JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 51, No. 5 AMERICAN WATER RESOURCES ASSOCIATION October 2015

IMPROVED WEATHER GENERATOR ALGORITHM FOR MULTISITE SIMULATION OF PRECIPITATION AND TEMPERATURE[1]

Leanna M. King, A. Ian McLeod, and Slobodan P. Simonovic[2]

ABSTRACT: The KnnCAD Version 4 weather generator algorithm for nonparametric, multisite simulations of temperature and precipitation data is presented. The K-nearest neighbor weather generator essentially reshuffles the historical data, with replacement. In KnnCAD Version 4, a block resampling scheme is introduced to preserve the temporal correlation structure in temperature data. Perturbation of the reshuffled variable data is also added to enhance the generation of extreme values. The Upper Thames River Basin in Ontario, Canada is used as a case study and the model is shown to simulate effectively the historical characteristics at the site. The KnnCAD Version 4 approach is shown to improve on the previous versions of the model and offers a major advantage over many parametric and semiparametric weather generators in that multisite use can be easily achieved without making statistical assumptions dealing with the spatial correlations and probability distributions of each variable.

(KEY TERMS: stochastic; simulation; weather generator; climate.)

King, Leanna M., A. Ian McLeod, and Slobodan P. Simonovic, 2015. Improved Weather Generator Algorithm for Multisite Simulation of Precipitation and Temperature. *Journal of the American Water Resources Association* (JAWRA) 51(5): 1305-1320. DOI: 10.1111/1752-1688.12307

INTRODUCTION

Weather generators are stochastic simulation tools that are commonly used to produce synthetic climate data of any length with the same characteristics as the input record. Such algorithms are often used for hydrological applications (Dibike and Coulibaly, 2005; Charles *et al.*, 2007; Kwon *et al.*, 2011). In recent years, their application has been extended for statistical downscaling of atmosphere-ocean coupled global circulation model (AOGCM) outputs to investigate the impacts of climate change at a basin scale (Dibike and Coulibaly, 2005; Eum and Simonovic, 2012; Hashmi *et al.*, 2011).

There are several categories of stochastic weather generators. Parametric models typically follow the WGEN approach (Richardson, 1981; Soltani and Hoogenboom, 2003; Kuchar, 2004; Craigmile and Guttorp, 2011), which uses Markov chains to simulate wet and dry spells and probability distributions for temperature and precipitation amounts. Some other parametric weather generator examples include SIMMETEO (Geng *et al.*, 1988; Soltani and Hoogenboom, 2003; Elshamy *et al.*, 2006), WGENK (Kuchar, 2004), AAFC-WG (Qian *et al.*, 2004, 2008), and GEM (Hanson and Johnson, 1998). A drawback of the parametric models is that they require careful statistical checks to ensure the developed probability distributions are suitable to the study area (Sharif and Burn,

[1]Paper No. JAWRA-13-0264-P of the *Journal of the American Water Resources Association* (JAWRA). Received December 11, 2013; accepted March 6, 2015. **Discussions are open until six months from issue publication**.

[2]Ph.D. Candidate (King) and Professor (Simonovic), Department of Civil and Environmental Engineering, and Professor (McLeod), Department of Statistical and Actuarial Sciences, The University of Western Ontario, London, Ontario, Canada N6A3K7; Contact author mailing address: 448 Grangeover Crescent, London Ontario N6G 4P7 (E-Mail/King: leannamichelleking@gmail.com).

2006). Furthermore, when the models are used with low-order Markov dependence, they cannot adequately simulate wet and dry spell lengths, underestimating the occurrence of prolonged drought or rainfall events (Semenov and Barrow, 1997; Dibike and Coulibaly, 2005; Mehrotra and Sharma, 2007; Sharif and Burn, 2007). It is difficult to preserve multisite correlations across all variables of interest in a parametric weather generator.

Several semiparametric models have been developed to overcome some of the limitations of the parametric models. Semiparametric weather generators are generally comprised of both parametric and nonparametric components (Apipattanavis *et al.*, 2007). Two of the commonly used models are SDSM (Wilby and Dawson, 2007) and LARS-WG (Semenov and Barrow, 2002). A drawback of these models is that they may only be used for one site at a time, with a requirement for additional modeling approaches to utilize simulation results in a multisite application (Wilby *et al.*, 2003).

Due to the limitations associated with parametric and semiparametric weather generators, nonparametric models have become increasingly popular. Nonparametric weather generators can create multisite, multivariate climate simulations without making assumptions regarding the inter-site spatial correlations or the probability distributions of the variables (Sharif and Burn, 2006, 2007). Nonparametric models typically use a K-Nearest Neighbor (K-NN) procedure which resamples from the historical record, with replacement (Yates *et al.*, 2003; Sharif and Burn, 2006; Eum *et al.*, 2010). The nearest neighbors to the current day are selected by calculating a distance metric between the current day and each of the days within a temporal window centered on that day from the entire N years of record. The closest K days are retained and one is randomly chosen as the next day's weather, with a higher probability given to closer days (Yates *et al.*, 2003; Sharif and Burn, 2006, 2007).

The K-NN models are capable of simulating several climate variables at a time by employing unequal weights for the contributing variables in the distance calculation, as suggested by Karlsson and Yakowitz (1987). A regional average is typically used to minimize the dimensions when choosing the nearest neighbor (Eum and Simonovic, 2012). The use of a regional average presents a potential drawback as the average daily climate condition in a very large study area may not be representative of climatic conditions at the more remote stations. Users of K-NN weather generators should ensure that the spatial correlations between climate stations are sufficiently high so that a regional average is representative of conditions throughout the basin.

Traditional K-NN approaches have been found to underestimate the occurrence of wet and dry spells (Apipattanavis *et al.*, 2007). Another drawback is that the reshuffling procedure results in a loss of the temporal correlation structure of daily climate variables. The extreme values in the output data are also limited by those in the input dataset since the values are resampled (Yates *et al.*, 2003). Sharif and Burn (2006, 2007) developed a perturbation method to generate alternative extreme values for precipitation in KnnCAD Version 1, however Eum and Simonovic (2012) found that their methodology could not be easily extended for generation of alternative temperature extremes because it would be capable of generating unrealistically high or low temperature values. Prodanovic and Simonovic (2008) improved on KnnCAD Version 1 by adding a leap year modification in KnnCAD Version 2. Eum and Simonovic (2012) improved on KnnCAD Version 2 by using principal components analysis to allow for inclusion of multiple climate variables without increasing computational effort. Lee *et al.* (2012), used a gamma kernel density estimate to successfully perturb the reshuffled precipitation data; however, their approach could not easily be used for temperature outputs.

In this study, the KnnCAD Version 4 algorithm is presented to address the issues associated with the previous versions of the algorithm: (1) loss of temporal correlation in simulated temperatures and (2) simulation of alternative extreme variable values. A block resampling procedure is added to improve temporal correlations of temperature variables. A perturbation procedure is introduced to enhance the generation of extreme temperature and precipitation values. The model outputs for the CLIMDEX set of statistics (CLIMDEX, 2012) are presented in a comparison with the previous version of KnnCAD (Version 3). The following section presents the details of the KnnCAD Version 4 weather generator algorithm, with a description of the improvements made over previous versions of the algorithm. The weather generators KnnCAD V3 and V4 are applied to simulate precipitation and temperature of the upper Thames River basin (UTRB) in Ontario, Canada. Finally conclusions of the article are presented.

METHODS

The KnnCAD Version 4 weather generator is an extension of the model of Yates *et al.* (2003). The model reshuffles the observed daily data, so application to multiple sites is achieved by selecting the corresponding day's weather at all stations. In this way, the spatial correlations of the climate variables are inherently preserved.

The model proceeds in steps by creating a subset of days from each year in the historical record that are centered within a temporal window on the current day of the simulation. The current day is removed to prevent repeated daily values. This subset of "potential neighbours" has length $L = N \times (w + 1) - 1$ for N years of record and a temporal window of length w. The regional average from all stations is computed for each variable and day in the potential neighbors. These potential neighbor averages are then compared to the current day's regional average using a distance metric, the Mahalanobis distance (Yates *et al.*, 2003; Sharif and Burn, 2006). Based on their distance from the current day, the potential neighbors are ranked and the first K are selected, the "K-NNs." Based on the days' ranks, a cumulative probability distribution is developed. The next day's weather is then selected by generating a random number u(0,1) and comparing this to the probability distribution, selecting the closest day. As such, days which are more similar to the current day have a greater probability of selection. In order to improve on KnnCAD V3 (Eum and Simonovic, 2012), KnnCAD V4 resamples a block of B days following the chosen neighbor in the historical record in order to preserve the temporal correlation structures of variables such as temperature. Each of the resampled values is then perturbed to ensure unique values are generated that do not necessarily occur in the historical record. The simulation proceeds until a dataset of reshuffled values is generated, with the same length as the input dataset. The simulation can be repeated several times to generate ensembles of synthetic daily data. AOGCM simulations can be generated by applying monthly change factors to the input dataset.

KnnCAD V4 presents two major advantages over KnnCAD V3: (1) an improvement in the temporal correlation structure of simulated temperatures and simulation of wet and dry spell sequences through the use of block resampling and (2) an improved perturbation method that is also applicable for reshuffled temperature values. The perturbation methodology of KnnCAD V4 differs from V3 in that it uses the standard deviation from the subset of nearest neighbors along with a random variable to perturb the daily climate variable. The user selects the amount of perturbation desired to ensure unique values are generated but daily autocorrelations are maintained. The approach can be applied to both temperature and precipitation values and generates reasonable extreme values.

Detailed steps of KnnCAD V4 are presented below. Full details of KnnCAD Version 3 can be found in Eum *et al.* (2009).

1. Compute the regional means of p variables (x) across all q stations for each day (t) in the historic record of length T, following Equations (1) and (2):

$$\bar{\boldsymbol{X}}_t = [\bar{x}_{1,t}, \bar{x}_{2,t}, \ldots, \bar{x}_{p,t}] \qquad \forall t = \{1, 2, \ldots, T\} \tag{1}$$

where

$$\bar{x}_{i,t} = \frac{1}{q}\sum_{j=1}^{q} x_{i,t}^{j} \qquad \forall i = \{1, 2, \ldots, p\} \tag{2}$$

The variables that are typically used in the KnnCAD V4 approach are precipitation, maximum temperature, and minimum temperature. Other variables may also be considered in some applications.

2. Choose a temporal window of length w, and select a subset of L potential neighbors to the current day of simulation, where $L = N \times (w + 1) - 1$ for N years of record. The potential neighbors are the days within the temporal window centered on the current day of the simulation, t, and contain p variables for a total of L days. Yates *et al.* (2003) used a temporal window of 14 days in the Great Lakes region, so if January 20 is the current day, the potential neighbors are all days that fall between January 13 and January 27 for all N years, excluding the value of the current day to prevent repeating weather sequences.
3. Randomly choose p variables at q stations from one of the N years of historical record for the first time step day (e.g., January 1).
4. Compute the regional means $\bar{\boldsymbol{X}}_l$, of the L potential neighbors (l = 1, 2, …, L) for each day across all q stations.
5. Compute the covariance matrix, C_t, for day t using the potential neighbors from (3) with a standardized data block of size L by p.

6a. Calculate the eigenvector and eigenvalue from the covariance matrix C_t.

6b. Retain the eigenvector $\boldsymbol{E}$ which corresponds to the highest eigenvalue which explains the largest fraction of variance in the p variables.

6c. Calculate the first principal component using $\boldsymbol{E}$ from (6b):

$$PC_t = \bar{\boldsymbol{X}}_t \boldsymbol{E} \tag{3}$$

$$PC_l = \bar{\boldsymbol{X}}_l \boldsymbol{E}, \qquad \forall l = \{1, 2, \ldots, L\} \tag{4}$$

where PC_t and PC_l are one-dimensional values transferred from the eigenvector in (6b) for the current day, t, and the lth of L potential neighbors. Only one principal component is retained following the recommendation of Eum and Simonovic (2012).

6d. Calculation of the Mahalanobis distance using the values obtained in Equations (3) and (4) as well as the variance, Var(PC), between all L values of PC_l.

$$d_l = \sqrt{\frac{(\mathrm{PC}_t - \mathrm{PC}_l)^2)}{\mathrm{Var}(\mathbf{PC})}} \qquad \forall l = \{1, 2, \ldots, L\} \qquad (5)$$

7. Select the number K of nearest neighbors to retain out of the L potential values. A value of $K = \sqrt{L}$ is recommended (Lall and Sharma, 1996; Rajagopalan and Lall, 1999; Yates *et al.*, 2003). Sort the Mahalanobis distance metric from smallest to largest, and retain the first K neighbors on the list. Use a discrete probability distribution as described in Equation (6) from Yates *et al.* (2003), which weights closest neighbors highest for resampling one of the K values.
8. Generate a random number, $u(0,1)$, and compare this to the cumulative probability, p_{m}, to determine the current day's nearest neighbor. The day m for which u is closest to p_{m} is selected as the nearest neighbor and the corresponding weather is used for all stations in the region. Through this step, spatial correlation among the variables is preserved.
9. Resample B days from the historical record which follow the selected day (m) from Step (8). B is the block length and is selected based on the observed autocorrelations of daily temperatures (lag 1, lag 2, etc.). It should be as large as required for the model to reproduce the observed temperature autocorrelations. If the selected day (m) is January 1, and the block length is $B = 5$, the climate variables from January 1, and the following daily variables on January 2, January 3, January 4, and January 5 will be sampled. The next day to be sampled from the nearest neighbors is then January 6 and this will occur following perturbation (Step 10). In KnnCAD V3, $B = 1$ and only one day is resampled at a time.
10. Perturbation of the reshuffled variable values for days t to $t + b$ (where $b = 0, 1, \ldots, B$). A new perturbation component is introduced to ensure simulation of unique but reasonable values for temperature and precipitation amount that can lie outside of the observed ranges. See Eum and Simonovic (2012) for a description of the perturbation methodology used in KnnCAD V3. Because precipitation has a non-negativity constraint, it must be dealt with differently from temperature. As such, the same interpolation equation is used for both precipitation and temperature with different randomly distributed variables, Z_{t+b}, as shown in Steps (10a) and (10b).

10a. Perturbation of the reshuffled temperature values $x^j_{i,t+b}$ for temperature variable i, station j and day $t+b$ following Equation (6):

$$y^j_{i,t+b} = \lambda_{\mathrm{temp}} x^j_{i,t+b} + (1 - \lambda_{\mathrm{temp}}) Z_{t+b} \qquad (6)$$

where $y^j_{i,t+b}$ is the simulated perturbed value and λ_{temp} is chosen between 0 and 1 during calibration (1 gives an unperturbed result and 0 yields a result based entirely on perturbation). For preservation of temporal correlations, λ_{temp} should be as large as is reasonable. Z_{t+b} is a normally distributed value with a mean of $x^j_{i,t+b}$ and a standard deviation of $\sigma^j_{i,t}$, calculated from the K-NNs for day t, station j, and temperature variable i. To prevent minimum temperature from exceeding maximum temperature, the same random normal variable z is used for both maximum and minimum temperature across all stations and its value is transformed using the variables' corresponding $x^j_{i,t+b}$ and $\sigma^j_{i,t}$ values.

10b. Perturbation of the reshuffled non-zero precipitation values $x^j_{\mathrm{ppt},t+b}$ for station j and day $t+b$ (where $b = 1, 2, \ldots, B$), following Equation (7):

$$y^j_{\mathrm{ppt},t+b} = \lambda_{\mathrm{ppt}} x^j_{\mathrm{ppt},t+b} + (1 - \lambda_{\mathrm{ppt}}) Z_{t+b} \qquad (7)$$

where Z_{t+b} comes from a two-parameter lognormal distribution. The parameters for the lognormal distribution are calculated using the method of moments from Singh (1998). The Singh (1998) equations use a mean equal to the unperturbed precipitation value and a standard deviation equal to that of the non-zero values in the potential neighbors subset (the days that lie within a temporal window centered on the current day t). λ_{ppt} is chosen between 0 and 1 and should be as large as is reasonable to preserve spatial correlation. The proposed perturbation scheme inherently produces values above zero while still producing perturbed precipitation amounts that can be either higher or lower than the unperturbed value.

11. Repeat Steps 6 through 10 for time step $t + B + 1$. The simulated value on day $t + B$ will be compared to the potential neighbors of day $t + B + 1$ to determine the selection of nearest neighbors for day $t + B + 1$ as outlined

FIGURE 1. The Upper Thames River Basin.

in Steps 6 through 7. Selection of the next day and perturbation of the block will follow as outlined in Steps 8 through 10. This process is repeated until the end of the historical record is reached, at which point additional simulations can be performed to generate long synthetic datasets.

The KnnCAD program is coded in R programming language and has a Visual Basic decision support system to aid researchers in applying the program for the study area. The user may vary parameters such as the block length, B, and the interpolation parameters, λ_{temp} and λ_{ppt}, to determine which combination of parameters provides the best calibration based on the outputs provided by the decision support system. B should be selected as large as required to reproduce daily temporal correlations in the observed temperature series. Parameters λ_{temp} and λ_{ppt} (which are chosen between 0 and 1) should be selected as large as is reasonable to preserve spatial correlations while still producing values outside of the historical range.

APPLICATION

To illustrate the utility of the KnnCAD V4 model, a case study of the UTRB is presented. KnnCAD V4 outputs are compared with outputs from the previous version of the weather generator, KnnCAD V3. The UTRB, shown in Figure 1, is located between the Great Lakes of Erie and Huron and has a population of 515,640, the majority residing in London, Ontario,

TABLE 1. Upper Thames River Stations.

Station	Latitude (°N)	Longitude (°W)	Elevation (m)
Blyth	43.72	81.38	350.5
Brantford	43.13	80.23	196.0
Chatham	42.38	82.2	198.0
Delhi CS	42.87	80.55	255.1
Dorchester	43.00	81.03	271.3
Embro	43.25	80.93	358.1
Exeter	43.35	81.50	262.1
Fergus	43.73	80.33	410.0
Foldens	43.02	80.78	328.0
Glen Allan	43.68	80.71	404.0
Hamilton A	43.17	79.93	238.0
Ilderton	43.05	81.43	266.7
London A	43.03	81.16	278.0
Petrolia Town	42.86	82.17	201.2
Ridgetown	42.45	81.88	210.3
Sarnia	43.00	82.32	191.0
Stratford	43.37	81.00	354.0
St. Thomas	42.78	81.21	209.0
Tillsonburg	42.86	80.72	270.0
Waterloo Wellington	43.46	80.38	317.0
Woodstock	43.14	80.77	282.0
Wroxeter	43.86	81.15	355.0

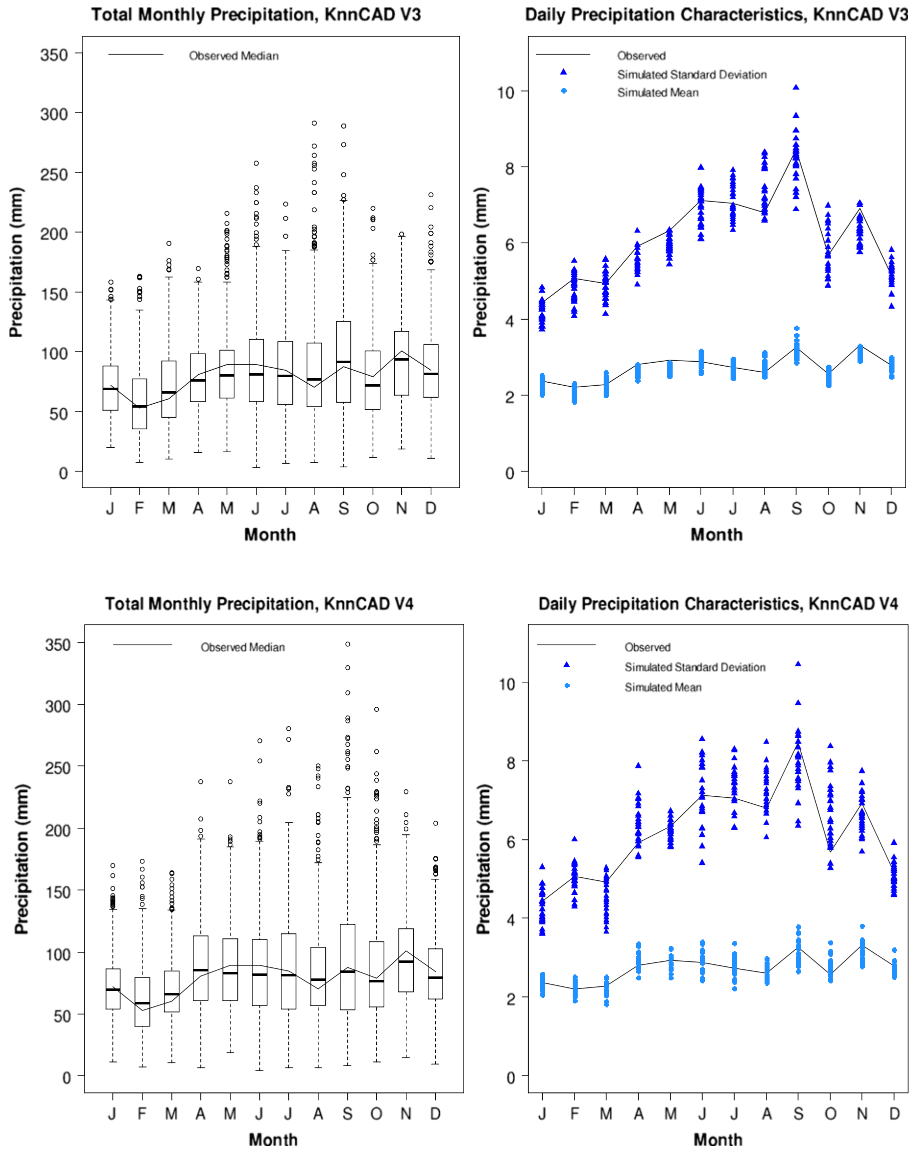

FIGURE 2. KnnCAD V3 and V4-Simulated and Observed Total Monthly Precipitation and Daily Precipitation Characteristics at London A.

the major urban center in the region. The basin has a history of major flooding events, often occurring in March or April following snowmelt. Floods can also occur after a sudden peak in temperatures during the winter or in the summer after extreme precipitation events (Wilcox *et al.*, 1998). There have been a number of studies assessing the potential impacts of climate change on the UTRB, indicating vulnerability of the basin to future extreme precipitation events and flooding (Sharif and Burn, 2006, 2007; Simonovic, 2010; Solaiman *et al.*, 2010; Eum and Simonovic, 2012; King *et al.*, 2012).

A total of 22 stations around the basin are used in this study based on the availability and completeness of the Environment Canada datasets (Government of Canada, 2014). A 27-year historical record from 1979 to 2005 is gathered from each station. The selected record length is chosen based on the availability and completeness of Environment Canada data at each of the 22 weather stations. The station locations can be found in Figure 1. Table 1 provides the names, elevations, latitudes, and longitudes of each station. The historical data from each of these stations is used as input to the KnnCAD V3 and V4 algorithms which are then used to produce 25 ensembles of synthetic historical climate data. Each ensemble has the same length as the input dataset (27 years) so in total 675 years of synthetic historical climate data is output by the weather generator. It is important to generate a number of ensembles due to the random component of the program as each simulation is different. By varying the KnnCAD V4 parameters for block length and interpolation, it was found that $B = 10$ days and $\lambda_{ppt} = \lambda_{temp} = 0.9$ provided the best calibration result. For more details on the validation procedure for KnnCAD V4, please refer to King (2012).

The ability of the algorithm to simulate historical climate characteristics is investigated in terms of total monthly precipitation, daily precipitation characteristics, extreme precipitation events, wet spell lengths, as well as mean and extreme daily temperatures. Monthly boxplots are used to demonstrate some of the results, with historical medians shown as a line plot. The upper and lower lines in the boxplots represent the quartiles, the middle line represents the median and the whiskers extend to 1.5 times the interquartile range. Dot plots are also used in order to show the spread of results from the different simulation ensembles. The observed values are shown as a line plot and each dot clustered over the historical observation represents the result for one ensemble of climate data, for a total of 25 dots in each grouping.

In order to demonstrate the utility of KnnCAD V4 as a climate simulation tool, the results for the London A station have been compared with simulated outputs from KnnCAD V3. Historical data are compared to KnnCAD V3 using CLIMDEX indices.

Figure 2 shows the precipitation results for total monthly precipitation and daily precipitation characteristics from the London A station in the UTRB from KnnCAD V3 and V4. Results at other climate stations are similar. Results indicate that both KnnCAD V3 and V4 perform well at simulating mean total monthly precipitation amounts. KnnCAD V4 is able to generate slightly higher extreme monthly precipitation values as a result of the improved perturbation

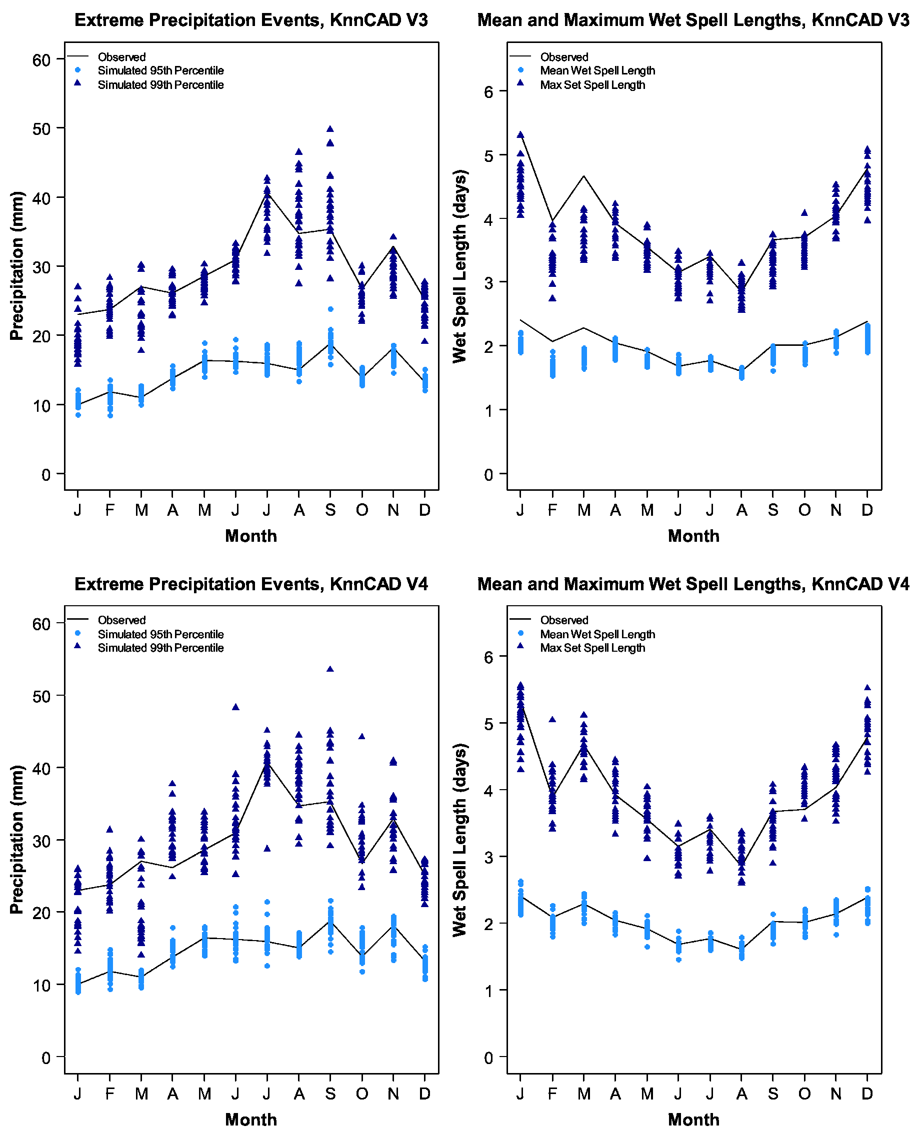

FIGURE 3. KnnCAD V3 and V4-Simulated and Observed Extreme Daily Precipitation and Wet Spell Lengths at London A.

method. For the plots showing monthly means and standard deviations of the daily precipitation amounts, the spread of points is generally centered on the historical observation indicating good performance of both KnnCAD V3 and V4. KnnCAD V4 simulates increased variability in the standard deviations over V3 due to the perturbation methodology applied; however, there is a slight overestimation in the standard deviations for April and October. KnnCAD V3 slightly underestimates the standard deviations in April, May, and November. It is clear from the figures that while there are slight over- and underestimations in some months, both KnnCAD V3 and V4 simulate daily and monthly precipitation characteristics quite well.

Figure 3 shows the KnnCAD V3 and V4-simulated extreme daily precipitation events at London A on the left and wet spell lengths on the right. Both KnnCAD V3 and V4 simulate 95th and 99th percentile precipitation events quite well, with the spread of the points generally centered on the historical observation. KnnCAD V3 slightly underestimates the 99th percentile events for the months of May, June, and October to December. There is more variation observed in the 99th percentile simulations from KnnCAD V4, and this may be a result of the perturbation method used. KnnCAD V4 overestimates 99th percentile precipitation in April but for most of the other months the spread of ensemble points is generally centered over the historical value. KnnCAD V4 outperforms KnnCAD V3 in the

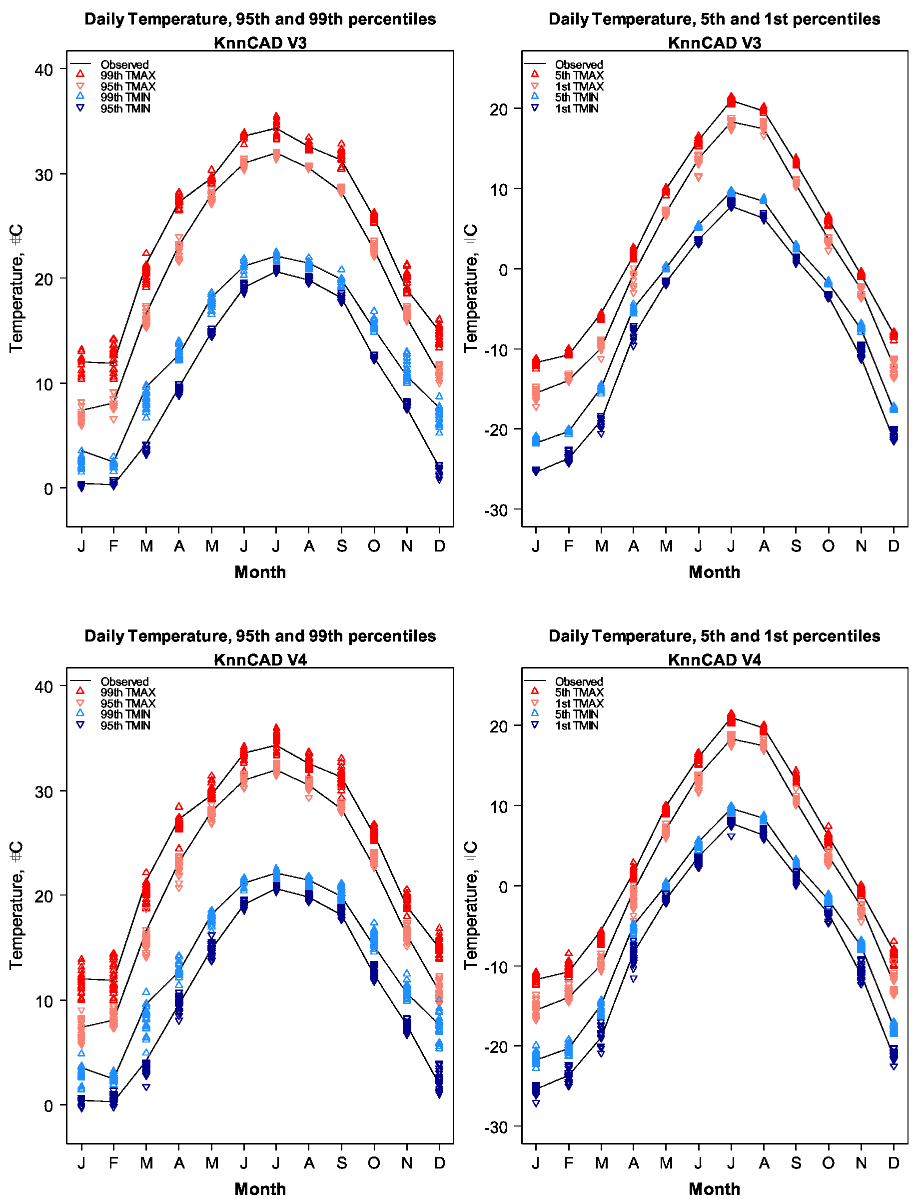

FIGURE 4. KnnCAD V3 and V4-Simulated and Observed Extreme Daily Maximum and Minimum Temperatures at London A.

simulation of mean and maximum monthly wet spell lengths, shown on the right of Figure 3. KnnCAD V3 underestimates mean and maximum wet spell lengths in several months. KnnCAD Version 4 performs very well as the points are centered on the historical observation in all months but the maximum wet spell length for October. Results for dry spell lengths are similar. The improvement over KnnCAD V3 is likely due to the block resampling addition in the program which helps preserve wet and dry spell structure. In KnnCAD V3, resampling occurs on a daily basis and therefore spell sequences are less likely to be preserved.

Figure 4 shows the results for extreme high and low temperature values from KnnCAD V3 and V4. Because the median values are simulated quite well by both KnnCAD V3 and V4, they are not presented. It is clear in the figures that the simulated dots for both KnnCAD V3 and V4 are very close to the observed values which are shown as line plots. There is increased variability in the simulated extreme temperature values from KnnCAD V4 due to the addition of temperature perturbation. KnnCAD V3 does not perturb reshuffled temperature values and is unable to produce temperature values which fall outside of the observed range.

Figures 5 and 6 show the KnnCAD V3 and V4-simulated 95th and 5th percentile temperatures on wet and dry days, respectively. Both models are able

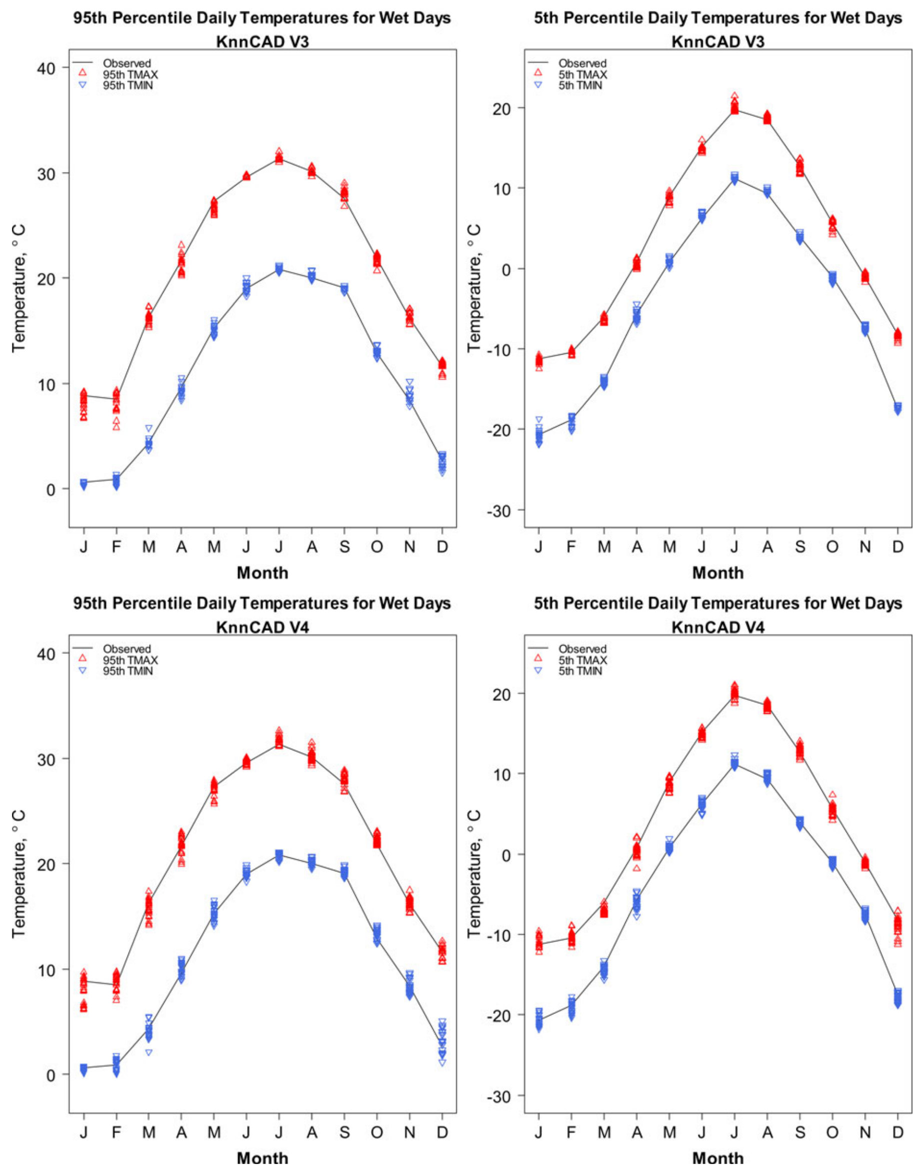

FIGURE 5. KnnCAD V3 and V4-Simulated and Observed 95th Percentile Daily Maximum and Minimum Temperatures on Wet Days at London A.

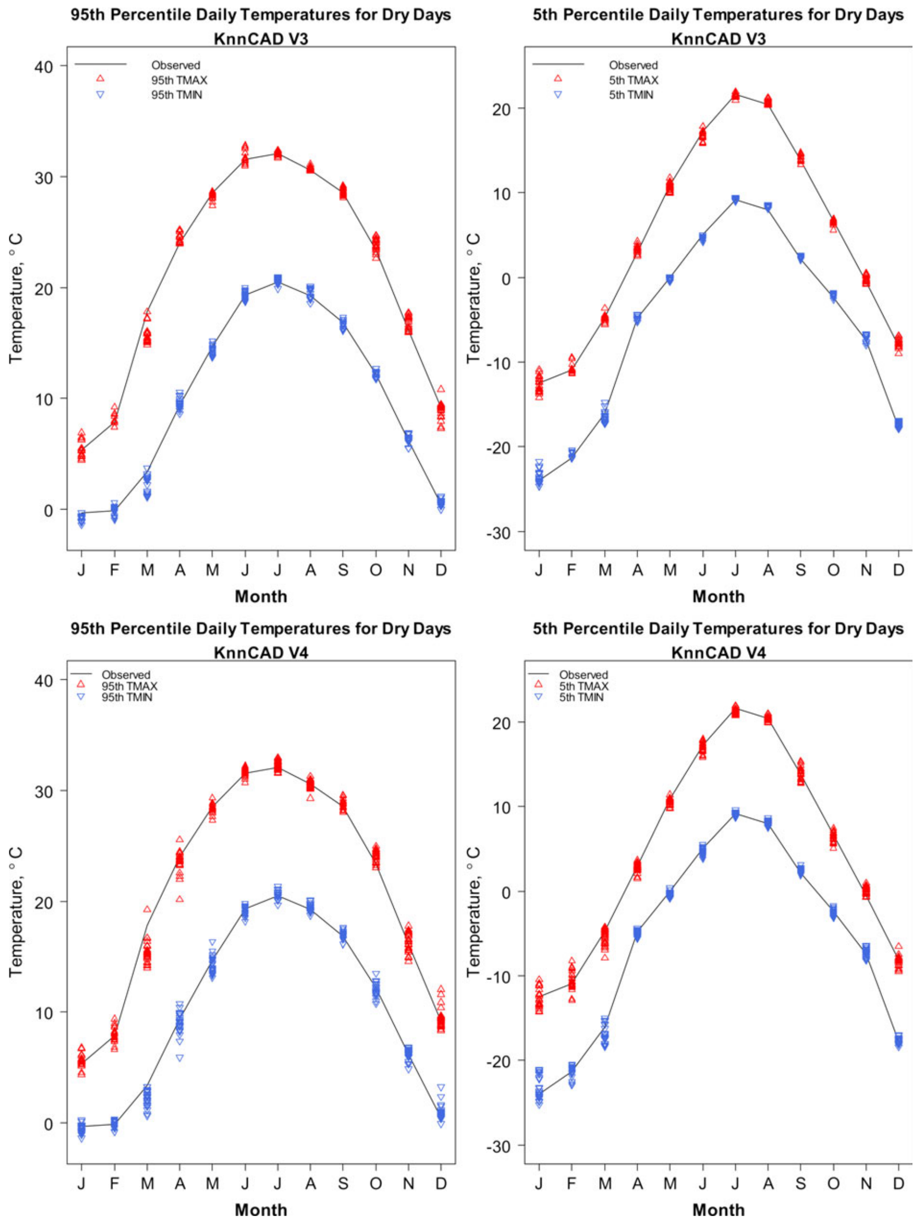

FIGURE 6. KnnCAD V3 and V4-Simulated and Observed 95th Percentile Daily Maximum and Minimum Temperatures on Dry Days at London A.

to simulate temperatures on wet and dry days very well, with slightly more spread in the values simulated by KnnCAD V4.

Figure 7 shows the lag-1 autocorrelations for daily maximum and minimum temperatures at the London A station. It is clear that by adding block resampling, the ability of the model to preserve the temporal correlation structure in the observed record is significantly improved. This is an important factor in the UTRB where snow accumulation and melt lead to major flooding events. Hydrologic models often use temperature-index methods for prediction of snowmelt (Buttle, 2009; Jenicek *et al.*, 2012; Kumar *et al.*, 2013) so the use of a temporally correlated series of temperatures could significantly improve snowpack prediction and accuracy of snowmelt timing and magnitude. Other weather generators, such as SDSM and LARS-WG, are inherently unable to simulate temporally correlated temperature series due to their stochastic nature (King, 2012).

Figure 8 contains plots of snow accumulation, simulated using a rudimentary temperature-index model (see Cunderlik and Simonovic, 2004) to demonstrate the improvement in timing and quantity of snowmelt resulting from the implementation of block resampling. The mean, 95th, and 5th percentile daily snowpack amount was calculated from the observed record and simulated KnnCAD V3 and V4 results are shown as a

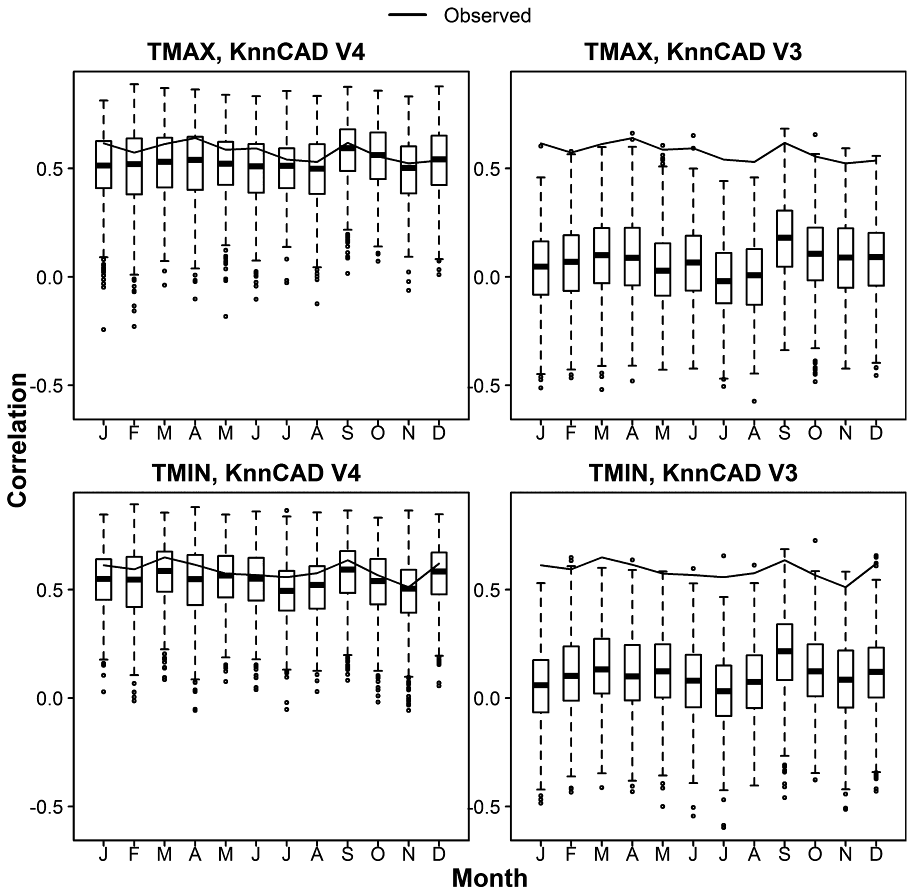

FIGURE 7. Simulated and Observed Lag-1 Autocorrelations in Daily Maximum and Minimum Temperatures at London A from KnnCAD V3 and V4.

scatterplot overlaying the historical observation. It is clear that KnnCAD V4 offers a significant improvement in simulation of snowpack and runoff timing.

CLIMDEX extreme climate indices (CLIMDEX, 2012) have been included in Table 2 to provide a comparison between KnnCAD V3 and V4. A description of each index is provided in the table. The mean values of selected indices are presented for the historical dataset, as well as for the simulated outputs. For some of the monthly indices, mean annual values are shown in order to effectively compare outputs from KnnCAD V3 and V4.

Both models performed quite well in reproducing CLIMDEX extreme climate indices, with mean values that were quite close to the historical values in most cases. KnnCAD V4 simulations produced values that were closer to the historical record for annual total precipitation amount as well as maximum dry and wet spell lengths. Both models performed very well in simulation of r10 mm and r20 mm indices, with V3 slightly outperforming V4. Monthly maximum one-day and five-day values were slightly overestimated by both V3 and V4, with V3 generating values that were closer to the historical average. This is likely due to KnnCAD V4 simulating increased variability in extreme daily precipitation events, as shown in Figures 1 and 2. The simple precipitation intensity index was slightly overestimated by both models. For temperature indices TNn, TNx, TXn, TXx, and DTR, both models performed very well and KnnCAD V4 outperformed V3 in most cases. Both models performed well in the simulation of frost days, icing days, summer days, and tropical nights, however, KnnCAD V3 slightly outperformed V4 for each of these indices excluding summer days. For warm spell duration index and cold spell duration index, it is clear that KnnCAD V4 simulates values much closer to the historical average. This is likely a result of the block resampling which preserves temperature autocorrelations, as shown in Figure 8. Overall, both models performed very well in simulation of the CLIMDEX indices, with the most major discrepancy in the KnnCAD V3 simulation of warm and cold spells.

Figures 9 and 10 present the simulated and historical correlations of daily maximum and minimum temperatures across all station pairs from KnnCAD

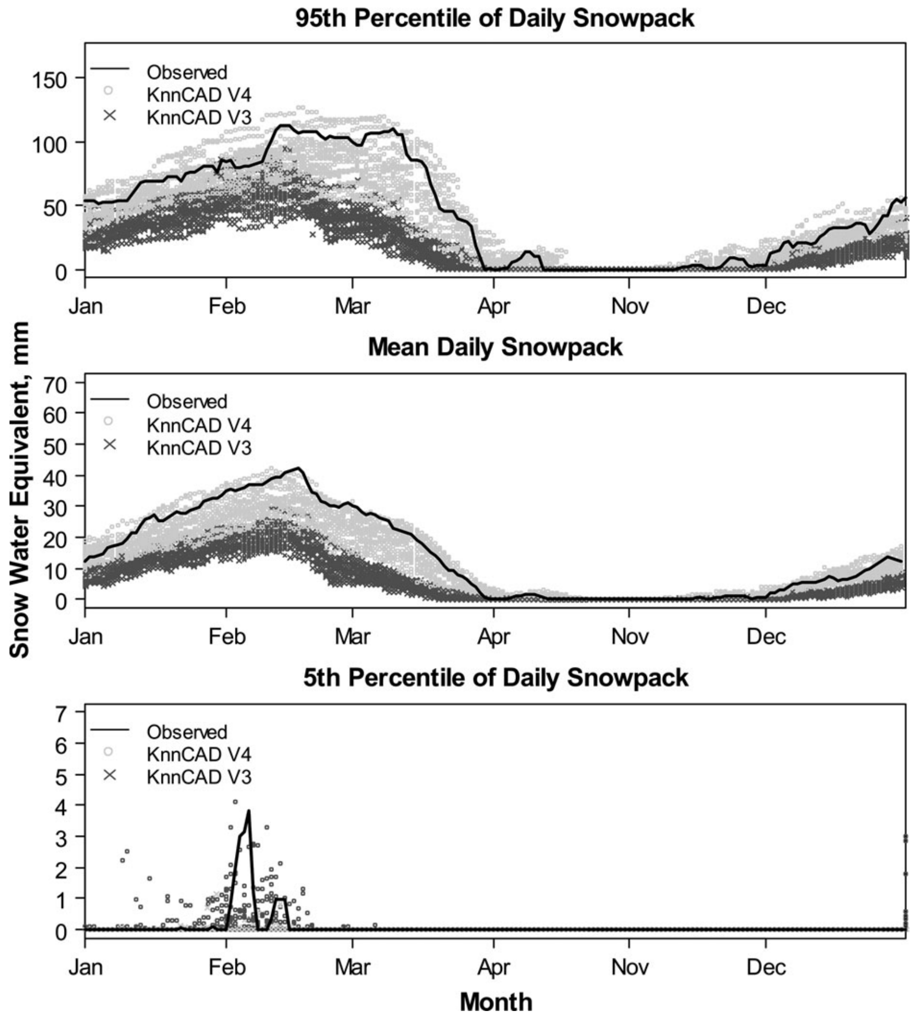

FIGURE 8. Mean, 5th, and 95th Percentiles of Observed Daily Snowpack (January through April, November and December) Compared with Simulation Outputs from KnnCAD V3 and V4.

V3 and V4, respectively. Results are presented for one ensemble (of a total of 25) as results from the other ensembles are similar. For 22 stations there are a total of 253 station pairs and the correlations between daily maximum and minimum temperature between each of these pairs was computed from the historical and simulated dataset. It is clear from the plots that correlations are preserved very well. This is an inherent advantage of the KnnCAD algorithm over the parametric and semiparametric weather generators; it is able to accurately simulate spatial correlations without making any statistical assumptions.

Figures 11 and 12 present the simulated and historical correlations of maximum and minimum temperatures at the monthly time step, across all station pairs for KnnCAD V3 and V4, respectively. Both models slightly overestimate the correlations; however, KnnCAD V4 outperforms V3 with correlation values that are closer to the 1:1 line.

Figures 13 and 14 present the simulated and historical correlations of annual precipitation totals across all station pairs for KnnCAD V3 and V4, respectively. Both models generally simulate the spatial correlations well. However, there is a slight improvement in the results for KnnCAD V4.

CONCLUSIONS

The KnnCAD Version 4 weather generator provides an improvement over Version 3 by adding block resampling to improve the temporal correlation structure in temperatures. KnnCAD V4 is able to simulate temperature autocorrelations quite accurately while these are not reproduced well in V3. Temporally correlated temperature data is of particular importance for inflow forecasting in study basins such as

TABLE 2. Historical and Simulated Mean Values of CLIMDEX Climate Extreme Indices from KnnCAD V3 and V4.

CLIMDEX Extreme Climate Index	Description	Historical	KnnCAD-V3	KnnCAD-V4
PRCPTOT (mm)	Annual total precipitation	977.18	951.96	973.92
CDD (days)	Maximum dry spell length (PPT < 1 mm)	14.37	15.61	14.81
CWD (days)	Maximum wet spell length (PPT > 1 mm)	6.56	5.98	6.04
r10 mm (days)	Annual count of days with PPT < 10 mm	30.93	30.93	30.47
r20 mm (days)	Annual count of days with PPT < 20 mm	9.85	9.96	10.18
rx1 day (days)	Monthly maximum 1-day PPT	49.71	50.99	55.61
rx5 day (days)	Monthly maximum 5-day PPT	78.91	81.23	84.22
SDII (mm/day)	Simple precipitation intensity index	7.56	7.89	7.76
TNn (°C)	Annual minimum value of daily minimum temperature	−23.45	−23.41	−23.97
TNx (°C)	Annual maximum value of daily minimum temperature	21.78	22.04	21.99
TXn (°C)	Annual minimum value of daily maximum temperature	−14.05	−14.25	−14.18
TXx (°C)	Annual maximum value of daily maximum temperature	32.77	34.22	33.95
DTR (°C)	Diurnal temperature range	9.71	9.73	9.72
FD (days)	Frost days (annual count of days with $T_{min} < 0$)	140.78	141.15	143.27
ID (days)	Icing days (annual count of days with $T_{max} < 0$)	57.93	58.17	58.88
SU (days)	Summer days (annual count of days with $T_{max} > 25$)	62.63	63.21	62.96
TR (days)	Tropical nights ($T_{min} > 20$)	4.96	4.89	5.37
WSDI (days)	Warm spell duration index (annual count of 6-day periods with T_{max} > 90th percentile)	3.67	0.37	2.15
CSDI (days)	Cold spell duration index (annual count of 6-day periods with T_{min} < 90th percentile)	0.78	0.28	0.79

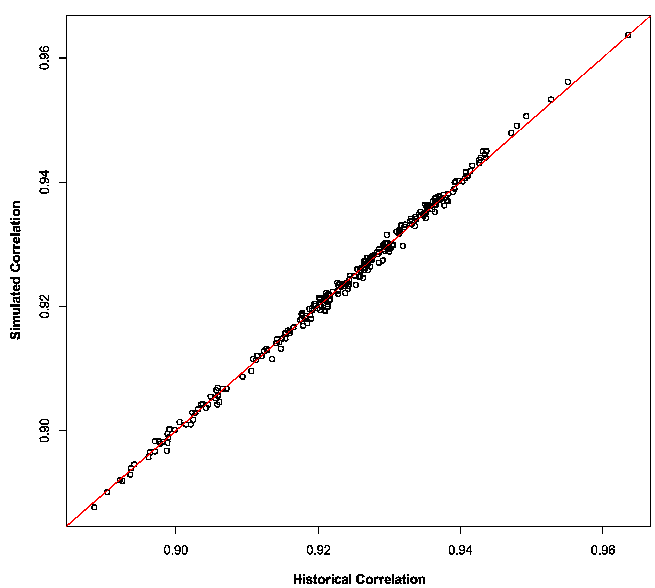

FIGURE 9. KnnCAD V3-Simulated and Historical Correlations of Daily Maximum and Minimum Temperatures for All Possible Station Pairs, First Ensemble (results similar for other ensembles).

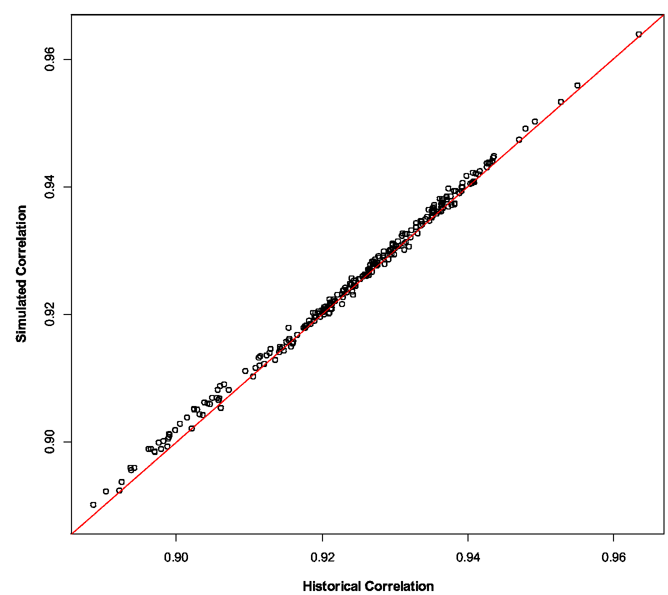

FIGURE 10. KnnCAD V4-Simulated and Historical Correlations of Daily Maximum and Minimum Temperatures for All Possible Station Pairs, First Ensemble (results similar for other ensembles).

the UTRB, where snowmelt events can cause major floods. Temperature-index hydrologic models rely on accurate time series of temperature data to predict snow accumulation and melting. KnnCAD V4 was shown to improve simulation of snow accumulation and melt in the basin. The block resampling in V4 is also shown to enhance simulation of wet and dry spell lengths as well as cold and warm spells. KnnCAD V4 also includes an improved perturbation scheme to enhance the simulation of extreme temperature and precipitation values. Both of the KnnCAD models are shown to simulate effectively the historical climate variables and extreme climate indices at several sites simultaneously, while preserving spatial correlations. The nearest neighbor models can be applied to multiple sites without making statistical assumptions regarding variables' probability distributions and spatial correlations between weather

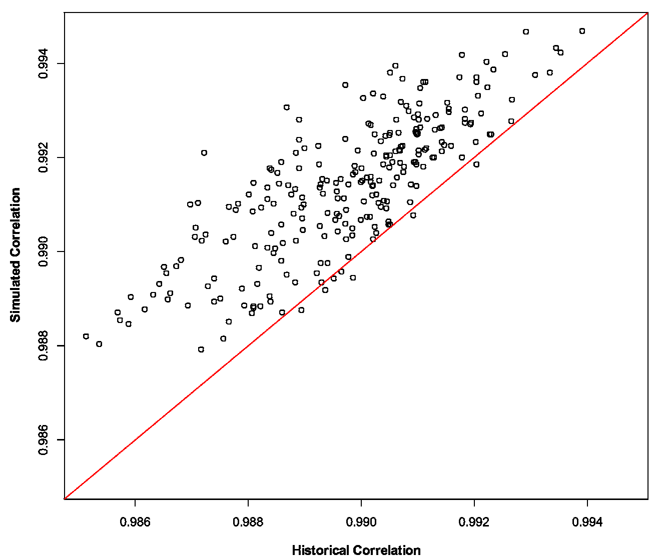

FIGURE 11. KnnCAD V3-Simulated and Historical Correlations of Monthly Maximum and Minimum Temperatures for All Possible Station Pairs, First Ensemble (results similar for other ensembles).

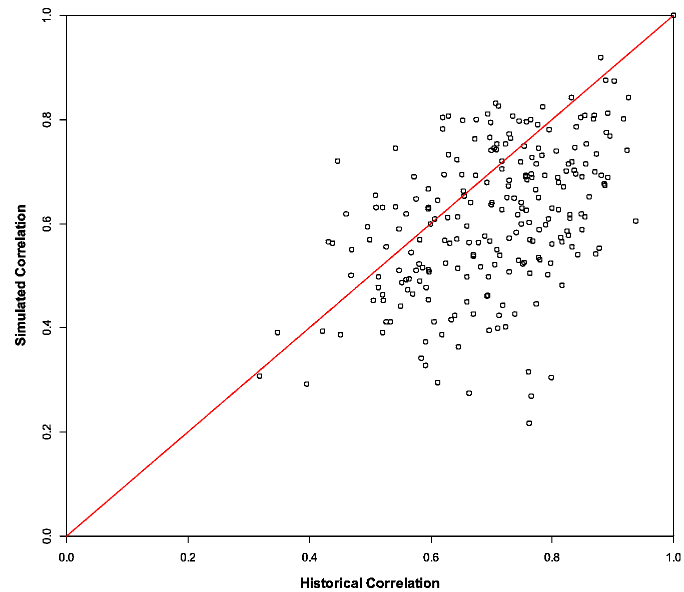

FIGURE 13. KnnCAD V3-Simulated and Historical Correlations of Annual Precipitation Totals for All Possible Station Pairs, First Ensemble (results similar for other ensembles).

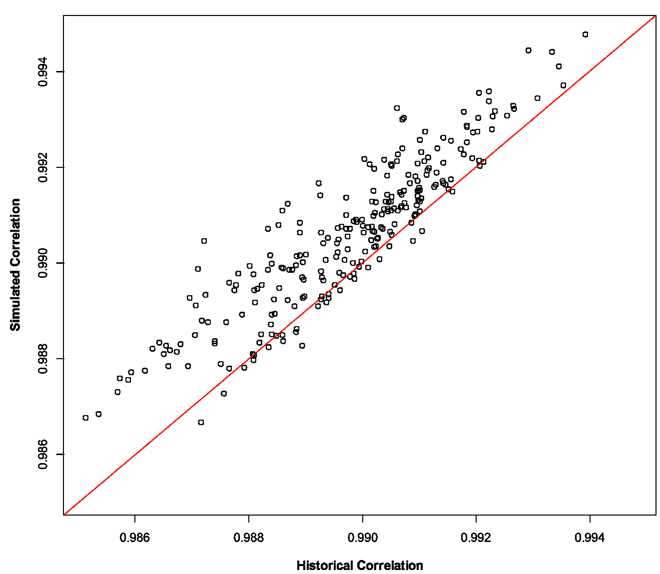

FIGURE 12. KnnCAD V4-Simulated and Historical Correlations of Monthly Maximum and Minimum Temperatures for All Possible Station Pairs, First Ensemble (results similar for other ensembles).

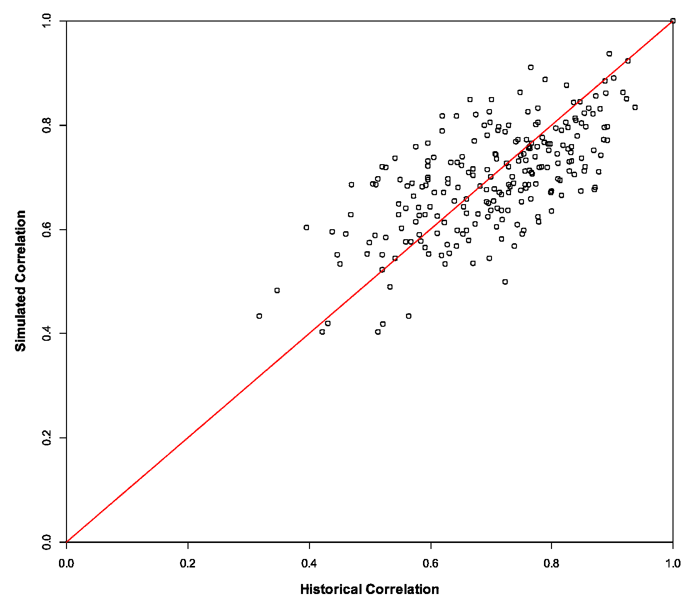

FIGURE 14. KnnCAD V4-Simulated and Historical Correlations of Annual Precipitation Totals for All Possible Station Pairs, First Ensemble (results similar for other ensembles).

stations. As such, the nearest-neighbor models provide a major advantage over the semiparametric and parametric weather generators.

Currently, the KnnCAD Version 4 model is being validated for other climatic regions such as the Grand River Basin, Ontario and two watersheds in Brazil. Applications of the model employing a wavelet autoregressive model could improve simulation of low-frequency variability in the simulated data (see Kwon *et al.*, 2007; Steinschneider and Brown, 2013), and should be considered in future research. A new methodology for developing AOGCM-modified input datasets for KnnCAD Version 4 from daily AOGCM data is another important area for future research, as the current weather generator cannot create future climate scenarios that take into account changes in the variability of future daily temperatures and precipitation amounts.

ACKNOWLEDGMENTS

Dr. Ian McLeod and Dr. Slobodan Simonovic acknowledge with thanks support from the NSERC Discovery Grant Program. The authors wish to thank Environment Canada for the climate data used in this study, as well as Ontario Graduate Scholarships (OGS) and the Canadian Federation for Climate and Atmospheric Sciences (CFCAS) for funding this research.

LITERATURE CITED

Apipattanavis, S., G. Podestá, B. Rajagopalan, and R.W. Katz, 2007. A Semiparametric Multivariate and Multisite Weather Generator. Water Resources Research 43(11):W11401, doi:10.1029/2006 WR005714.

Buttle, J.M., 2009. Using a Temperature-Based Model of Snow Accumulation and Melt to Assess the Long Term Hydrologic Behaviour of Forested Headwater Basins in South-Central Ontario. In Proceedings of 66th Eastern Snow Conference, Niagara on the Lake.

Charles, S.P., M.A. Bari, A. Kitsios, and B.C. Bates, 2007. Effect of GCM Bias on Downscaled Precipitation and Runoff Projections for the Serpentine Catchment, Western Australia. International Journal of Climatology 27:1673-1690, doi:10.1002/joc.1508.

CLIMDEX, 2012. Climate Extremes Indices. http://www.climdex.org/indices.html, *accessed* February 2013.

Craigmile, P.F. and P. Guttorp, 2011. Space-Time Modelling of Trends in Temperature Series. Journal of Time Series Analysis 32:378-395, doi:10.1111/j.1467-9892.2011.00733.x.

Cunderlik, J. and Simonovic, S.P., 2004 Calibraion, Verification and Sensitivity Analysis of the HEC-HMS Hydrologic Model CFCAS Project: Assessment of Water Resources Risk and Vulnerability to Changing Climatic Condition. Water Resources Research Report No. 048, Facility for Intelligent Decision Support, Department of Civil and Environmental Engineering, The University of Western Ontario, London, Ontario, Canada

Dibike, Y.B. and P. Coulibaly, 2005. Hydrologic Impact of Climate Change in the Saugenay Watershed: Comparison of Downscaling Methods and Hydrologic Models. Journal of Hydrology 307:144-163, doi:10.1016/j.jhydrol.2004.10.012.

Elshamy, M.E., H.S. Wheater, N. Gedney, and C. Huntingford, 2006. Evaluation of the Rainfall Component of a Weather Generator for Climate Impact Studies. Journal of Hydrology 326: 1-24, doi:10.1016/j.jhydrol.2005.09.017.

Eum, H.-I. and S.P. Simonovic, 2012. Assessment on Variability of Extreme Climatic Events for the Upper Thames River Basin in Canada. Hydrological Processes 26(4):485-499, doi:10.1002/hyp.8145.

Eum, H.-I., S.P. Simonovic, and Y.-O. Kim, 2010. Climate Change Impact Assessment Using K-Nearest Neighbour Weather Generator: Case Study of the Nakdong River Basin in Korea. Journal of Hydrologic Engineering 15:772-785, doi:10.1061/_ASCE_HE.1943-5584.0000251.

Eum, H.-I., A. Vassn, and S.P. Simonovic, 2009. Integrated Reservoir Management System for Adaptation to Climate Change Impacts in the Upper Thames River Basin. Water Resources Research Report No. 062, Facility for Intelligent Decision Support. Department of Civil and Environmental Engineering, London, Ontario.

Geng, S., J.S. Auburn, E. Brandstetter, and B. Li, 1988. A Program to Simulate Meteorological Variables: Documentation for SIMMETEO, Agronomy Progress Rep 204. Department of Agronomy and Range Science, University of California, Davis, California.

Government of Canada, 2014. National Climate Data and Information Archive. http://climate.weather.gc.ca/index_e.html, *accessed* April 2015.

Hanson, C.L. and G.L. Johnson, 1998. GEM (Generation of Weather Elements for Multiple Applications): Its Application in Areas of Complex Terrain. Hydrology, Water Resources and Ecology in Headwaters IAHS 248:27-32.

Hashmi, M.Z., A.Y. Shamseldin, and B.W. Melville, 2011. Comparison of LARS-WG and SDSM for Simulation and Downscaling of Extreme Precipitation Events in a Watershed. Stochastic Environmental Research and Risk Assessment 25:475-484, doi:10.1007/s00477-010-0416-x.

Jenicek, M., H. Beitlerova, M. Hasa, D. Kucerova, H. Pevna, and S. Podzimek, 2012. Modelling Snow Accumulation and Snowmelt Runoff-Present Approaches and Results. AUC Geographica 47(2): 15-24.

Karlsson, M. and S. Yakowitz, 1987. Nearest Neighbour Methods for Nonparametric Rainfall-Runoff Forecasting. Water Resources Research 23(7):1300-1308.

King, L.M., 2012. Application of a K-Nearest Neighbor Weather Generator for Simulation of Historical and Future Climate Variables in the Upper Thames River Basin. MESc Thesis, Department of Civil and Environmental Engineering, University of Western Ontario, London, Ontario, Canada.

King, L.M., S. Irwin, R. Sarwar, A.I. McLeod, and S.P. Simonovic, 2012. The Effects of Climate Change on Extreme Precipitation Events in the Upper Thames River Basin: A Comparison of Downscaling Approaches. Canadian Water Resources Journal 37(3):253-274, doi:10.4296/cwrj2011-938.

Kuchar, L., 2004. Using WGENK to Generate Synthetic Daily Weather Data for Modelling of Agricultural Processes. Mathematics and Computers in Simulation 65:69-75, doi:10.1016/j.matcom.2003.09.009.

Kumar, M., D. Marks, J. Dozier, M. Reba, and A. Winstral, 2013. Evaluation of Distributed Hydrologic Impacts of Temperature-Index and Energy-Based Snow Models. Advances in Water Resources 56:77-89, doi:10.1016/j.advwatres.2013.03.006.

Kwon, H.-H., U. Lall, and A.F. Khalil, 2007. Stochastic Simulation Model for Nonstationary Time Series Using an Autoregressive Wavelet Decomposition: Applications to Rainfall and Temperature. Water Resources Research 43(5):W05407, doi:10.1029/2006 WR005258.

Kwon, H.-H., B. Sivakumar, Y.-I. Moon, and B.-S. Kim, 2011. Assessment of Change in Design Flood Frequency Under Climate Change Using a Multivariate Downscaling Model and a Precipitation-Runoff Model. Stochastic Environmental Research and Risk Assessment 25:567-581, doi:10.1007/s00477-010-0422-z.

Lall, U. and A. Sharma, 1996. A Nearest Neighbor Bootstrap for Resampling Hydrologic Time Series. Water Resources Research 32(3):679-693.

Lee, T., T.B.M.J. Ouarda, and C. Jeong, 2012. Nonparametric Multivariate Weather Generator and an Extreme Value Theory for Bandwidth Selection. Journal of Hydrology 452-453:161-171.

Mehrotra, R. and A. Sharma, 2007. Preserving Low-Frequency Variability in Generated Daily Rainfall Sequences. Journal of Hydrology 345:102-120, doi:10.1016/j.jhydrol.2007.08.003.

Prodanovic, P. and S.P. Simonovic, 2008. Intensity Duraton Frequency Analysis under Changing Climatic Conditions. *In*: 4th International Symposium on Flood Defense: Managing Flood Risk, Reliability and Vulnerability. S.P. Simonovic and P. Bourget (Editors), Toronto, Canada.

Qian, B., S. Gameda, and H. Hayhoe, 2008. Performance of Stochastic Weather Generators LARS-WG and AAFC-WG for

Reproducing Daily Extremes of Diverse Canadian Climates. Climate Research 37:17-33, doi:10.3354/cr00755.

Qian, B., S. Gameda, H. Hayhoe, R. De Jong, and A. Bootsma, 2004. Comparison of LARS-WG and AAFC-WG Stochastic Weather Generators for Diverse Canadian Climates. Climate Research 26:175-191, doi:10.3354/cr026175.

Rajagopalan, B. and U. Lall, 1999. A k-Nearest Neighbour Simulator for Daily Precipitation and Other Weather Variables. Water Resources Research 35(10):3089-3101.

Richardson, C.W., 1981. Stochastic Simulation of Daily Precipitation, Temperature, and Solar Radiation. Water Resources Research 17:182-190.

Semenov, M.A. and E.M. Barrow, 2002. LARS-WG: A Stochastic Weather Generator for Use in Climate Impact Studies, Version 3.0 User Manual. http://www.rothamsted.ac.uk/mas-models/download/LARS-WG-Manual.pdf, *accessed* August, 2012.

Semenov, M.A. and E.M. Barrow, 1997. Use of a Stochastic Weather Generator in the Development of Climate Change Scenarios. Climatic Change 35:397-414.

Sharif, M. and D.H. Burn, 2006. Simulating Climate Change Scenarios Using an Improved K-Nearest Neighbor Model. Journal of Hydrology 325:179-196, doi:10.1016/j.jhydrol.2005.10.015.

Sharif, M. and D.H. Burn, 2007. Improved K-Nearest Neighbour Weather Generating Model. Journal of Hydrologic Engineering 12:42-51, doi:10.1061/(ASCE)1084-0699(2007) 12:1(42).

Simonovic, S.P., 2010. A New Methodology for the Assessment of Climate Change Impacts on a Watershed Scale. Current Science 98(8):1047-1055.

Singh, V.P., 1998. Entropy-Based Parameter Estimation in Hydrology, Kluwer Academic Publishers, Dordrecht, the Netherlands, 79 pp.

Solaiman, T., L. King, and S.P. Simonovic, 2010. Extreme Precipitation Vulnerability in the Upper Thames River Basin: Uncertainty in Climate Model Projections. International Journal of Climatology 31(15):2350-2364, doi:10.1002/joc.2244.

Soltani, A. and G. Hoogenboom, 2003. A Statistical Comparison of the Stochastic Weather Generators WGEN and SIMMENTO. Climate Research 24:215-230, doi:10.3354/cr024215.

Steinschneider, S. and C. Brown, 2013. A Semiparametric Multivariate, Multisite Weather Generator with Low-Frequency Variability for Use in Climate Risk Assessments. Water Resources Research 49:7205-7220, doi:10.1002/wrcr.20528.

Wilby, R.L. and C.W. Dawson, 2007. SDSM 4.2- A Decision Support Tool for the Assessment of Regional Climate Change Impacts: User Manual. http://co-public.lboro.ac.uk/cocwd/SDSM/SDSM-Manual.pdf, *accessed* October, 2011.

Wilby, R.L., O.J. Tomlinson, and C.W. Dawson, 2003. Multisite Simulation of Precipitation by Conditional Resampling. Climate Research 23:183-194.

Wilcox, I., C. Quinlan, C. Rogers, M. Troughton, I. McCallum, A. Quenneville, E. Heagy, and D. Dool, 1998. The Thames River Watershed: Background Study for Nomination under the Canadian Heritage Rivers System. Upper Thames Conservation Authority, London, Canada.

Yates, D., S. Gangopadhyay, B. Rajagopalan, and K. Strzepek, 2003. A Technique for Generating Regional Climate Scenarios Using a Nearest-Neighbour Algorithm. Water Resources Research 39(7):1199-1213, doi:10.1029/2002WR001769.