missing records for Embro, Dorchester, Tavistock and Fullarton were filled in. At these four stations the precipitation records were available but maximum temperature and minimum temperature records were missing for the entire period. The weighted average inverse distance square method was used to estimate the missing temperature data for these four stations. The method has the advantage in that the estimated values will always be less than the greatest and greater than the smallest value of the temperature at the surrounding stations. The calculation of distances between various stations was based on the latitude and longitude of the stations which are shown in Table 1. It was not possible to use the weighted average inverse distance square method for London, Foldens, Startford and Woodstock data as there were many days in the record for which the data is missing for either one or more of the remaining stations i. e. Embro, Dorchester, Tavistock and Fullarton.

Table 1 Station Characteristics

S. No.	Station	Latitude (Deg N)	Longitude (Deg W)	Mean Ann. TMX. (C)	Mean Ann. TMN (C)	Total Ann. PPT (mm)
1	Foldens	$43^{\circ}\ 1'$	$80^{\circ}\ 47'$	11.98	3.21	945
2	Ilderton	$43^{\circ}\ 3'$	$81^{\circ}\ 26'$	12.71	3.27	1010
3	London	$43^{\circ}\ 2'$	$81^{\circ}\ 9'$	12.37	2.42	980
4	Stratford	$43^{\circ}\ 22'$	$81^{\circ}\ 0'$	11.43	2.37	1056
5	Woodstock	$43^{\circ}\ 8'$	$80^{\circ}\ 46'$	12.45	2.52	942
6	Embro	$43^{\circ}\ 15'$	$80^{\circ}\ 56'$	11.88	2.52	984
7	Dorchester	$43^{\circ}\ 0'$	$81^{\circ}\ 2'$	12.28	2.57	1035
8	Tavistock	$43^{\circ}\ 19'$	$80^{\circ}\ 50'$	11.83	2.50	1048
9	Fullarton	$43^{\circ}\ 23'$	$80^{\circ}\ 47'$	11.80	2.50	1013

1.6 Model Application

The K-NN model described above was applied to the data from Upper Thames River Basin in southwestern Ontario province of Canada. A simulation period of around 20 times the length of historical record is generally considered sufficient to provide a reliable estimate of the desired statistics. Model runs were therefore carried to generate 800 years of synthetic weather data for various potential climate change scenarios. For each scenario considered, the statistics of interest are computed from the simulated sequence and compared to the statistics of observed record using box plots. Box plots are preferred method of data analysis in many applications as they show the range of variation in the statistics of simulations and provide a straightforward method of comparing the statistics of simulations with the historical data.

The statistics of interest considered here are mean, standard deviation and correlation coefficients of the data. The bottom and top horizontal lines in the box in a box plot indicate the 25th and 75th percentile respectively of the statistics computed from the simulated data. Median is represented by the horizontal line within the box. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. The whisker extends to the most extreme data value within 1.5 times the interquartile range of the data. Outliers are data with values beyond the ends of the whiskers and have been shown by dots. The statistics of the historical record are represented by dots and joined by solid lines. The dependence nature is evaluated using lag-0 cross correlation values across the variables. Time dependence of variables is evaluated using lag-1 autocorrelation. Interstation correlations are described by pairwise scatter plots. Alternative climate change scenarios are described in the next section.

1.6.1 Case 1: Reproduction of Historic Data Statistics

The first simulation was carried out to produce a series with nearly the same statistical attributes as the historic data series. A new subset of years that constitute the driving data for the model is obtained by using an integer function that returns integers between the specified upper and lower bounds. In our model, the upper bound was set to N, and lower bound was 1. To generate N years of data, the integer function was called N number of times. With this method, each year has equal probability of being selected. However, some years may be selected more than once. A new data set is thus obtained and the K-NN algorithm uses this data set to generate required number of years of synthetic data. Comparison of various statistical attributes computed from the K-NN simulations with those for the observed record is presented below. Results are presented for London only since the results for other stations are almost similar.

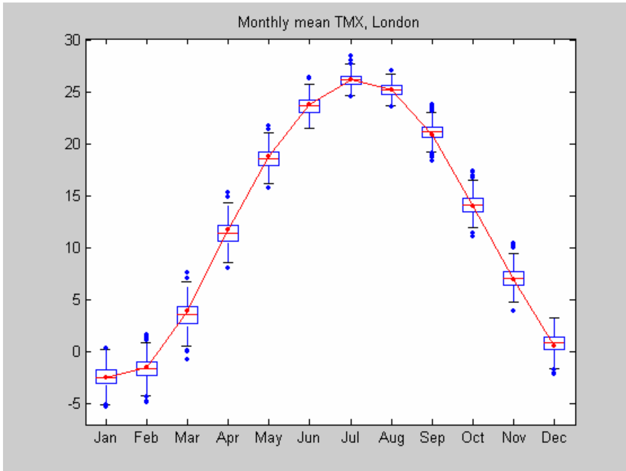


Figure 2 Box plots of monthly mean maximum temperature

Figure 2 shows the box plots of simulated values of mean TMX values for London. Although the model was applied on daily data, the statistics from the daily data have been aggregated to the monthly timescale to facilitate presentation of the results. The statistics of simulations are shown by box plots while the solid lines with dots show the same statistics for the historical data. Comparison of historical monthly values with the simulated values clearly showed that the model was able to adequately reproduce the historical values. The model slightly underestimated the mean TMX for the month of April but for the rest of the months, the simulated values are very close to the observed historical values. This is highly satisfactory given the fact that monthly statistics are not explicitly specified in fitting the K-NN model, unlike the parametric models that are fitted separately to each month. It may be recalled that since the K-NN model is driven by daily weather data, reproduction of monthly statistics may be considered as the indication of the robustness of the algorithm.

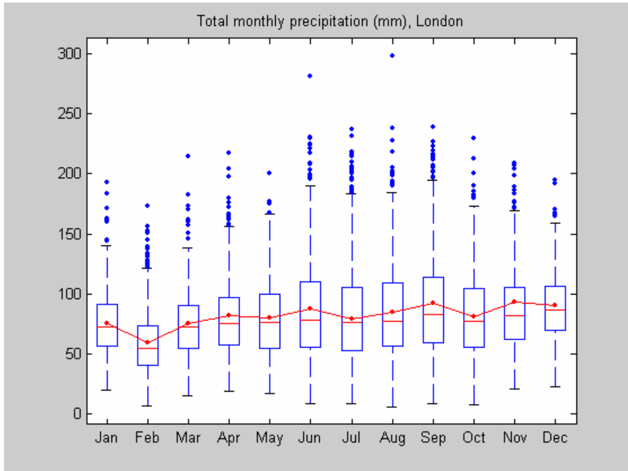


Figure 3 Box plots of total monthly precipitation

Figure 3 provides the box plots of total monthly precipitation for London. It can be seen from the box plots that the historical mean of the total precipitation is close to the median of the simulated data for all the months. A number of values were found to lie beyond the whiskers but these outliers constitute only a fraction of the total number of years (800) simulated. Moreover, the outliers give an indication of the variety in the simulated data. The total annual precipitation simulated by the model (984 mm) matched very closely with the historical value (980 mm). Overall the performance of the model in simulating the total monthly precipitation was very good.

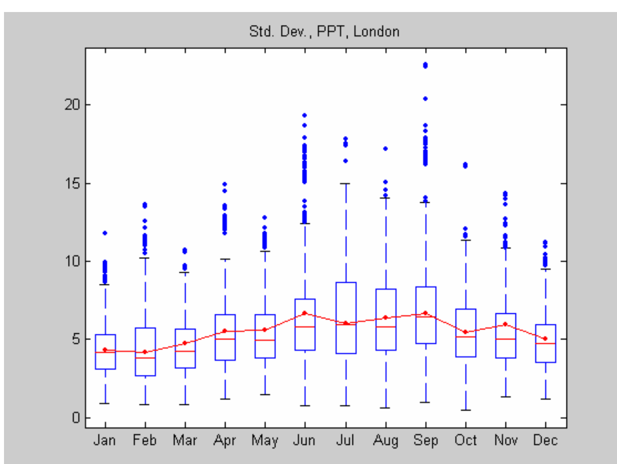


Figure 4 Box plots of standard deviation of precipitation

Simple parametric time series models for daily precipitation often underestimate the standard deviation of monthly totals (Buishand, 1978, Katz and Parlange, 1998). Therefore, it is important to compare the standard deviations of monthly totals when a non-parametric technique such as the K-NN algorithm is used. The box plots of standard deviations of precipitation aggregated to monthly time scale are shown in Figure 4. As can be seen from the box plots the model satisfactorily reproduced the standard deviations of the historical data.

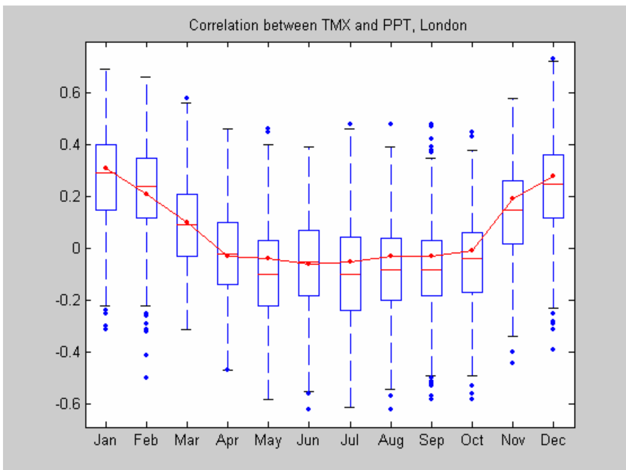


Figure 5 Box plots of correlation between TMAX and PPT

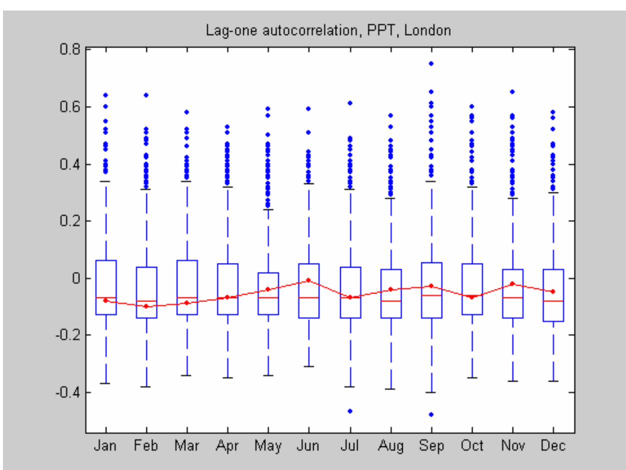


Figure 6 Lag-one autocorrelation of PPT at London

Parametric models often fail to reproduce the correlation structure among the variables. To evaluate the performance of K-NN algorithm based nonparametric model with respect to reproduction of correlation statistics, it was decided to analyse the correlation structure produced by the model, and to compare it with the historical structure. Box plots for correlation between TMX and precipitation, and autocorrelation of PPT are shown in Figure 5 and Figure 6 respectively. It can be observed from the box plots shown in Figure 5 that there is a positive correlation between TMX and PPT during the winter months while the correlation is negative during the summer months. The K-NN model adequately reproduced the historical correlation structure as shown by the box plots. It can be seen from Figure 6 that the mean values of autocorrelation coefficients for different months of historical record are close to 0 which implies a very weak lag-1 autocorrelation of PPT, and the model adequately captured this characteristic of the observed data.

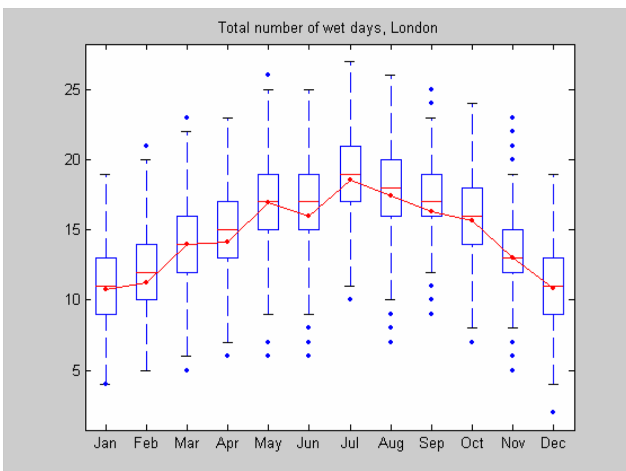


Figure 7 Box plots of total number of wet days

Figure 7 shows the box plots of total number of days with precipitation events for London. This statistics is important for sequences that are generated with the intention of use in the crop production and flood management models. It can be seen that the model reproduced the historical statistics very well. There was a slight overestimation for the months of April and September, and underestimation for June, however.

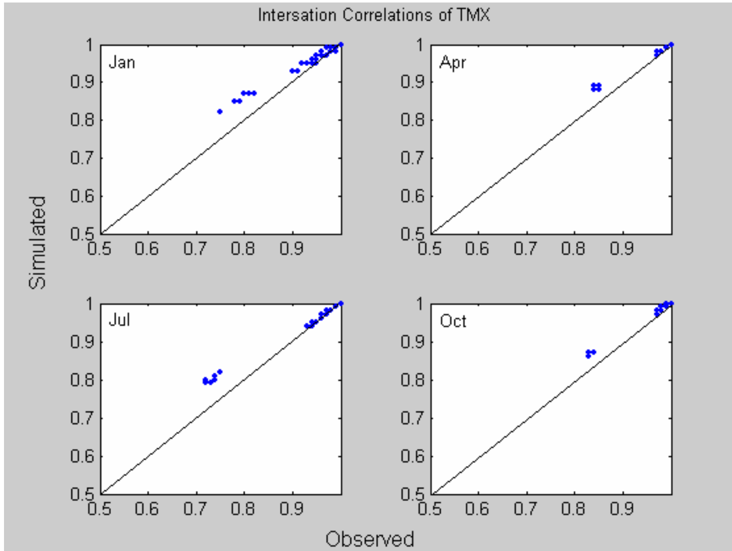


Figure 8 Comparison of observed versus synthetic interstation correlations for mean monthly TMX between all station pairs for four representative months

Figure 8 shows scatter plots of interstation correlation coefficients for mean monthly TMX values in the simulated and the observed data. For q stations, there are $q(q-1)/2$ pairwise correlations resulting in 36 such correlation coefficients for each month. The scatter plots have been shown for four representative months. The observed correlation coefficients are plotted on the horizontal scale while the simulated values are plotted on the vertical scale. It can be seen from Figure 8 that there are strong interstation correlations between TMX values, mostly in the range of 0.9 to 1.0. Almost all data points lie in the close vicinity of the 45° sloping solid line shown in the box plots. Clearly, the performance of K-NN model in reproducing the historical interstation correlation structure is very good.

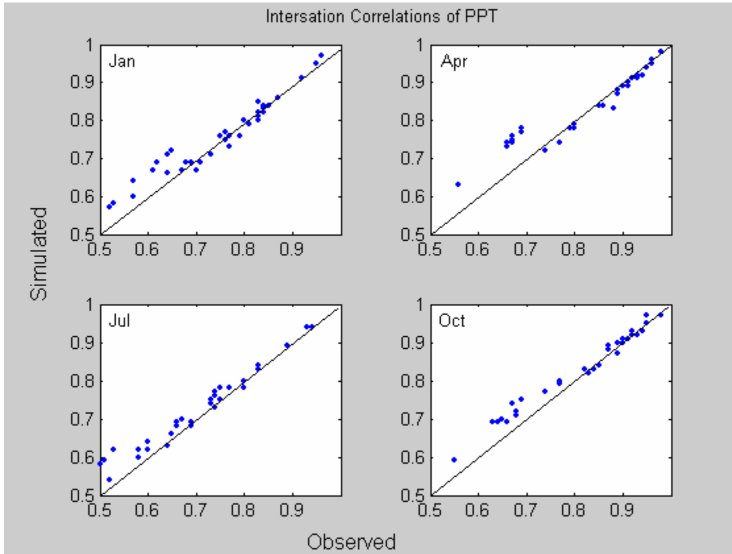


Figure 9 Comparison of observed versus synthetic correlations for monthly total PPT between all station pairs for four representative months

The scatter plots of interstation correlations of total monthly PPT between the observed and the simulated data are shown in Figure 9. Although the correlations are not as strong as observed in the case of TMAX, the model reproduced the historical structure very well. The performance of the K-NN with regard to the reproduction of interstation correlations is extremely good, both for the TMX and PPT. It is a well known problem that parametric techniques are often unable to reproduce the correlation structures adequately but the manner in which the K-NN algorithm works makes it possible for it to reproduce extremely well both the temporal and spatial correlation structure.

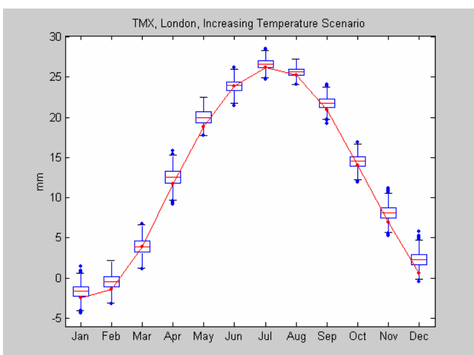
1.6.2 Case 2: Increasing Average Temperature Scenario

To assess the performance of water resource systems under a gradual warming trend over a certain period of time, a new data set comprising of years with increased average temperatures is required. Such a data set can be obtained by resampling strategically from a ranked list generated on the basis of the deviations of mean annual average temperature from the long term historical mean. Since an increasing average temperature scenario is required, the average temperature (TAV) for each station was calculated as a weighted average of TMX and TMIN

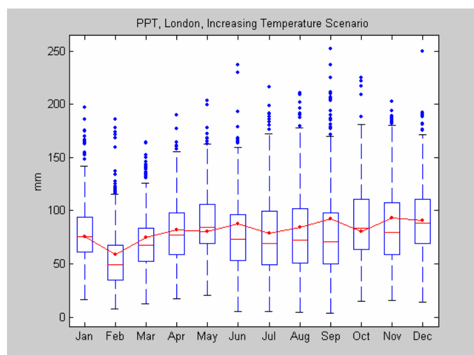
(TAV= TMX*0.6+TMN*0.4). The mean annual temperature for year i, TAV_i was computed for each year of the historical record. The overall long term mean of TAV was computed as follows.

$$\overline{T} = \frac{1}{N}\sum_{i=1}^{N} TAV_i$$

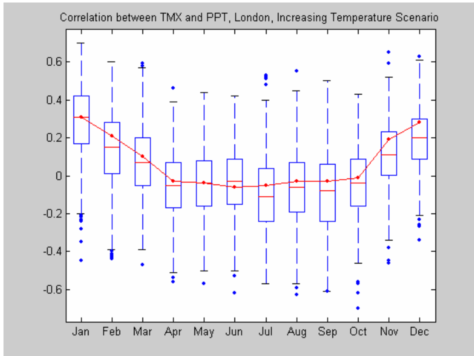
To compute the deviation for each year, the overall long term mean, $\overline{T}$ is subtracted from the mean yearly value, TAV_i for that particular year. On the basis of deviations, a ranked list of years is generated with the first rank corresponding to the year with the lowest deviation and the last rank corresponding to the year with the highest deviation. Using the integer function described earlier, index values are generated which directly corresponds to certain years in the ranked list. Biasing of certain years over others can be carried out by choosing appropriate values of the shape parameter S_i^l. With a value of shape parameter of 3, an increase in the overall mean historical TMX of approximately 1°C was obtained. Once the new data set with increased values of TAV is obtained, K-NN model is executed to generate 800 years of synthetic data. The results of simulation for London are presented through the box plots shown in Figure 10.

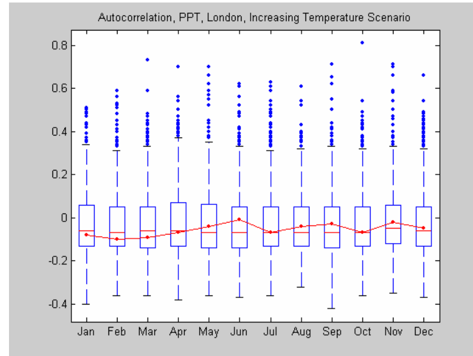


(a)



(b)





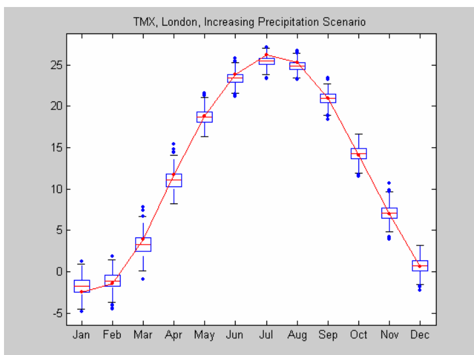
(c) (d)

Figure 10 Box plots for increasing TAV scenario: (a) monthly TMX, (b) total monthly PPT (c) correlation between TMX and PPT, and (d) autocorrelation of PPT

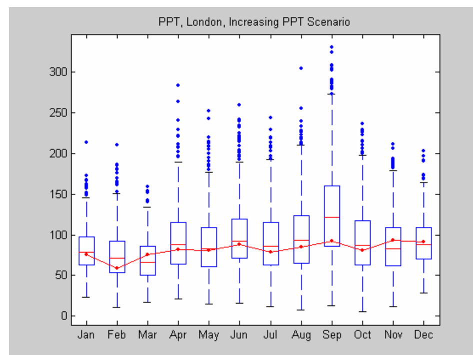
The box plots in Figure 10 (a) clearly show that the model did produce increases in TMX over the historical values for all the months. The historical data here represents the actual observed data and not the data series obtained by using the index function. Increases in TMN were also seen thereby indicating that the increase in TAV has been achieved. This implies that with strategic resampling, the model is capable of producing alternative climate change scenarios with desired attributes. The effect of increasing TAV on total monthly precipitation is shown in Figure 10(b). It appears that there has been a decrease in precipitation for most of the months except for January, October and December when the precipitation remained nearly the same as the historical. Correlation between TMX and PPT is shown in Figure 10(c). As expected, the correlation structure of historical values is preserved in the simulations. The correlation between TMX and PPT for winter months is positive and this trend has been well captured by the simulation model as shown in Figure 10 (c). Box plots of autocorrelation coefficients for the scenario are presented in Figure 10(d). There is a weak autocorrelation for precipitation and the model reflected this trend.

1.6.3 Case 3: Increasing Precipitation Scenario

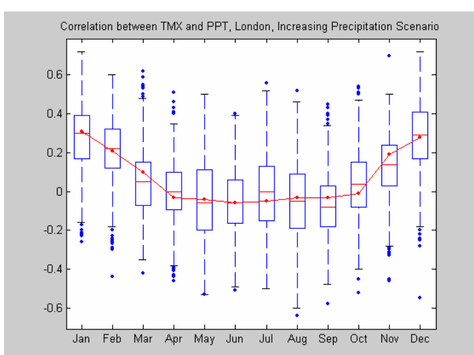
The next simulation was carried out to generate a scenario with increased precipitation over different months in a year. Resampling procedure similar to the one outlined in the previous section is followed. The deviations are, however, computed for the precipitation rather than for the average temperature. Once the ranked list of years is generated, strategic resampling is carried out to bias wetter years over dry years to generate an increased precipitation scenario. A value of S_i of 2 resulted in an increase of mean annual precipitation to 1033 mm while a value of 3 increased the mean annual precipitation to 1088 mm. The historical value of mean annual PPT is 980 mm. A new data set comprising of years with increased annual precipitation is obtained and used as the driving data set for the K-NN model.



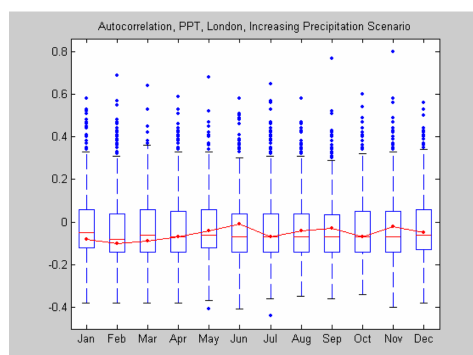
(a)



(b)



(c)



(d)

Figure 11 Box plots of statistics for increasing precipitation scenario: (a) Monthly TMX, (b) Total Monthly PPT (c) Correlation between TMX and PPT, and (d) Autocorrelation of PPT

Figure 11 shows the box plots for K-NN simulations as well as historical values for the increasing precipitation scenario for $S_i = 3$. It can be seen from Figure 11(a) that model produced some increase in TMX values for the month of January and February. As seen earlier (Figure 5) historically there is a positive correlation between TMX and PPT during the winter months (November, December, January and February). Owing to this positive correlation, increase in precipitation in January and February is accompanied by an increase in TMX values. There appears to be a slight increase in PPT for November and December but the corresponding increase in TMX is insignificant for these months. The maximum increase in PPT is obtained for September (nearly 40 mm) but similar increase in TMX is not visible, which might be due to the lack of correlation (Figure 11c) between historical values of TMX and PPT for September. Figure 11 (d) provides box plots of autocorrelation which clearly indicate that the historical values are adequately reproduced by the model. It may be recalled that the correlation structure was adequately preserved for the scenario of increasing TAV as well.

It is interesting to note that the correlation structure is mostly preserved irrespective of the climate scenario considered. This is realistically consistent with the observed climate relationships, but may be viewed as an arguable drawback of the K-NN algorithm (Yates et al., 2003). For example, if a scenario of increasing TAV is considered, above average minimum temperatures may be accompanied either by an increase or a decrease in precipitation, which will then be a characteristic of an "increasing average temperature scenario". It appears that it is not possible to perform true partial derivative experiments that generate climate scenarios with say, higher average temperatures but hold the other variables constant. Recall that with the K-NN algorithm, a block of values rather a single value of the variable is resampled. Hence, it would be difficult to produce scenarios with correlation structure that is significantly different to the one observed in historical record.

1.6.4 Case 4: Extreme Precipitation Events Simulation

This section describes the simulation carried out to assess the performance of the model in reproducing the persistence character of the observed data. The output from weather generator developed here will be used as an input to hydrological models. The intent is to assess the vulnerability of the basin to floods and droughts on the basis of outputs obtained from hydrological models. Particular attention is therefore given to the simulation of extreme precipitation events that are responsible for floods and droughts in the basin. Prolonged precipitation events during winter season combined with heavy rainfall during summer are the most probable cause of flooding in the basin. The ability of the model to simulate the occurrence of extreme events, both high precipitation and low precipitation, was therefore investigated.

Figure 12 shows the box plots of total precipitation that occurred during the most extreme precipitation event in each year of the historical and the simulated record. It can be seen from the box plots that the median of the simulated data matches very closely the median of the historical data. The interannual variability in the simulated data is quite prominent with a highest total precipitation of about 240 mm compared to a corresponding value of around 200 mm in the historical record. Figure 13 shows the total number of days with zero precipitation during extreme event in each year of the historical and the simulated record. Again, the median of the simulated data matches well with the observed data although the simulation produced a slightly higher number of days with zero precipitation. However, the variety in the simulated sequences is quite evident. This clearly indicates that the model is capable of producing extreme events other than those observed in the historical record while preserving the historical mean. The reproduction of unprecedented extreme events by the model is crucial for the intended application in rainfall-runoff models. The results of analysis of extreme events further indicate that the tendency of wet and dry days to exhibit persistence is adequately represented by the model.

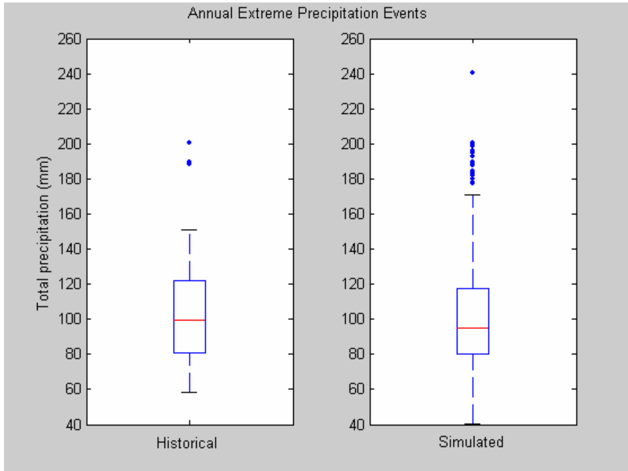


Figure 12 Box plots of total precipitation during extreme events in each year of the historical and simulated data

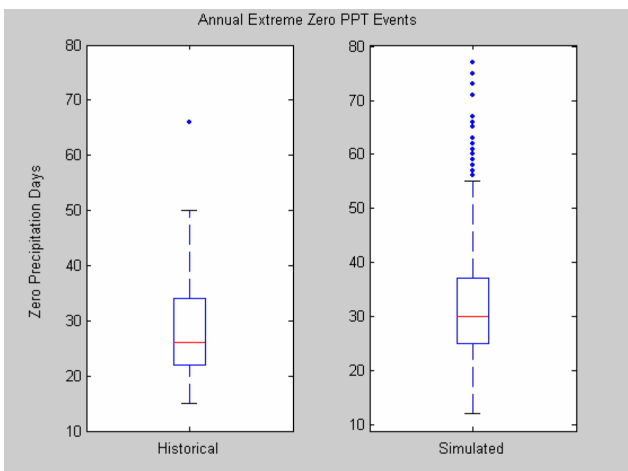


Figure 13 Box plots of zero precipitation days during extreme events in each year of the historical and simulated data

1.7 Summary and Conclusions

The development and application of a generic K-NN algorithm based weather generator in simulating potential climate change scenarios for the Upper Thames

River Basin in Ontario has been presented. The observed data set for different meteorological stations in the basin has been completed by estimating the missing values of the variables. In case of an incomplete data set, application of the K-NN algorithm can be quite problematic. Since development of weather generator is a part of a larger study to develop better flood and drought management practices in the basin under potential climate change scenarios, the completed data can now be used for other applications in the basin. The output from the weather generator developed here can be directly used as inputs to hydrological models.

Application of the weather generator to the data from Upper Thames River basin has clearly demonstrated the practicality of the approach in generating potential climate change scenarios for the basin. A major advantage of the approach is that non Gaussian features in the probability distribution of the variables are retained. As such, no prior assumptions regarding the probability distribution of the variables is required in the algorithm. Comparison of observed and synthetic clearly indicated that the model performance was very good with regards to reproduction to various statistics of interest to a hydrologist. Important properties of precipitation spell structure and amounts were preserved. Spatial and temporal dependencies were also well preserved which is the most distinguishing feature of the model presented here as most parametric techniques are unable to reproduce the correlation properties of the observed data series. The ability of the model to reproduce the correlation structure is particularly important for erosion, crop production and rainfall runoff models where the output of these models is greatly impacted by the right combination of meteorological variables.

Although the K-NN algorithm was designed to model daily statistics, the monthly statistics also appear to be reproduced adequately for the application presented here. Adequate reproduction of monthly statistics can be viewed as a challenging test of the statistical properties of the daily weather generator described here. A distinct practical advantage of the model developed here is that it is fairly generic and hence easily transportable to any other basin with very few modifications.

References

Bardosy, A. (1998). "Generating precipitation time series using simulated annealing" *Water Resour. Res.*, 34, 1737-1744.

Bardosy, A., and Plate, E. J. (1992). "Space-time model for daily rainfall using atmospheric circulation patterns." *Water Resour. Res.*, 28(5), 1247-1259.

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

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(11), 2761-2776.

Burn, D. H., and Simonovic, S. P. (1996). "Sensitivity of reservoir operation performance to climatic change." *Water Resour. Mgmt.*, 10, 463-478.

Burn, D. H. (1994). "Hydrological effects of climatic change In west-central Canada." *J. Hydrol.*, 160, 53-70.

Clair, T., Beltaos, S., Brimley, W., and Diamond, A. (1996). "Climate change sensitivities of Atlantic canada's Hydrological and Ecological Systems." *Proc., Climate Variability and Climatic Change in Atlantic Canada,* Nova Scotia, Canada.

Davis, J. (1986). *Statistics and Data Analysis in Geology*, John Wiley, Princeton, Hoboken, N. J.

Gleick, P. H. (1989). "Climate change, hydrology, and water resources." *Reviews of Geophysics*, 27(3), 329-344.

Gleick, P. H. (1987). "The development and testing of a water balance model for climate impact assessment: Modelling the Sacramento basin." *Water Resour. Res.*, 26, 1049-1061.

Gleick, P. H. (1986). "Methods for evaluating the regional hydrological impacts of global climatic changes." *J. Hydrol.*, 88, 97-116.

Gullet, D. W., and Skinner, W. R. (1992). The state of Canada's climate: Temperature changes in Canada 1895-1991. A state of the Environment report, SOE Rep No. 92-2, Atmos. Environ. Serv. Environment Canada, pp 3.

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

Hanson, C. L., and Johnson, G. L. (1998). GEM (Generation of weather Elements for Multiple applications): Its application in areas of complex terrain, *Hydrology Water Resources and Ecology in Headwaters,* K. Kovar, U. Tappeiner, N. E. Peters, and R. G. Craig (eds.), International Association of Hydrological Sciences Press, 27-32.

Intergovernmental Panel on Climate Change (IPCC). 1996. *Climate Change*: *The Science of Climate Change*. R. Watson, M. Zinyowera, and R. Moss (eds.), Cambridge University Press, UK.

Intergovernmental Panel on Climate Change (IPCC). 2001. *Impacts Adaptation and Vulnerability*: Report of Working Group 2, WMO and UNEP.

Jenkins, G. J. and Derwint, R. G. (1990) "Climate consequences of emissions." In *Climate Change*. The IPCC scientific assessment. J. T. Houghton, G. J. Jenkins and J. J. Ephraums (eds.) WMO, UNEP, Cambridge University Press, UK.

Katz, R. W. (1977). "Precipitation as a chain-dependent process." *J. Appl. Meteorol.*, 16, 671-676.

Katz, R., and Parlange, M. B. (1995). "Generalization of chain-depenedent processes: Applications to hourly precipitation." *Water Resour. Res.*, 31, 1331-1341.

Katz, R. W. (1996). "Use of conditional stochastic models to generate climate change scenarios." *Clim. Change*, 32, 237- 255.

Kite, G. W. (1993). "Application of a land class hydrological model to climatic change." *Water Resour. Res.*, 29, 2377-2384.

Lall, U., and Sharma, A. (1996). "A nearest neighbour bootstrap for time series resampling." *Water Resour. Res.*, 32(3), 679-693.

Lall, U., Rajagopalan, B., and Torboton, D. G. (1996). "A nonparametric wet/dry spell model for resampling daily precipitation." *Water Resour. Res.*, 32(9), 2803-2823.

Lattenmaier, D. P., and Gan, T. Y. (1990). "Hydrologic sensitivities of the Sacramento-San Joaquin River Basin, California, to global warming." *Water Resour. Res.*, 26, 69-86.

Lewis, J. E. (1989). "Climate change and its effects on water resources on Canada: A review." *Canadian Water Resources Journal*, 14(1), 34-55.

Mann, M. E., Bradley, R. S., and Hughes, M. K. (1999). "Northern hemisphere temperatures during the past millenium: Inferences, uncertainities, and limitations." *Geophysical Research Letters*, 26, 759-762.

Marsh, P., and Lesack, L. F. W. (1997). "Climate change and the hydrological regime of Lakes in the Mackenzie Delta." *Limnology and Oceanography*, 41, 849-856.

Mortsch, L., Hengeveld, H., Lister, M., Lofgren, B., Quinn, F., Slivitzky, M., and Wenger, L. (2000). "Climate change impacts on the hydrology of the Great Lakes – St. Lawrence System." *Canadian Water Resources Journal*, 25(2), 153-179.

Nicks, A. D., Richardson, C. W., and Williams, J. R. (1990). Evaluation of EPIC model weather generator: Erosion/Productivity Impact Calculator, 1. Model Documentation, A. N. Sharpley and J. R. Williams (eds.), USDA-ARS Tech. Bull. 1768, 235 pp.

Nicks, A. D., and Harp, J. F. (1980). "Stochastic generation of temperature and solar radiation data." *J. Hydrol.*, 48, 1-7.

Parlange, M. B., and Katz, R. W. (2000). "An extended version of the Richardson model for simulating daily weather variables." *J. Appl. Meteorol.*, 39, 610-622.

Rackso, P., Szeidi, L., and Semenov, M. (1991). "A serial approach to local stochastic weather models." *Ecol. Model.*, 57, 27-41.

Rajagopalan, B., and Lall, U. (1999). "A k-nearest neighbour simulator for daily precipitation and other variables." *Water Resour. Res.*, 35(10), 3089-3101.

Ravelle, R. R. and Waggoner, P. E. (1983). "Effects of a carbon dioxide-induced climatic change on water supplies in the western United States." in *Changing Climate*, Report of the Carbon Dioxide Assessment Committee, National Academic Press, Washington, D. C.

Richardson, C. W. (1981). "Stochastic simulation of daily precipitation, temperature and solar radiation." *Water Resour. Res.*, 17(1), 182-190.

Richardson, C.W., and D.A. Wright. 1984. "WGEN: a model for generating daily weather variables". U.S. Department of Agriculture, Agricultural Research Service, Washington, D.C., ARS-8, 88p.

Semenov, M. A., and Barrow, E. M. (1997). "Use of a stochastic weather generator in the development of climate change scenarios." *Climate Change*, 35, 397-414.

Semenov, M. A., Brooks, R. J., Barrow, E. M., and Richardson, C. W. (1998). "Comparison of WGEN and LARS-WG stochastic weather generators for diverse climates." *Climate Research*, 10, 95-107.

Simonovic, S. P. (2001). "Assessment of the impact of climate change and variability on the reliability, resiliency and vulnerability of complex flood protection systems." *Report to Natural Resources Canada.*, 91 pp.

Smith, R. E., and Schreiber, H. A. (1974). "Point processes of seasonal thunderstorm rainfall, 2, rainfall depth probablities." *Water Resour. Res.*, (10), 418-423.

Smith, R. L. (1994). "Spatial modelling of rainfall data." In *Statistics for the Environment 2: Water Related Issues*, V. Barnett and K. F. Turkman (eds.), John Wiley, NY, pp 19-41.

Southam, C. F., Mills, B. N., Moulton, R. J., and Brown, D. W. (1999). "The potential impact of climate change in Ontario's Grand River basin: Water supply and demand issues." *Canadian Water Resources Journal*, 24(4), 307-330.

Stern, R. D., and Coe, R. (1984). "A model fitting anlaysis of rainfall data." *Stat. Soc., Ser. A.*, 147, 1-34.

Todorovic, P. and Woolhiser, D. A. (1975). "A stochastic model of n-day precipitation." *J. Appl. Meteorol.*, 14, 17-24.

Wilby, R. L. (1994). "Stochastic weather type simulation for regional climate change impact." *Water Resour. Res.*, 30(12),3395-3403.

Wilks, D. S. (1999). "Interannual variability and extreme value characteristics of several stochastic daily precipitation models." *Agric. For Meteorol.,* 93, 153-169.

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

Wilks, D. S. (1992). "Adapting stochastic weather generation algorithms for climate change studies." *Climate Change*, 22, 67-84.

Wilks, D. S., and Wilby, R. L. (1999). "The weather generation game: a review of stochastic weather models." *Progress in Physical Geography*, 23, 329-357.

Westmacott, J. R., and Burn, D. H. (1997). "Climate change effects on the hydrological regime within the Churchill-Nelson river Basin." *J. Hydrol.*, 202, 263-279.

Woolhiser, D. A., and Roldan, J. (1982). "Stochastic daily precipitation models, 2, A comparison of distribution of amounts." *Water Resour. Res.*, (18), 1461-1468.

Yates, D., Gangopadhyay, S., Rajagopalan, B., Strzepek, K. (2003). "A technique for generating regional climate scenarios using a nearest-neighbor algorithm." *Water Resour. Res.*, 39(7), SWC 7-1 – 7-14.

Young, K. C. (1994). "A multivariate chain model for simulating climatic parameters with daily data." *J. Appl. Meteorol.*, 33, 661-671.