18th Canadian Hydrotechnical Conference
18ième congrès spécialisé hydrotechnique canadienne

Challenges for Water Resources Engineering in a Changing World
Winnipeg, Manitoba, August 22 – 24, 2007 / *22 – 24 août 2007*

Daily and Hourly Weather Data Generation using a K-Nearest Neighbour Approach

Mohammed Sharif[1], Donald H. Burn[2] and Karen M. Wey[2]
[1] Department of Civil Engineering, Jamia University, New Delhi, INDIA
[2] Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario

Abstract: This paper describes the development and application of a weather generator capable of generating meteorological inputs for hydrological models at both a daily and hourly time step. The weather generator is based on K-nearest neighbour resampling of the historical data. A strategy is introduced that resamples the historical data, with an added perturbation, while preserving the prominent statistical characteristics, including the inter-station correlations. An advantage of the approach is that unprecedented meteorological data are generated that are important for the simulation of extreme events. The basic weather generator operates at a daily time step; however daily values can be disaggregated to an hourly time step, for specific events of interest. The model can be used to generate realistic sequences of weather data reflecting historic conditions and can also be used to create an ensemble of climate scenarios that can be used for the assessment of the vulnerability of a watershed to extreme events, including floods and droughts. Climate scenarios are derived using change fields from GCM output. The weather generator model, in conjunction with a rainfall-runoff model, provides valuable aid in developing efficient management strategies for a watershed. The approach is demonstrated through application to the Upper Thames River Basin in Ontario. Daily and hourly weather variables (maximum and minimum temperature, and precipitation) were simulated at multiple stations in and around the basin. Analysis of the simulated data demonstrated the capability of the model to reproduce important statistical parameters of the observed data series while allowing perturbations to the observed data points.

1. Introduction

This paper describes the development and application of an improved K-nearest neighbour weather generating model that allows nearest neighbour resampling with perturbation of the historic data. The modelling approach involves a two-stage process of generating daily weather data followed by disaggregation to an hourly time step for select variables for some events. The practicality of the improved model is demonstrated through simulation of climate change scenarios based on the prediction of Global Climate Models (GCMs) for the Upper Thames River Basin (UTRB) in Ontario.

Stochastic weather generators are often used to simulate synthetic weather data that provide valuable aid in the formulation of water resource management policies. Synthetic sequences derived from stochastic generators are typically based on the assumption that the past will be representative of the future. However, the assumption of stationarity of meteorological data is not valid if the impacts of climate change are considered.

Due to the lack of an ideal method for generating climate change scenarios, weather generators have been employed for producing alternative climate data sets based on prescribed conditioning criteria (e.g., Semenov et al. 1998). Daily weather generators are most common due to the wide availability of meteorological data at a daily time step and due to the fact that most impact assessment models are driven by daily data. Results from a weather generator at a daily time step may need to be disaggregated to a finer temporal scale for some applications. The traditional weather generating approach (Nicks and Harp 1980) focuses first on the independent generation of precipitation while the remaining variables are modelled conditioned on 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 the cross-correlations. More elaborate models have been proposed for the distribution of precipitation amounts given the occurrence of a wet day. Buishand (1978) and Stern and Coe (1984) used the two-parameter gamma distribution to describe the precipitation amount on wet days. Wilks (1999) fitted the 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. The 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 stochastically generate daily values of maximum and minimum temperature, precipitation and solar radiation for any required length of time.

A drawback associated with the 'Richardson type' weather generators is that persistent events, such as drought or prolonged rainfall, are not well reproduced. To overcome this problem, a serial approach to weather generation was presented by Racsko 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 status. Racsko 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 observation in the modelling process, it is expected to perform better than models that are based on fitting of a predefined distribution to the observed data. Semenov et al. (1998) confirmed the superiority of LARS-WG through a performance evaluation of WGEN and LARS-WG at 18 sites from different parts of the world. LARS-WG matched the observed data more closely than did 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. Smith (1994) presented an extension of the Markov model to bivariate time series of daily precipitation at two stations. Further extension to more stations was limited by the number of parameters needed 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 the first order Markov model with the mixed exponential distribution for wet day precipitation amounts. Means, variances, and inter-station 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, they do not adequately reproduce various aspects of spatial and temporal dependence of variables. 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 (MAR) models implicitly assume a normal distribution. 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 problems associated with the parametric methods. Simple nonparametric techniques essentially involve random resampling from the historical data to generate synthetic sequences of the required duration. Such sequences often fail to capture the time correlation of the data series. Procedures have been developed that can capture the prominent time correlation between the weather data. The most promising nonparametric technique for generating weather data is the K-nearest neighbour (K-NN) resampling approach. The works of Lall et al. (1996), Rajagopalan and Lall (1999) and Buishand and Brandsma (2001) describe various forms of K-NN resampling. 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.

A nonparametric weather generator can be used to create synthetic meteorological data for given climate scenarios and operates on a daily time step. The output from this model functions as input for hydrological models used to simulate streamflow. While the daily weather generator data can be input directly to the continuous hydrological model used in this project, temporal disaggregation and spatial interpolation are required before these data can be used as input to the event-based hydrological model (Cunderlik and Simonovic 2007). The approaches used for disaggregation and interpolation of the daily precipitation data are also discussed in this paper.

2. Nonparametric Weather Generator

A basic K-NN algorithm involves selecting a specified number of days similar in characteristics to the day of interest. One of these days is randomly resampled to generate the weather of the next day in the simulation period. The nearest neighbour approach involves simultaneous sampling of weather variables, such as precipitation and temperature. The sampling is carried out from the observed data, with replacement. To simulate weather variables for a new day (t+1), days with similar characteristics as those simulated for day t are selected from the historical record. One of these nearest neighbours is then selected according to a defined probability distribution or kernel and the observed values for the day subsequent to that nearest neighbour are adopted as the simulated values for day t+1. Models based on the K-NN approach can be easily extended to multisite simulation of weather data while keeping the spatial correlation structure virtually intact. Spatial dependencies are preserved because the same day's weather is adopted as the weather for all stations. Temporal dependence is also likely to be preserved as the simulated values for day t+1 are conditioned on the values for day t. The cross-correlation among the variables at any given site is preserved since a block of variables is resampled from the observed data.

A limitation of K-NN weather generators is that they do not produce new values but merely reshuffle the historical data to generate realistic weather sequences. A modified approach was developed that entails resampling with perturbation of the historic data. A strategy is introduced that resamples the historical data with perturbations while preserving the prominent statistical characteristics, including the inter-station correlations. The improved model is capable of extrapolating beyond the observed record to produce precipitation and temperature values that are different from the observed values. Details of this daily weather generation model are presented in Sharif and Burn (2006). A practical advantage of the approach is that unprecedented precipitation amounts are generated that are important for the prediction of hydrologic extremes. The disaggregation of the daily data to an hourly time step further facilitates the use of the weather data for event based modelling, which typically operates at a small time interval.

Precipitation data are often required at a finer time scale than is observed, such as hourly rather than daily. One option for creating hourly precipitation data is to modify the weather generator to produce hourly data, while another is to disaggregate the daily data produced by the weather generator to hourly data. Porter and Pink (1991) recognized that generation of monthly data rarely preserves the annual statistics of the time series and concluded that disaggregation was a good way to maintain the necessary statistical properties of the data, at both the annual and monthly scale. Wójcik and Buishand (2003) extended this theory to the disaggregation of daily climate data and concluded that disaggregation of daily data preserved second order statistics much better than directly generating data at the finer time scale.

For precipitation events requiring hourly data, a second modelling stage is used herein to disaggregate the daily precipitation values obtained for select precipitation events generated using the daily weather generation model. Since the hourly data are intended to estimate potential flood events, only data for events with a substantial amount of precipitation were disaggregated, where an event is defined as all consecutive non-zero precipitation days, with at least one day with zero precipitation separating events. The approach used is based on the method of fragments (Svanidze 1977). The method of fragments requires data at both the original and the disaggregated time scale. For a daily data value that is to be disaggregated, a set of fragments must be formed. The fragments are the fraction of daily precipitation that occurred in each hour of the day, thus the fragments sum to unity. Each fragment is then multiplied by the daily data being disaggregated such that the total precipitation in the day is not altered. This produces the new hourly precipitation values.

The series of hourly data to be used as fragments was selected such that the total daily precipitation closely matches the daily data being disaggregated. The method used is a hybrid between the key site approach (McMahon and Mein 1986; Porter and Pink 1991) and the method used by Maheepala and Perera (1996). The approach involves multiple key sites, thereby representing the distribution across the watershed without the computational expense of examining all sites in the study area. The multiple key site method requires the user to select the number of key sites to be used for a specific application. By increasing the number of key sites, the user decreases the ambiguity of the selection process; however, by decreasing the number of keys sites, the user decreases the required computational time for fragment selection. It is important to select enough sites to accurately represent the region of study.

For the disaggregation of the first day of any event there are no hourly data in the previous day for comparison because, by definition, the day prior to the beginning of an event must be a non-event day. There is no precedence put on the time of day when the precipitation occurs during the fragment selection process for the first day of an event. By selecting the fragments for the first day of an event in this manner, no assumptions were made regarding the precipitation on the previous day. The fragments to be used for disaggregation of a generated daily precipitation, values are selected based on a nearest-neighbour day for the generated daily precipitation. While the first event day is not influenced by preceding precipitation, all subsequent event days must maintain some continuity with the former portion of the event. All subsequent event days had fragments chosen by selecting the day that resulted in minimum deviation from current day precipitation and the hourly precipitation on the previous day. Precipitation patterns vary depending on the season. To maintain an appropriate seasonal distinction in the disaggregated data, it is vital to choose fragments, and hence precipitation patterns, that originate in the correct season. For this model a season was defined as 30 days in either direction of the day being disaggregated and nearest neighbours for selecting fragments were restricted to days within this window.

3. Application

The Thames River Basin is located in the agricultural heartland of the south-western region of Ontario. The Thames River is 273 km long and has a catchment area of around 5,825 km^2, making it the second largest watershed in south-western Ontario. Most of the precipitation comes in the form of winter snow. Daily maximum temperature (TMX), daily minimum temperature (TMN) and daily precipitation depth (PPT) data from 15 stations in and around the basin were used for the period from 1964 to 2001. The meteorological stations in and around the basin are distributed across an area of approximate dimension 80 km (west to east) by 120 km (north to south). The inter-station distances range from approximately 10 km to 120 km. In addition to the daily data, hourly precipitation data are available for some of the sites plus additional sites that are not part of the daily data network (for a total of 28 sites with hourly data). As a result of poor temporal coverage for many of the hourly data sites and the need for co-located hourly and daily data sites, spatial interpolation with the inverse distance method was used to create a consistent network of 15 sites with both daily and hourly data.

3.1 Reproduction of Historical Statistics

The model was initially evaluated for capability to reproduce the characteristics of historical data. A new subset of years that constitute the driving data for the model was obtained by using an integer function that returned integers between specified upper and lower bounds. To generate N years of data, the integer function was called N times. With this method, each year has an equal probability of being selected but some years may be selected more than once. A new data set was thus obtained and the K-NN algorithm was used to generate 800 years of synthetic data with this data set. The goal of simulation was to produce a data series that preserved the statistical attributes of the historic data while perturbing the data points. The statistics of interest were computed from the simulated sequence and compared to the statistics of the observed record using box plots. Results are presented below for London; the results for other stations are similar.

Figure 1 shows the box plots of 800 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 a monthly time scale to facilitate presentation of the results. The simulations are summarized using box plots. Box plots show the range of variation in statistics of simulations and provide a straightforward method of comparing the statistics of simulations with historical 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. The horizontal line within the box represents the median. 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 inter-quartile range of the data. The values beyond the ends of the whiskers are called outliers and are shown by dots. The statistics of the historical record are represented by dots and joined by solid lines. Comparison of historical monthly values with the simulated values clearly showed that the model was capable of adequately reproducing the historical values. This is highly satisfactory given that monthly statistics were not explicitly specified in fitting the K-NN model.

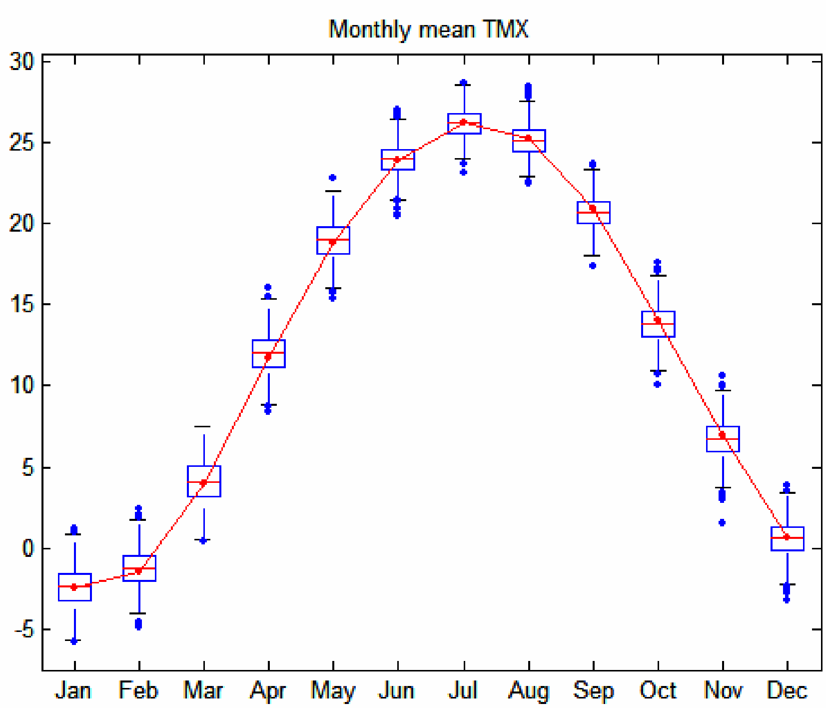

Figure 1. Box plots of monthly mean maximum temperature for London

Evaluation of total monthly precipitation revealed that the historical mean of the total precipitation is close to the median of the simulated data for all months. The simulated total annual precipitation (987 mm) closely matched the historical value (980 mm). The model slightly overestimated the monthly totals for February, August, September and November. For the rest of the months, model results were close to the observed values. Among all weather variables, precipitation has the greatest variability in time and space and therefore the performance of the model in simulating the total monthly precipitation may be considered to be very good. The total number of wet days was also evaluated with results indicating that the model adequately reproduced the total number of wet days in different months of the year.

Parametric models often fail to reproduce the correlation structure of the observed data. Due to the inherent structure of the basic K-NN model, there is a strong likelihood of the correlation structure being preserved. Many studies have shown that the lack of spatially distributed precipitation amounts can have a serious impact on basin runoff generation (Shah et al. 1996). The assumption of uniform spatial precipitation distribution is often invalid, even for small basins, because runoff simulation is significantly impacted by the distribution of precipitation over the basin. Therefore, it was considered important to evaluate the performance of the model in reproducing the spatial dependencies of the observed data, especially since the proposed approach involves perturbing the observed data points. Scatter plots of inter-station correlations for daily precipitation values are presented in Figure 2, which shows correlation coefficients for daily PPT values for the observed and the simulated data. The model reproduced the historical correlation structure very well. The spatial correlation of the disaggregated historical generated data matched the spatial correlation of spatially interpolated data closely across the entire watershed. It was also found that spatial variation was maintained within extreme events as well. Isohyetal plots revealed that plausible spatial variation of hourly data is obtained within extreme events.

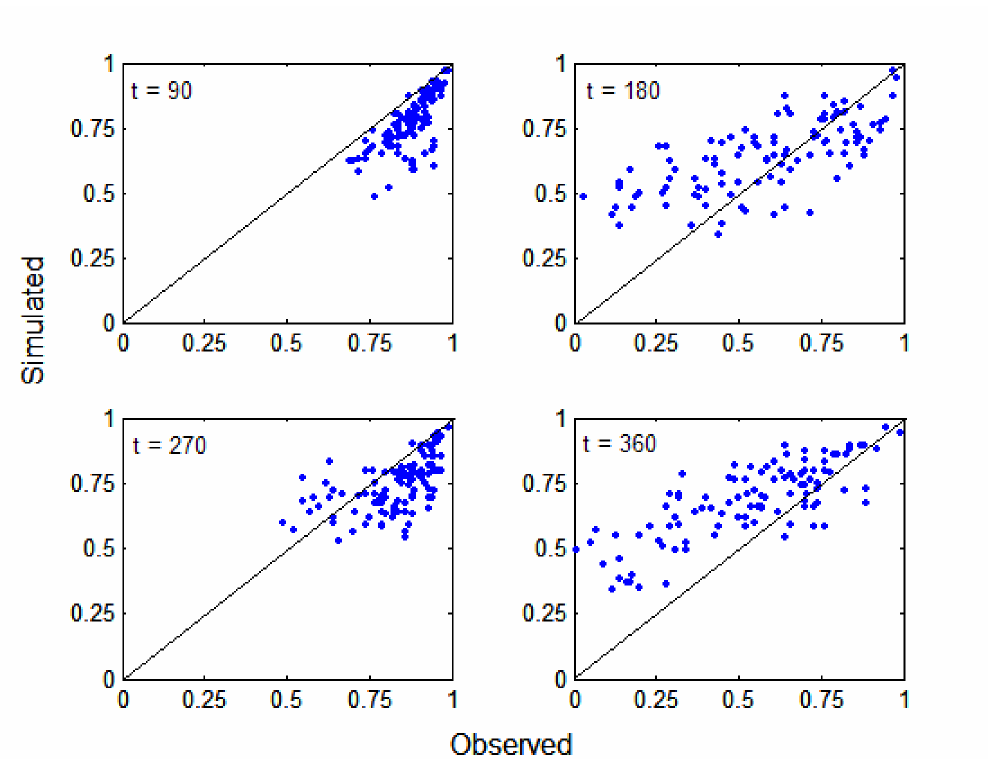

Figure 2. Observed versus simulated correlations for daily PPT values for four days

3.2 Climate Change Scenarios

For the present study, output from two GCMs (CCSRNIES and CSIRO Mk2b) for scenarios B11 and B21, respectively, have been considered (see Nakicenovic and Swart (2000) for a description of the scenarios). The time-series of three climate variables: TMX, TMN, and PPT were obtained for a 30 year time slice (2040-2069). The data obtained may be taken as representing the average climatic conditions in the year 2050. The baseline simulations were carried out for the period 1961-1990 and the change fields corresponding to the year 2050 for both scenarios were obtained for TMX, TMN, and PPT. These changes have been expressed on a monthly time scale. All the observed daily temperatures of a given month have been increased/decreased by the amount predicted by the GCM for that particular month. Similarly, observed daily precipitation values of each month have been modified by multiplying them with the fractional change predicted by the GCM for that particular month. A modified data set is, therefore, obtained after applying the GCM-predicted change fields to the observed data set. Both models predict increases in TMN for all months although the magnitude of change is different for each model. For example, there is a difference of about 4° C between the change in TMN predicted by the two GCMs for the month of January. For the month of March and July, the predictions match very well. There are, however, substantial differences between the predicted values for the rest of the months with maximum deviations observed for the month of March. The variation between the change predicted by CSIRO Mk2b B11 and CCSRNIES B21 scenarios is less for TMX as compared to TMN. The difference between the predicted values of PPT for the two scenarios is large when compared to the change fields for TMX and TMN. The deviation in predictions is to the extent that one scenario predicts an increase while the other predicts a decrease for the same month. The PPT change fields deviate substantially from each other except for the months of February, May and July. For April and September, CSIRO Mk2b B11 predicts a decrease in PPT while CCSRNIES B21 predicts an increase. For October and December the change predicted by CSIRO Mk2b B11 is negative in comparison to a positive change predicted by scenario CCSRNIES B21. It is only for the month of November that both models predict a decrease in the PPT values. The two GCM scenarios produce very different projections of future climate. These scenarios should be interpreted as plausible scenarios rather than predictions of future climatic conditions.

The KNN-model was then applied to simulate weather data conditioned upon plausible climate change scenarios derived from the two GCMs. Box plots have been used to present the statistics of interest obtained from the simulated sequences. Results are again presented for the London station. Two 800-year simulations were carried out, each corresponding to the GCM scenarios considered. Simulation 1 was carried out to reproduce the statistics of the modified data set obtained after applying the change fields from CSIRO Mk2b B11 to the observed data. Simulation 2 represents the results obtained for CCSRNIES B21. The results of the simulations are described below.

Relatively modest changes in precipitation can have proportionally large impacts on runoff, which directly affects the occurrence of floods. Under scenarios of increasing temperatures, it is reasonable to expect reductions in spring and summer runoff, increases in winter runoff and earlier peak runoff. It is therefore important to evaluate the impacts of potential change in precipitation amounts on the occurrence of extreme events. The analysis of simulated data representative of CSIRO Mk2b B11 and CCSRNIES B21 was, therefore, conducted. The capability of the model to simulate the occurrence of extreme events, both high precipitation and low precipitation, was investigated with particular emphasis on generating realistic unprecedented events for the basin. Figure 3 shows the box plots of total precipitation that occurred during the most extreme precipitation event in each year of the simulated record for simulations under CSIRO Mk2b B11 and CCSRNIES B21. An extreme precipitation event is defined as consisting of a continuous sequence of wet days. For each year of the data, a single multi-day extreme event was determined along with the total precipitation that occurred during the event. The first box plot in Figure 3 shows the total precipitation during the most extreme event in each year of the observed data while the remaining box plots show the same statistics for two different simulations.

The box plot for CSIRO Mk2b B11 shows that the median of the simulated data matches closely with the median of the observed data. A total precipitation of the order of 250 mm in the most extreme precipitation event was observed in the simulated data compared to a corresponding value of 200 mm in the observed record. Additionally, greater variability in the simulation of extreme events is achieved. A total precipitation amount of around 225 mm was simulated in the most extreme event for CCSRNIES B21, which is about 25 mm higher than the highest observed value. The median value is slightly higher than the observed value as well as that simulated for scenario CSIRO Mk2b B11. Of the two scenarios, CSIRO Mk2b B11 appears to be critical with respect to simulation of extreme high precipitation events. The inter-annual variability for both scenarios is quite high as is the case with all other simulations thus providing a variety of extreme events that can be used as input to hydrological models.

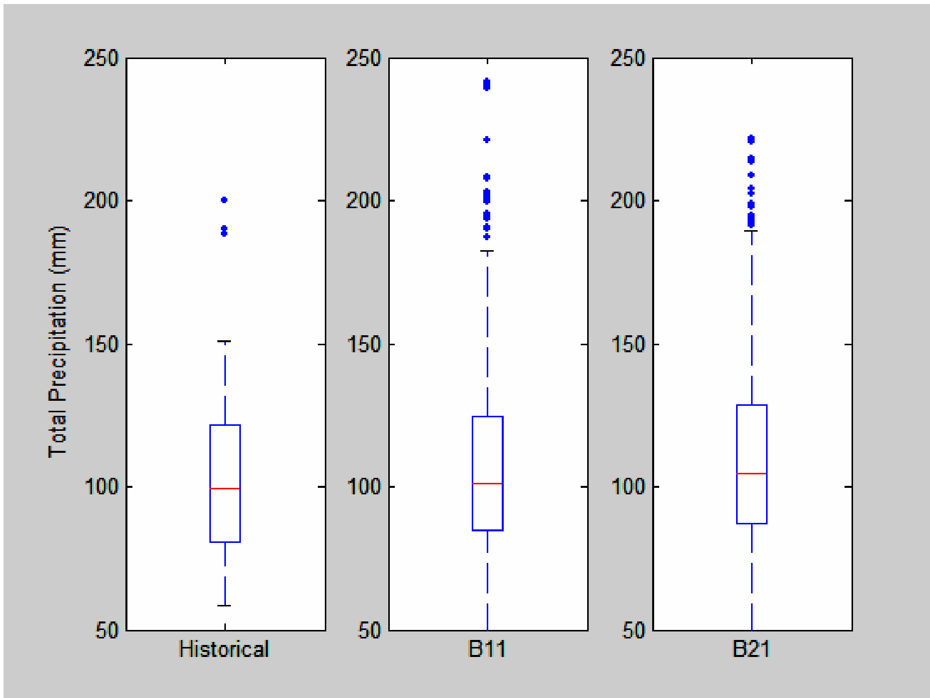

Figure 3. Box plots of total precipitation during extreme events in each year for different simulations

Apart from floods, the vulnerability of UTRB to the occurrence of droughts needs to be assessed. Figure 4 shows box plots of dry spell durations for the observed and the simulated data for both scenarios. The first box plot in Figure 4 presents the distribution of extreme dry spells computed from 38 years of observed data. It can be seen from the box plot of simulation 1 that while the median dry spell duration for the observed data is 25, the corresponding statistic in the simulated data for CSIRO Mk2b B11 is 29. The median of the simulated data for both scenarios was substantially higher than that for the observed data. The longest dry spell for CSIRO Mk2b B11 lasted around 90 days, which is substantially higher in comparison to the observed value of 67 days. For CCSRNIES B21, the most extreme dry spell duration has about the same value as for the observed data. Therefore, CSIRO Mk2b B11 is critical with respect to dry spell durations. Several events that are more severe than the observed events were simulated thus providing a wider range of events as input to a hydrologic model. The strength of the approach presented here lies in the simulation of extreme dry spells for potential climate change scenarios that are important for evaluation of effective drought management policies for the basin.

4. Conclusions

The potential impact of climatic change, as predicted by two GCMs, on the occurrence of extreme precipitation events in the UTRB has been investigated. Weather sequences based on two plausible climate change scenarios for the basin were simulated using a K-NN model. It should be noted that the

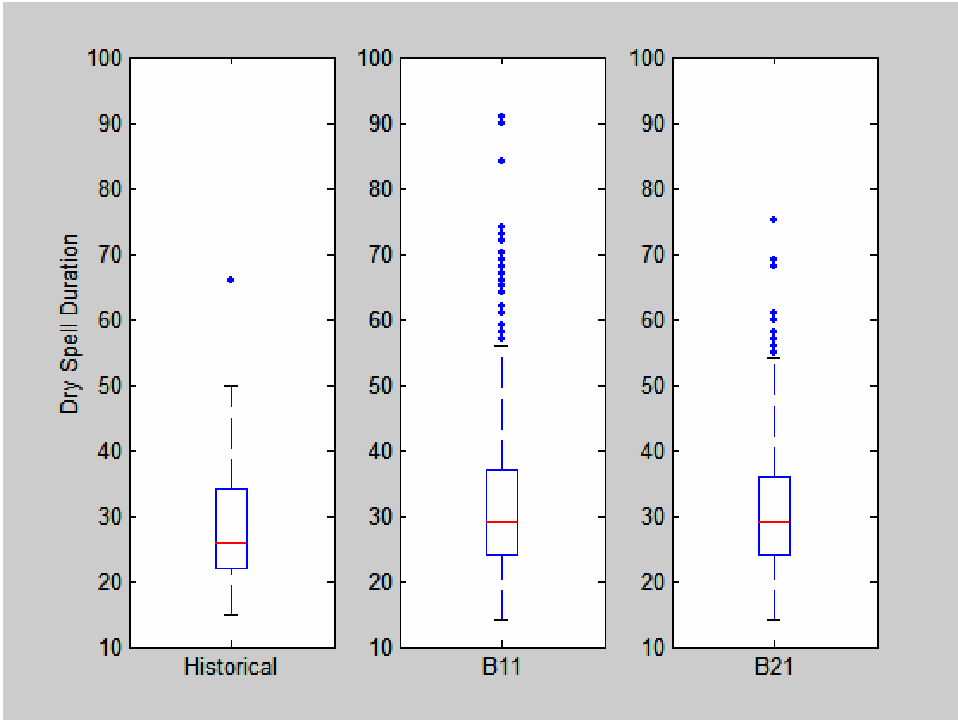

Figure 4. Box plots of dry spell durations during extreme events in each year for different simulations

intent herein was not to conduct an extensive climate change sensitivity analysis for the UTRB. Doing so would have required the consideration of a larger number of GCMs and scenarios. Rather the intent was to assess the effectiveness of the weather generation scheme for application to this watershed. The model framework has been shown to be effective in producing simulated sequences that are representative of future climate scenarios predicted by the GCMs for the UTRB. An advantage of the approach presented here is that unprecedented precipitation amounts can be simulated for a given climate change scenario. Consequently, extreme events, both high precipitation and low precipitation, that are more severe than the observed events can be easily simulated, which is important for assessing the vulnerability of the basin to floods and droughts under changing climatic conditions. A distinct advantage of the model is that it produces spatially correlated data for the basin, which is crucial for evaluating the response of hydrological models to watershed-level processes.

There are, however, limitations associated with the methodology adopted in this work. First, validation exercises have revealed that GCMs suffer from several deficiencies in the simulation of present day climate conditions. Therefore, the accuracy of the results presented is directly related to the accuracy of the GCM simulations. Second, a climate change scenario based on the output from a GCM represents only one of many future climate change scenarios whereas exploring many alternative climate scenarios could be more useful for effective management of water resource systems. Future work should explore the impacts of considering a larger number of GCMs and scenarios.

5. Acknowledgements

The research presented in this paper was funded by the Canadian Foundation for Climatic and Atmospheric Sciences (CFCAS). The authors gratefully acknowledge the cooperation received from Prof. S.P. Simonovic and Dr. J.M. Cunderlik, University of Western Ontario. The authors also thank Mr. M. Helsten for providing meteorological data for the Upper Thames River Basin.

6. References

Buishand, T.A. 1978. Some remarks on the use of daily rainfall models. *Journal of Hydrology*, 36(3-4): 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 Resources Research,* 37(11): 2761-2776.

Cunderlik, J. and Simonovic, S.P. 2007. Hydrologic models for inverse flood risk modeling. 18th Canadian Hydrotechnical Conference, CSCE, Winnipeg, MB, Canada.

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

Maheepala, S. and Perera, B.J.C. 1996. Monthly hydrologic data generation by disaggregation. *Journal of Hydrology*, 178: 277-291.

McMahon, T.A. and Mein, R.G. 1986. *River and Reservoir Yield*, Water Resources Publications, Littleton, Colorado.

Nakicenovic, N. and Swart, R. (eds.) 2000. *Special Report on Emissions Scenarios*, Cambridge University Press, Cambridge, UK, 612 p.

Nicks, A.D. and Harp, J.F. 1980. Stochastic generation of temperature and solar radiation data. *Journal of Hydrology*, 48: 1-7.

Porter, J.W. and Pink, B.J. 1991. A method of synthetic fragments for disaggregation in stochastic data generation. *International Hydrology and Water Resources Symposium*, Perth, Australia: 781-786.

Racsko, P., Szeidi, L. and Semenov, M. 1991. A serial approach to local stochastic weather models. *Ecological Modelling,* 57: 27-41.

Rajagopalan, B. and Lall, U. 1999. A K-nearest neighbor simulator for daily precipitation and other variables. *Water Resources Research*, 35(10): 3089-3101.

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

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

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.

Shah, S.M.S., O'Connell, P.E. and Hosking, J.R.M. 1996. Modelling the effect of spatial variability in rainfall on catchment response. 2: Experiments with distributed and lumped models. *Journal of Hydrology,* 175: 89-111.

Sharif, M. and Burn, D.H. 2006. Simulating climate change scenarios using an improved K-nearest neighbor model, *Journal of Hydrology*, 325: 179-196.

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: 19-41.

Stern, R.D. and Coe, R. 1984. A model fitting analysis of rainfall data. *Journal of the Royal Statistical Society Series A-Statistics in Society* 147: 1-34.

Svanidze, G.G. 1977. *Mathematical Modeling of Hydrologic Series*, T. Guerchon, translator, Water Resources Publications, Littleton, Colorado.

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

Wilks, D.S. 1998. Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1-4): 178-191.

Wilks, D.S. 1999. Multisite downscaling of daily precipitation with a stochastic weather generator. *Climate Research*, 11: 125-136.

Wilks, D.S., and Wilby, R.L. 1999. The weather generation game: A review of stochastic weather models. *Progress in Physical Geography*, 23(3): 329-357.

Wójcik, J. and Buishand, T.A. 2003. Simulation of 6-hourly rainfall and temperature by two resampling schemes. *Journal of Hydrology*, 273: 69-80.