5.6 SEASONAL AND SPATIAL VARIATIONS OF SERIAL AND CROSS-CORRELATION MATRICES USED BY STOCHASTIC WEATHER GENERATORS

Justin T. Schoof* and Scott M. Robeson
Indiana University, Bloomington, IN

1. INTRODUCTION

Stochastic weather generators have had many applications, including agricultural risk assessment and generation of regional climate change scenarios (Semenov *et al.* 1998; Wilks and Wilby 1999). As time-series models with several interconnected components, stochastic weather generators typically use Markov chains or wet/dry spell length distributions to simulate precipitation occurrence, gamma or mixed exponential distributions for precipitation amount, and a multivariate stochastic process for daily maximum and minimum air temperature (T_{max} and T_{min}) and total daily solar radiation (R). The generated sequences are designed to have the desired serial and cross-correlations between T_{max}, T_{min}, and R by two matrices, **A** and **B**, which are defined using lag-0 cross-correlations and lag-1 serial correlations (see Section 3.2 for details). In many weather-generator implementations, **A** and **B** are treated as constant with respect to location, time of year, and wet/dry status. However, deviations from these constant values have been documented (Schubert 1994; Hayhoe 1998) and several authors (e.g., Wilks and Wilby, 1999) have suggested using location-specific parameters in an effort to account for spatial and seasonal differences.

In this study, the effects of varying these stochastic model parameterizations are investigated. We examine the spatial and seasonal differences in the values of the lag-0 cross-correlations and lag-1 serial correlations, and hence **A** and **B**, as well as the differences between simulated weather series when **A** and **B** are held constant and when they are allowed to vary by location, wet/dry status, and time of year. In addition to the traditional statistics that are used to evaluate stochastic weather generators, diurnal temperature range (DTR = T_{max} - T_{min}), is used for model evaluation, as it is a derived variable that should have similar persistence in both the models and observations.

2. DATA

To estimate the impacts of seasonally and spatially varying autoregressive parameters, daily data were extracted for five climatically diverse locations across the contiguous U.S.A.: Boston, MA; Indianapolis, IN; Jacksonville, FL; Portland, OR; and Tucson, AZ. Hourly air temperature and solar radiation values for 1961 to 1990 were available at these five locations through the Solar and Meteorological Surface Observation Network dataset (i.e., SAMSON CD-ROMs), available from the National Climatic Data Center, Asheville, NC. To allow (future) comparisons and analysis with cooperative climatic data, daily T_{max} and T_{min} (°C) were calculated using a 7am observation time. Daily total solar radiation values (MJ m^{-2} day^{-1}) were produced by integration of hourly solar radiation observations.

3. DESCRIPTION OF WEATHER GENERATOR

The stochastic weather generator used in this research is based on the well-known WGEN model (Richardson and Wright 1984). Using a number of parameters estimated from observational data, the model traditionally generates daily values of precipitation occurrence, precipitation amount, T_{max}, T_{min}, and R. In this study, the objective was to evaluate the impact of the parameterizations of **A** and **B**. Therefore, precipitation amount was not simulated.

3.1. Precipitation Occurrence Component

Precipitation occurrence is simulated by a two-state, first-order Markov chain. The occurrence of precipitation depends on two parameters: p_{01}, the probability of a wet day following a dry day, and p_{11}, the probability of a wet day following a wet day. In this study, a wet day is defined as any day having precipitation > 0 mm. Depending on the precipitation occurrence simulation for the previous day, a uniform [0,1] random number is compared to the appropriate transition probability. If the random number is less than the transition probability, a wet day is simulated. Otherwise, a dry day is simulated.

3.2. Temperature and Radiation Component

Daily values of T_{max}, T_{min}, and R are simulated by a first-order multivariate stochastic process, as described by Matalas (1967). Harmonic analysis is used to construct annual cycles of the input variables and their standard deviations. Annual-cycle harmonics are fit to daily means and standard deviations for wet and dry days separately (in some cases, a given day of the year may have few wet or dry occurrences; therefore, a 15-day moving window was used to construct the daily means and standard deviations). The time series then are reduced to residual elements by subtracting the daily means, as defined by the harmonics.

The weather generator simulates daily residuals of T_{max}, T_{min}, and R for day *i* with the equation

$$\mathbf{X_i} = \mathbf{A\,X_{i-1}} + \mathbf{B}\boldsymbol{\varepsilon}_\mathbf{i} \qquad (1)$$

where $\mathbf{X_i}$ is a (3 × 1) matrix containing the current day's values of T_{max}, T_{min}, and R, $\mathbf{X_{i-1}}$ is a (3 × 1) matrix containing the previous day's values of T_{max}, T_{min}, and R,

* *Corresponding author address:* Justin T. Schoof, Indiana University, Dept. of Geography, Bloomington, IN 47405; e-mail: jschoof@indiana.edu

ε_i is a (3×1) vector of independent standard Normal values, and **A** and **B** are (3×3) matrices given by

$$\mathbf{A} = \mathbf{M_1 M_0^{-1}} \qquad (2)$$

$$\mathbf{BB^T} = \mathbf{M_0} - \mathbf{M_1 M_0^{-1} M_1^T} \qquad (3)$$

where $\mathbf{M_o}$ is the (3×3) matrix of lag-0 cross-correlations and $\mathbf{M_1}$ is the (3×3) matrix of lag-1 serial correlations. For example, $\mathbf{M_o}$ (1,2) is the correlation between T_{max} and T_{min} and $\mathbf{M_1}$ (1,2) is the correlation between T_{max} and T_{min} lagged by one day. While **A** can be directly computed, we used an iterative method to solve for **B**.

After generation of the residual series with Eq. 1, the daily harmonics described above are used to produce dimensional values of T_{max}, T_{min}, and R, based on wet/dry status.

4. OBSERVED RELATIONSHIPS

In most implementations, WGEN-type models use fixed values of **A** and **B** (and therefore $\mathbf{M_0}$ and $\mathbf{M_1}$), as given by (Richarson 1982). Data records from the five stations used in this study show substantial deviations from Richardson's values of $\mathbf{M}_0$ (Fig. 1). Large differences in correlation occur between wet and dry days. Seasonally varying calculations show even larger differences in the magnitude – and, in some cases, the sign – of the correlations (Fig. 1). While the literature-based values of the correlation may be appropriate during some seasons at some locations, the observed and literature-based (constant) correlations are quite different when location and the entire calendar year are considered (Fig. 1). The elements of $\mathbf{M_1}$ are less variable than the elements of $\mathbf{M_0}$ and show smaller differences between wet and dry days. However, for some variables at some locations, the differences in $\mathbf{M_1}$ may be important. For example, the observed and literature-based lag-1 correlation between T_{max} and R for January at Indianapolis differ by >0.3.

The differences in the correlations between variables ultimately dictate the variability in the elements of the **A** and **B** matrices. For each location, at least one element of the **B** matrix shows substantial differences between wet and dry observations. However, these differences are typically small (~0.1) and the literature-based values provide a reasonable estimate in both cases (wet/dry). This preliminary analysis suggests that seasonal and spatial variability of elements within the **B** matrix is negligible. Elements of the **A** matrix, however, show large differences both seasonally and between wet and dry status (e.g., Fig. 2). All five of the stations used here exhibit similar ranges of variability, although the seasonal patterns differ substantially. Variability in the elements of the **A** matrix suggests that using station defined parameters may have a large impact on the generated data values.

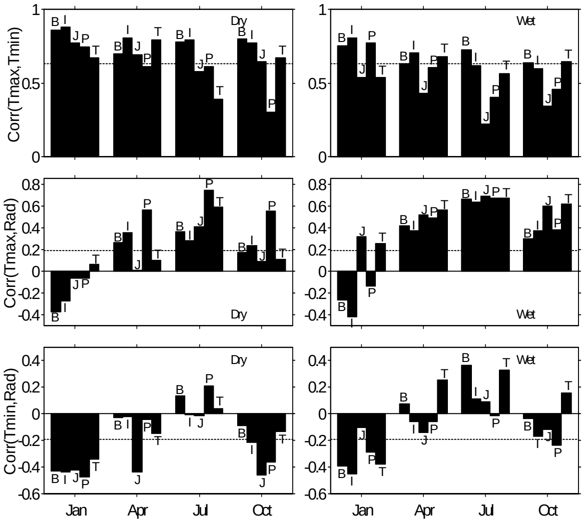

Fig 1. Observed lag-0 cross-correlations for wet and dry observations at Boston (B), Indianapolis (I), Jacksonville (J), Portland (P), and Tucson (T). The literature-based values are shown as dashed lines.

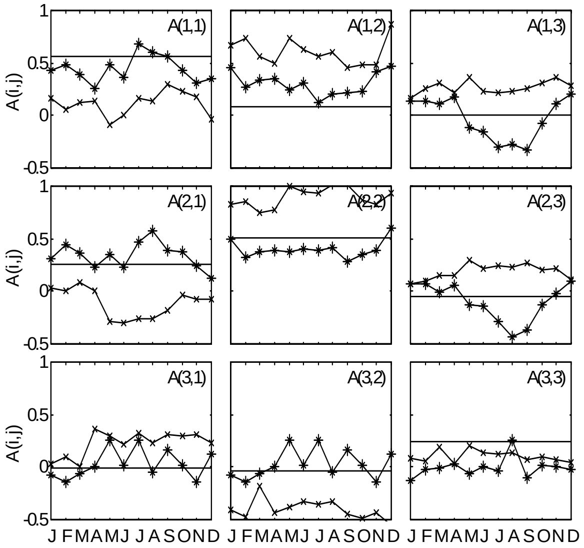

Fig 2. Elements of the **A** matrix for Indianapolis, IN showing seasonal variability and differences between wet (*) and dry (+) days. Literature-based values are shown with a solid black line.

5. WEATHER GENERATOR IMPLEMENTATION

To investigate the effects of the spatially and seasonally varying parameterizations described above, the weather generator described in Sec. 3 was used to generate 100-year sequences of daily T_{max}, T_{min}, and R for the stations described in Sec. 2. The weather generator was run in two modes. First, the elements of **A** and **B** were held constant according to the literature-based values (i.e., values given by Richardson, 1982). In the second mode, monthly values of **A** and **B** that depend on wet/dry status are computed using historical data from each individual station.

6. EVALUATION OF GENERATED DATA

6.1. Basic Monthly Statistics

Several traditional statistics were used to evaluate the simulated data. These included monthly means, medians, standard deviations, and maximum and minimum values of each of the generated variables. For both T_{max} and T_{min}, each of these statistics exhibited close agreement between both of the generated series and the observed data. However, for some locations, for some months, the upper tails of the generated T_{max} distributions are not in agreement with the observations (see Sec. 6.3). These results suggests that both generators successfully reproduce the probability distributions of the means and standard deviations of the temperature variables, but may fail to properly simulate extreme values during some months due to poorly modeled tails.

The generated R data show systematically overestimated monthly maximum values. As a result of describing R with a Normal distribution, both generators produce negative values, which are set to zero. Therefore, many months have a simulated minimum R value of 0. Several methods have been suggested for dealing with this problem, including the use of semi-empirical distributions (Semenov *et al.* 1998).

6.2. Correlations Between Generated Variables

The correlation structure between the generated variables is fundamentally dependent on the values of **A** and **B.** It was therefore expected that the station specific generator (with monthly parameterizations of **A** and **B** according to wet/dry status) would better replicate the observed correlations between variables. Computation of the lag-0 cross-correlations and lag-1 serial correlations confirmed that station-based, monthly parameterization of **A** and **B** results in a nearly perfect match between simulated and observed data. Data simulated with the constant, literature-based values for **A** and **B** resulted in large differences between observed and generated correlations (e.g., observed and generated lag-0 correlation between T_{min} and R at Portland, OR in June differ by ~0.4). The lag-1 serial correlations of the generated data are typically not as different as the lag-0 correlations, but substantial differences (>0.3) exist for the stations analyzed here.

6.3. Extreme Events and Persistence

The generated sequences were evaluated in terms of the mean number of days below and above certain thresholds, as well as the persistence of events below and above those thresholds. In general, the generated and observed series match well in terms of the number of extreme events. For three of the stations analyzed (Boston, MA, Indianapolis, IN, and Jacksonville, FL), the number of extremely warm events (T_{max}>35°C) was overestimated by both weather generators (Fig. 3). At Portland, the number of extreme warm events was slightly underestimated. At Tucson, there was good agreement between observed and generated occurrences of T_{max} >35°C. However, this is less of an extreme value in the warm climate of the Southwest.

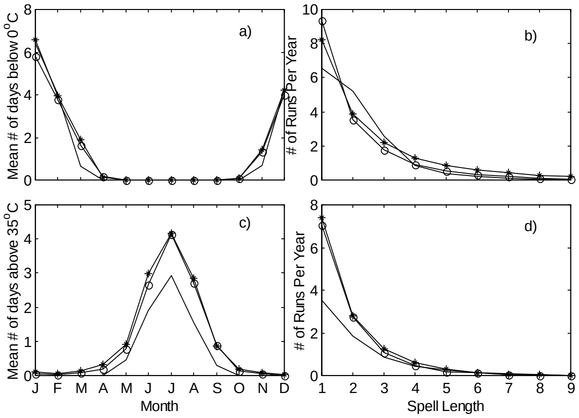

Fig 3. Evaluation of extreme events and persistence for Jacksonville, FL: a) Mean number of days with T_{min} below freezing, b) Mean number of spells of given length below freezing, c) Mean number of days with T_{max} above 35°C, and d) Mean number of spells of given length with T_{max} above 35°C. Literature-based data is symbolized by an 'o', station-specific data by a '*', and observed data by a solid line.

6.4. Diurnal Temperature Range (DTR)

As mentioned above, diurnal temperature range (DTR) was used as an evaluation tool in this research. Since DTR is dependent on both T_{max} and T_{min}, its proper simulation requires that the relationship between these two variables be preserved. For example, the literature-based generator (with constant **A** and **B**) produces many more days with negative DTR (i.e., a fundamental simulation error) than the station-specific generator at three of the locations used in this study (see Table 1). While both weather generators (literature and station-specific) simulated monthly means and standard deviations of generated variables well, DTR (and, therefore, the relationships between variables) is not simulated as well by the literature-based generator (Fig. 4). Although monthly mean DTR is similar in both models, the station-based generator achieves better agreement between observed and simulated standard deviation of DTR (Fig. 4).

Table 1. Average annual number of days with DTR<0 in a 100-year simulation

Station	# of days with DTR<0 (literature-based **A**, **B)**	# of days with DTR<0 (station-specific **A**, **B)**
Boston	9.78	4.47
Indianapolis	8.52	2.80
Jacksonville	0.69	0.86
Portland	3.73	1.89
Tucson	0.11	0.20

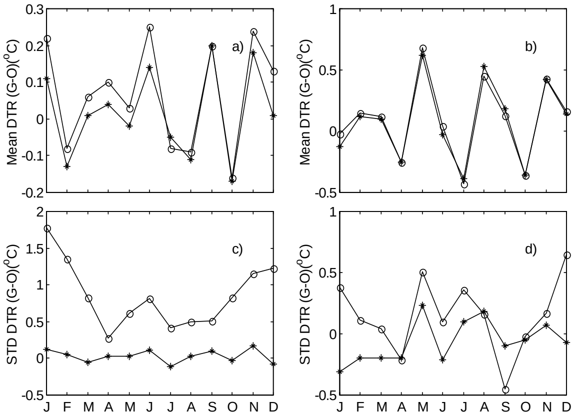

Fig 4. Monthly means and standard deviations of diurnal temperature range (DTR) expressed as the difference between generated and observed values at Boston, MA (literature-based data is symbolized by 'o', station-specific data is symbolized by '*'): a) means for dry observations, b) means for wet observations, c) standard deviations for dry observations, and d) standard deviations for wet observations.

The simulation of the standard deviation of DTR varies considerably between stations. For example, on dry days, the literature-based generator overestimated the standard deviation at two stations (Boston, MA and Indianapolis, IN), underestimated the standard deviation at one station (Portland, OR), and had a seasonally varying pattern at the remaining two stations (Jacksonville, FL and Tucson, AR). Similar relationships exist for the wet observations. The standard deviations of DTR are closely related to the persistence of DTR. For those stations that overestimate the standard deviation of DTR, a higher degree of persistence (as seen in the lag-1 serial correlation of DTR) is observed. This persistence is directly related to the values in the **A** matrix. Because **A** is multiplied by the previous days values in the generating equation (Eq. 1), the persistence in the generated data is dependent on the magnitudes of the elements of **A**. Examination of the **A** matrices for the five stations in this study confirms that larger **A** values are associated with greater persistence (and hence a larger standard deviation) of DTR.

The station-specific generator also reproduces the relationships between temperature and radiation more accurately. Because DTR is closely linked to cloud cover and precipitation (Leathers et al., 1998), and radiation is a reasonable surrogate for cloud cover, allowing the relationships between temperature and radiation to vary by location and time of year helps to improve the simulation of temporal variability in DTR.

7. DISCUSSION AND CONCLUSIONS

In this study, the effects of stochastic weather generator parameterizations have been investigated. Using historical data from five stations in the United States, we examined the spatial and seasonal differences in the lag-0 cross-correlations and lag-1 serial correlations between T_{max}, T_{min}, and R. These correlations ultimately determine the nature of the **A** and **B** matrices used in the stochastic weather generator, and were found to vary both spatially and seasonally.

To investigate the impacts of the seasonal and spatial variability in the elements of these matrices, 100-year simulations for the five stations were undertaken with 1) **A** and **B** assumed constant according to the literature-based values and 2) **A** and **B** computed for each individual station on a monthly basis. The second implementation also allowed **A** and **B** to vary according to wet/dry status.

The simulations were compared to observed data using statistical and graphical methods. The results suggested that monthly means and standard deviations of each simulated variable agree with observed values for both simulations. The two implementations of the weather generator exhibited only small differences in monthly maximum and minimum values. However, the literature-based generator failed to preserve relationships between variables. This shortcoming is evident in both the diurnal temperature range (DTR) and in the correlations between simulated variables.

These results suggest that literature-based values may be appropriate for applications where monthly values of the means and standard deviations of generated variables are of interest. For applications that require proper simulation of relationships between variables, station-specific parameterizations are recommended.

8. REFERENCES CITED

Hayhoe, H. N. (1998). Relationships between weather variables in observed and WXGEN generated data series. *Agricultural and Forest Meteorology* **90**: 203-214.

Leathers, D. J., Palecki, M. A., Robinson, D. A. and Dewey, K. F. (1998). Climatology of the daily temperature range annual cycle in the United States. *Climate Research* **9**: 197-211.

Matalas, N. C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Research* **3**(4): 937-945.

Richardson, C. W. and Wright, D. A. (1984). WGEN: A model for generating daily weather variables, USDA-ARS**:** 83.

Richardson, C. W. (1982). Dependence structure of daily temperature and solar radiation. *Transactions of the ASAE* **25**(3): 735-739.

Schubert, S. (1994) A weather generator based on the European 'Grosswetterlagen'. *Climate Research* **4**, 191-202.

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

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.