

Research Article

Stochastic Models to Generate Geospatial-, Temporal-, and Cross-Correlated Daily Maximum and Minimum Temperatures

Guillermo A. Baigorria

School of Natural Resources and Department of Agriculture & Horticulture, University of Nebraska-Lincoln, 823 Hardin Hall, 3310 Holdrege Street, Lincoln, NE 68583-0968, USA

Correspondence should be addressed to Guillermo A. Baigorria; gbaigorria@unl.edu

Received 17 October 2013; Revised 22 December 2013; Accepted 23 December 2013; Published 18 February 2014

Academic Editor: Julio Diaz

Copyright © 2014 Guillermo A. Baigorria. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Weather generators are tools used to downscale monthly to seasonal climate forecasts, from numerical climate models to daily values for use as inputs for crop and other environmental models. One main limitation of most of weather generators is that they do not incorporate neither the spatial/temporal correlations between/within sites nor the cross-correlations between variables, characteristics specially important when aggregating, for example, simulated crop yields, freeze events, or heat waves in a watershed or region. Three models were developed to generate realization of daily maximum and minimum temperatures for multiple sites. The first model incorporates only spatial correlation, whereas temporal correlation using a 1-day lag and cross-correlation between variables were added to model one, respectively, by the other two models. Vectors of correlated random numbers were rescaled to temperature values by multiplying each element with the standard deviation and adding the mean of the corresponding weather station. An extension of Crout's algorithm was developed to enable the factorization of nonpositive definite matrices. Monthly spatial correlations of generated daily maximum and minimum temperatures between all pairs of weather stations closely matched their observed counterparts. Performance was analyzed by comparing the root mean squared error, temporal semivariograms, correlation/cross-correlation matrices, multiannual monthly means, and standard deviations.

1. Introduction

Because of the geospatially homogeneous nature of daily temperatures, few publications have investigated the generation of daily temperatures while reproducing the observed spatial variability [1, 2]. Most of these methods [3, 4], including those presented in the current paper, are based on Wilks' parametric model [5]. Another common approach found in the literature is generally based on resampling methods [6–8]. The applicability of Wilks' methods is potentially limited if the correlation matrix cannot be factorized, such as the case when a matrix is nonpositive definite. The probability of facing nonpositive definite matrix increases exponentially for large number of locations or the use of more complex matrices, which includes lag time and cross-correlations.

Thus, the methodological requirements to factorize matrices cannot be overlooked.

When regional analysis is required, it is important to generate geospatially and temporally correlated maximum and minimum temperatures while taking into account the covariance between them and between other variables, such as rainfall [9, 10]. For instance, a water balance for a given region must be based on multisite estimations of daily evapotranspiration based on temperatures (among other variables) following the observed spatial structure and its relationship with rainfall occurrence [11]. Failing to correctly generate daily differences between maximum and minimum temperatures will cause over- or underestimation of water availability in a region [12]. Inaccuracies can increase even more, if multisite estimations of daily evapotranspiration use

temperature-based methods to estimate solar radiation [13, 14].

Over the last decade, there has been increasing use of weather generators to stochastically downscale the increasingly skilled monthly forecasts/projections from global and regional circulation models (G/RCMs) to forecast seasonal climate and to project climate change [15, 16]. Together with other available methods, such as the delta change method [17] and bootstrapping [18], weather generators are necessary to apply because the daily outputs from G/RCMs are mean values within the geographical domain of a grid cell, and the direct use of these daily outputs in environmental and agricultural models has shown large bias and errors in the final outcomes [19, 20].

The objectives of this study are as follows: (i) to develop a stochastic daily maximum and minimum temperature generator capable of reproducing both the daily spatial correlation between weather stations and the monthly statistics of each individual weather station, accounting for temporal correlation from a previous day and/or cross-correlation between maximum and minimum temperatures, characteristics that current weather generators are not capable to reproduce and (ii) to extend Crout’s algorithm to transform a non-positive definite correlation matrix into a positive definite matrix, thus allowing its factorization.

Several questions have arisen while considering how to incorporate spatial correlations in generating maximum and minimum temperatures. For example, what is the best way to factorize a nonpositive definite correlation matrix? Are observed monthly spatial correlation patterns accurately reproduced by generating daily spatially correlated maximum and minimum temperatures? Is it possible to replicate the daily spatial covariance between maximum temperatures and minimum temperatures? The models developed in this paper generate daily maximum and minimum temperatures for weather station locations rather than continuous temperatures over space. Weather data generated by a single-site weather generator or bootstrapping methods should not be spatially interpolated because of the assumption of independence among stations; however, data generated by the algorithms developed in this paper can be spatially interpolated. For the same reason, outcomes produced by crop, hydrological, and environmental models using weather data generated without taking into account its spatial structure should not be spatially aggregated.

2. Methods

The present approach is based on the assumption of spatial-temporal covariance stationarity, which implies that the autocovariance functions of the data series, the spatial correlations between the data series, and the cross-correlation between variables of the data series do not change during the time period considered. From a temporal point of view, this allows one to characterize a time series of a given variable, like temperatures, as samples from a probability distribution at each weather station. It also allows one to characterize spatial relationships between all pairs of weather stations

TABLE 1: Differences among models according to their properties.

Properties	Model		
	First	Second	Third
Covariance between maximum temperature and rainfall occurrence	Yes	Yes	Yes
Temperature generated on previous day	No	Yes	No
Cross-correlation between minimum and maximum temperatures	No	No	Yes

in a selected area with a correlation and cross-correlation matrices. This assumption allows one to perturb the main statistics of the geospatial weather generator for climate variability and climate change applications. The assumption of spatial-temporal covariance stationarity applies when the global and regional atmospheric circulation patterns across the study area remain unchanged within the range of the interannual variability. In some cases, extreme events like ENSO for instance, can significantly change these spatial and temporal patterns [21, 22]. Under these circumstances, it is recommended to split the dataset accordingly and to apply the assumption to each of the new categories (i.e., El Niño, La Niña, and neutral years). Applying the present approach to future climate projections must use the same rationale. Because future climate projections are generated using global and regional circulation models, these analyses are possible to perform.

Daily historical records of maximum and minimum temperatures and rainfall are needed for the model described in this paper because it uses statistics of observed weather data within and between locations for autocalibration. Generation of daily rainfall events and amounts used in the present paper follows the method described by Baigorria and Jones [23].

2.1. Models. Three models for generating spatially and temporally correlated maximum and minimum temperatures were developed (Table 1). The covariance between maximum temperature and rainfall occurrences was taken into account in all three models [24]. To account for this covariance, two Gaussian distributions were fitted for maximum temperatures for each location: (a) using maximum temperatures during rainy days only and (b) using maximum temperatures during nonrainy days.

The first and second models generate daily maximum and minimum temperatures independently, whereas the third model generates both variables at the same time based on the spatial correlation of the same variable and the cross-correlation between variables. The first and third models generate daily temperatures without taking into account the previous day’s generated values; the second model does take into account previously generated values.

The rationale of developing, evaluating, and comparing three models instead of a unified model is to emphasize the advantages and disadvantages of spatial, temporal, and cross-correlations approaches alone. This facilitates the understanding of each step of each model, as well as the uncertainties and errors of each approach alone. This makes

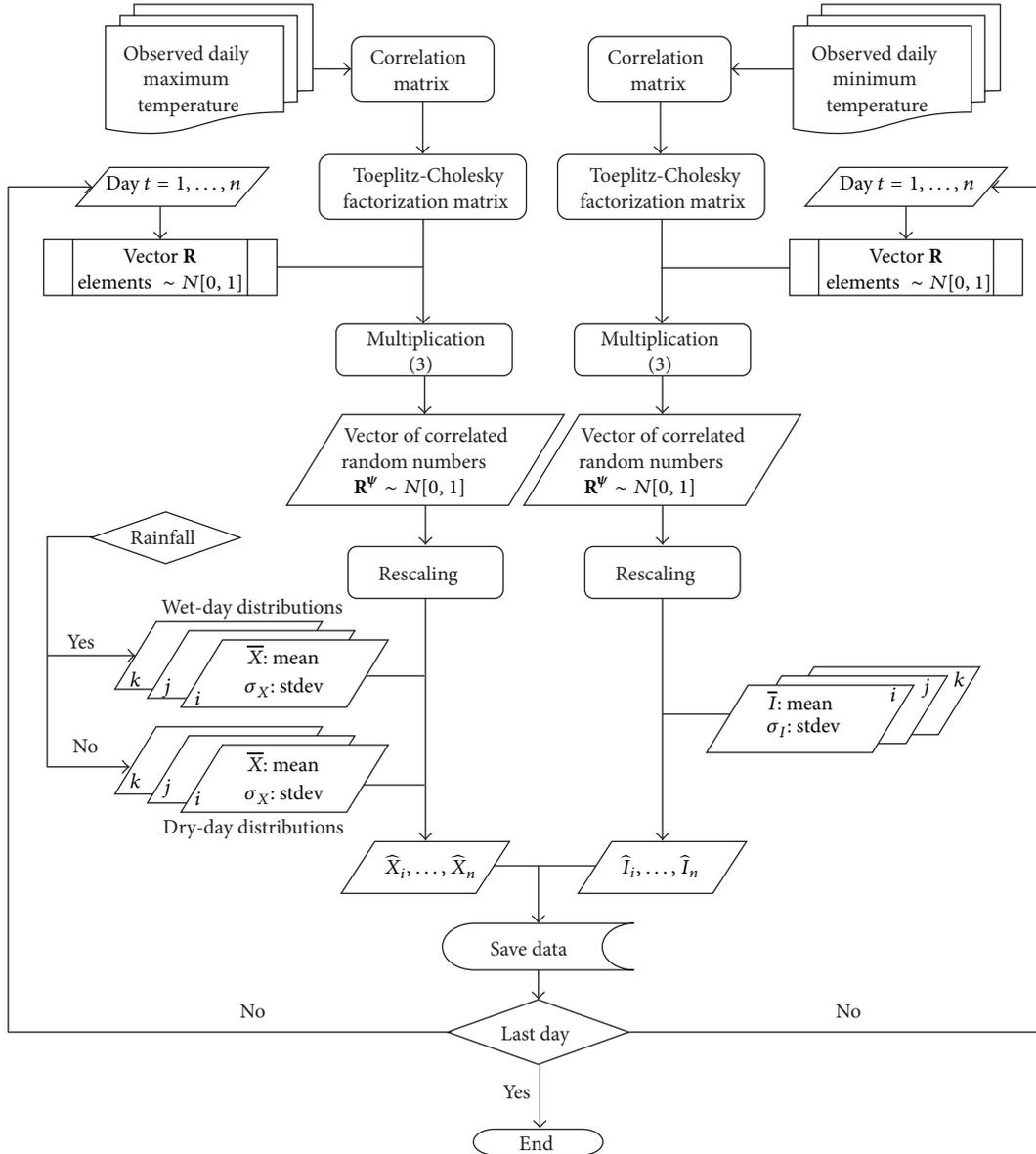


FIGURE 1: Flow diagram detailing the sequence of processes to generate maximum and minimum temperatures following the first model.

it possible for the readers to implement their own methods through the combination of the proposed models according to their own necessities.

The proposed models to generate maximum and minimum temperatures, together with previously published methods for generating rainfall [23], were implemented in a user-friendly software named Geo Spatial and Temporal weather generator (GiST-wg). The software can be downloaded at <http://www.agenviromodeling.com>.

2.1.1. First Model: Model with Spatial Correlation. The first model generates maximum temperatures independently of minimum temperatures (Figure 1). This model only takes into account the spatial correlations between weather stations; it

does not take into account temperatures generated on the previous day (Table 1).

Using historical data from all weather stations in the selected area, Pearson's correlation coefficients (ρ_{ij}), which are the elements of the correlation matrix, were calculated using the following equation (see Table 2 for variable definitions):

$$\rho_{ij} = \frac{1}{\eta} \frac{\sum_{t=1}^{\eta} (T_{it} - \bar{T}_i)(T_{jt} - \bar{T}_j)}{\sigma_i \sigma_j}. \quad (1)$$

According to the results of the Filliben Q-Q correlation test for Gaussian distribution (a computationally simple variant of the Shapiro-Wilk test [25]) (results not shown), the daily variability of maximum and minimum temperatures tended to follow a Gaussian distribution $\sim N[\bar{T}, \sigma]$ in our

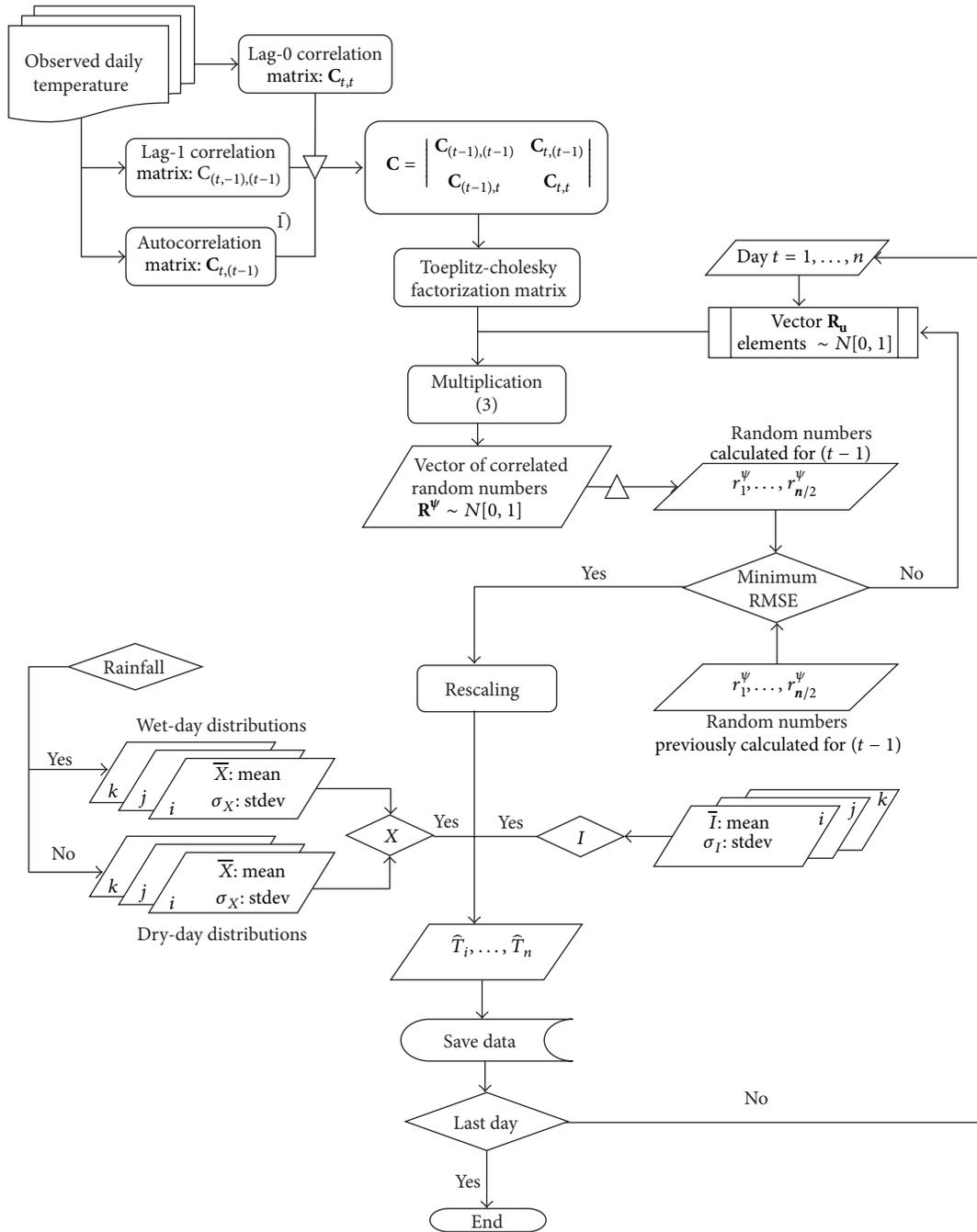


FIGURE 2: Flow diagram detailing the sequence of processes to generate maximum and minimum temperatures following the second model.

study regions. (Null hypotheses that the observed data were drawn from a fitted hypothetical Gaussian distribution were not rejected at levels ranging from 0.5% to 10% depending on month and weather station.) This is a requirement of the method [26]. If the data is not normally distributed, a previous transformation is recommended.

Taking into account the pairwise correlations between locations, a vector of independent standard Gaussian random numbers (\mathbf{R}) is transformed into a vector (\mathbf{R}^ψ). The elements

of \mathbf{R}^ψ are correlated random numbers (r^ψ) that follow the observed correlations between stations [27]. To obtain \mathbf{R}^ψ , a vector \mathbf{R} , with elements $r \sim N[0, 1]$ and with length equal to the number of locations, is matrix-multiplied by the Toeplitz-Cholesky factorization matrix (\mathbf{F} ; [28–31]), which is calculated, according to the case, based on the correlation matrix of maximum (\mathbf{C}_X) or minimum (\mathbf{C}_I) temperatures (An alternative to the Toeplitz-Cholesky factorization matrix is the eigendecomposition matrix (\mathbf{E}) [32, 33]. According to

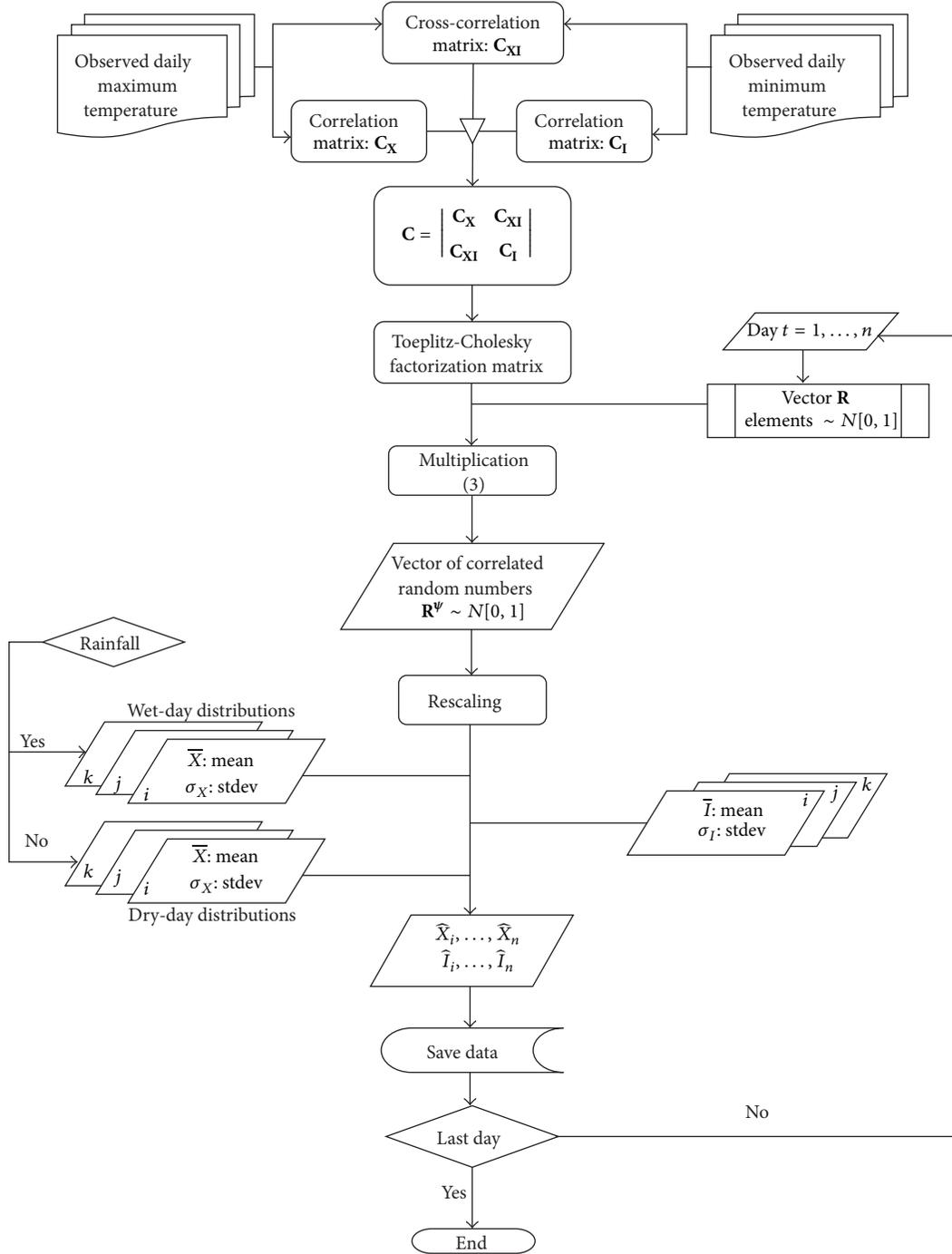


FIGURE 3: Flow diagram detailing the sequence of processes to generate maximum and minimum temperatures following the third model.

our experience, both methods perform equally well.) (see Section 2.2).

$$\mathbf{R}\mathbf{F} = \mathbf{R}^{\Psi}. \quad (2)$$

Afterward, the elements in \mathbf{R}^{Ψ} are rescaled according to the Gaussian distributed percentages of the population being included (e.g., from -1 to 1 for 68% or -1.96 to 1.96 for 95%). Equation (3) is then used to generate correlated

maximum temperatures (\widehat{X}) and minimum temperatures (\widehat{I}) using previously generated vector of correlated random numbers from a Gaussian distribution (r^{Ψ}) at each location (the variables are defined in Table 2):

$$\widehat{T}_{i_t} = \overline{T}_i + \sigma_i r_t^{\Psi}. \quad (3)$$

After the first set of maximum or minimum temperatures is generated for the first day of the time series, a new \mathbf{R} is then produced for each new generated day. Because \mathbf{C} is calculated

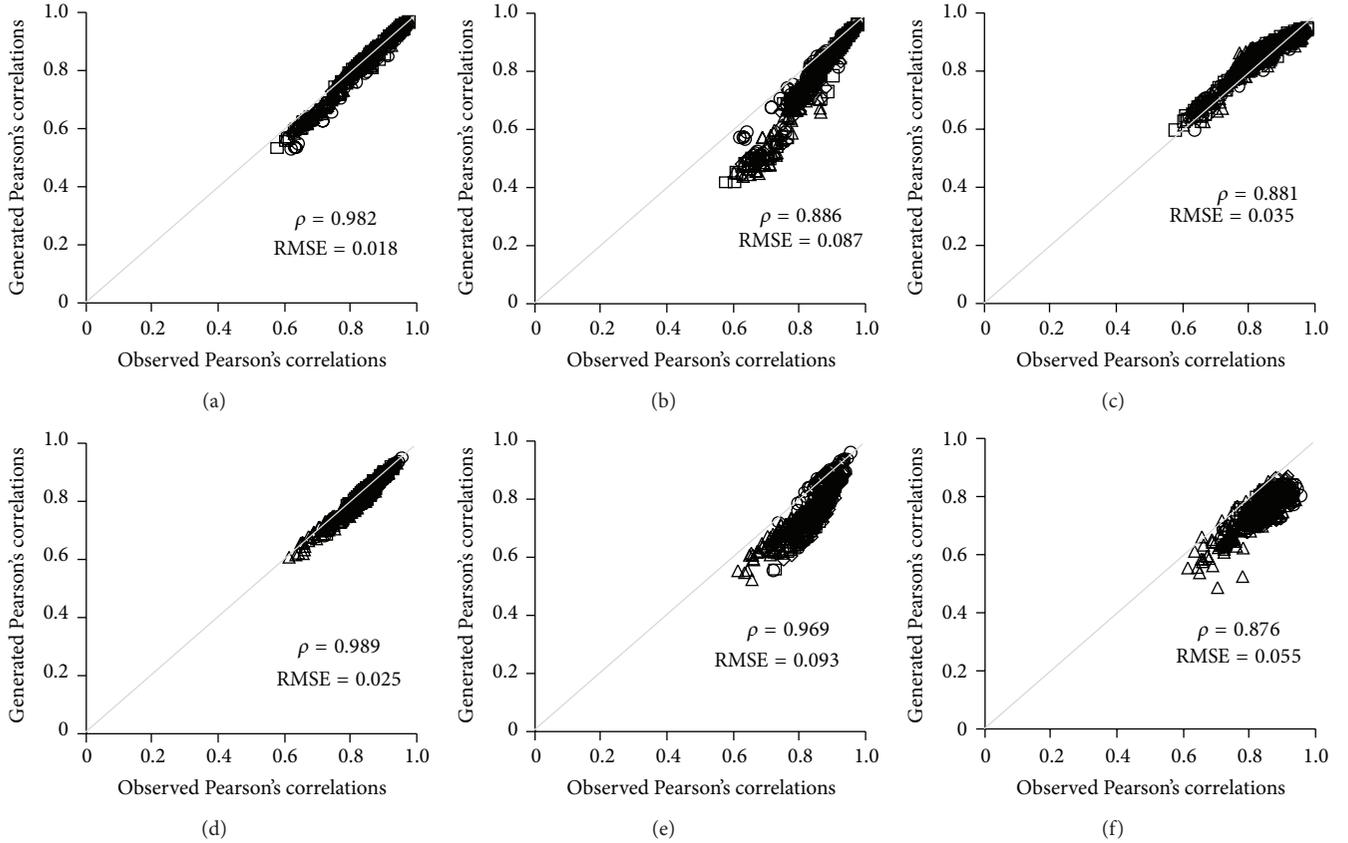


FIGURE 4: Comparison between observed and generated correlations of daily maximum temperatures ((a), (b), (c)) and minimum temperatures ((d), (e), (f)) among all pairs of weather stations for each month. ((a), (d)) first model, ((b), (e)) second model, and ((c), (f)) third model. (O) December-January-February, (□) March-April-May, (Δ) June-July-August, and (◇) September-October-November.

for each month, \mathbf{F} only needs to be calculated once per month, and it is the same for all the generated years.

2.1.2. Second Model: Model with Spatial and Temporal Correlation. As with the previous model, the second model generates maximum temperatures independently of the minimum temperatures. The main difference is that this model takes into account the spatial correlations between weather stations during the day being generated (t) and the previous day ($t-1$) (Table 1; Figure 2).

Taking into account for both days: (a) the pairwise correlations between locations and (b) the pairwise 1-day lag autocorrelations between locations, a vector of independent standard Gaussian random numbers (\mathbf{R}) with length equal to twice the number of locations is transformed into a vector (\mathbf{R}^Ψ) whose elements are correlated random numbers (r^Ψ) that follow the observed correlations and the autocorrelations between stations. To obtain \mathbf{R}^Ψ , a vector \mathbf{R} of random numbers ($r \sim N[0, 1]$) is matrix-multiplied by the Toeplitz-Cholesky factorization matrix (\mathbf{F}) calculated based on the correlation matrix of maximum (\mathbf{C}_X) or minimum (\mathbf{C}_I) temperature. In this model, however, the correlation matrix is composed of two correlation matrices: the 0-day lag ($\mathbf{C}_{t,t}$) and the 1-day lag ($\mathbf{C}_{(t-1),(t-1)}$), and two autocorrelation matrices:

$\mathbf{C}_{t,(t-1)}$ and $\mathbf{C}_{(t-1),t}$, as shown in (4), where $\mathbf{C}_{t,t} = \mathbf{C}_{(t-1),(t-1)}$ and $\mathbf{C}_{t,(t-1)} \approx \mathbf{C}_{(t-1),t}$

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{(t-1),(t-1)} & \mathbf{C}_{t,(t-1)} \\ \mathbf{C}_{(t-1),t} & \mathbf{C}_{t,t} \end{bmatrix}. \quad (4)$$

Elements in \mathbf{R}^Ψ are then rescaled according to the Gaussian distributed percentages of the population being included. Using (3), elements of this vector are used to finally obtain \hat{T} for all locations for two consecutive days. By producing two days at a time, the 1-day lag autocorrelation is considered in the process. However, if we proceed in the same manner to produce days 3 and 4 of the time series, an inconsistency between days 2 and 3 will be created. Then, only one day at a time can be added to the time series. Because the temperature (maximum or minimum) has already been generated for day 2, it is impossible to numerically solve the set of equations because half of the correlated random numbers resulting from the multiplication are already known from the initial calculation (days 1 and 2). To solve this set of equations, an iterative process is performed. Here, a given number of \mathbf{R}^Ψ are generated, forming an ensemble. Half of the vector's elements of each \mathbf{R}^Ψ are used to generate day 2 ($r_1^\Psi, \dots, r_n^\Psi$), and the other half are used to generate day

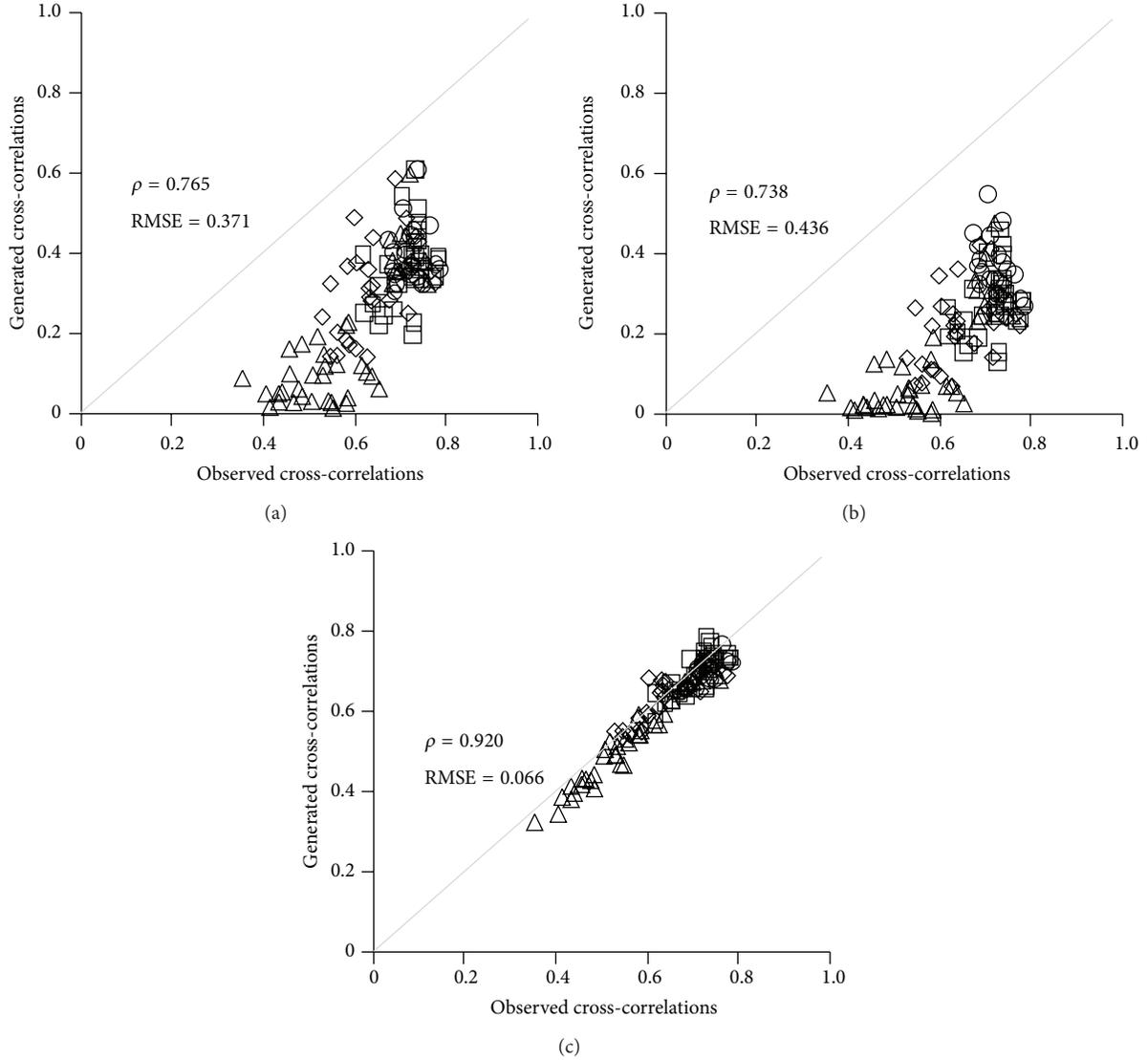


FIGURE 5: Comparison between observed and generated cross-correlations between daily maximum and minimum temperatures for all weather stations and for each month. (a) First model, (b) second model, and (c) third model. (O) December-January-February, (□) March-April-May, (Δ) June-July-August, and (◇) September-October-November.

3 ($r_{(n+1)}^\psi, \dots, r_{2n}^\psi$). Those elements r^ψ corresponding to day 2 are then compared to their counterparts previously calculated when generating days 1 and 2. For the comparison, the root mean square error (RMSE) is calculated. After the RMSEs are calculated for every ensemble member, the one that produces the lower RMSE is selected. This \mathbf{R}^Ψ is then rescaled and used in (2) to generate the temperatures of day 3 for all locations. The same procedure is repeated to generate the complete required time series.

2.1.3. Third Model: Model with Spatial and Cross-Correlation. The third model generates maximum and minimum temperatures simultaneously, taking into account the existing cross-correlations between these variables (Figure 3). This model takes into account the spatial correlations between weather

stations, but not temperatures generated the previous day (Table 1).

Using historical data from all weather stations in the selected area, Pearson's correlation coefficients are calculated using (1). The cross-correlations between maximum and minimum temperatures (φ_{ij}), which are the elements of the cross-correlation matrix, are calculated using:

$$\varphi_{ij} = \frac{1}{\eta} \frac{\sum_{t=1}^{\eta} (X_{it} - \bar{X}_i)(I_{jt} - \bar{I}_j)}{\sigma_{X_i} \sigma_{I_j}}. \quad (5)$$

As in the second model, two sets of temperatures are calculated simultaneously. In this model, these sets correspond to the maximum and minimum temperatures of the same

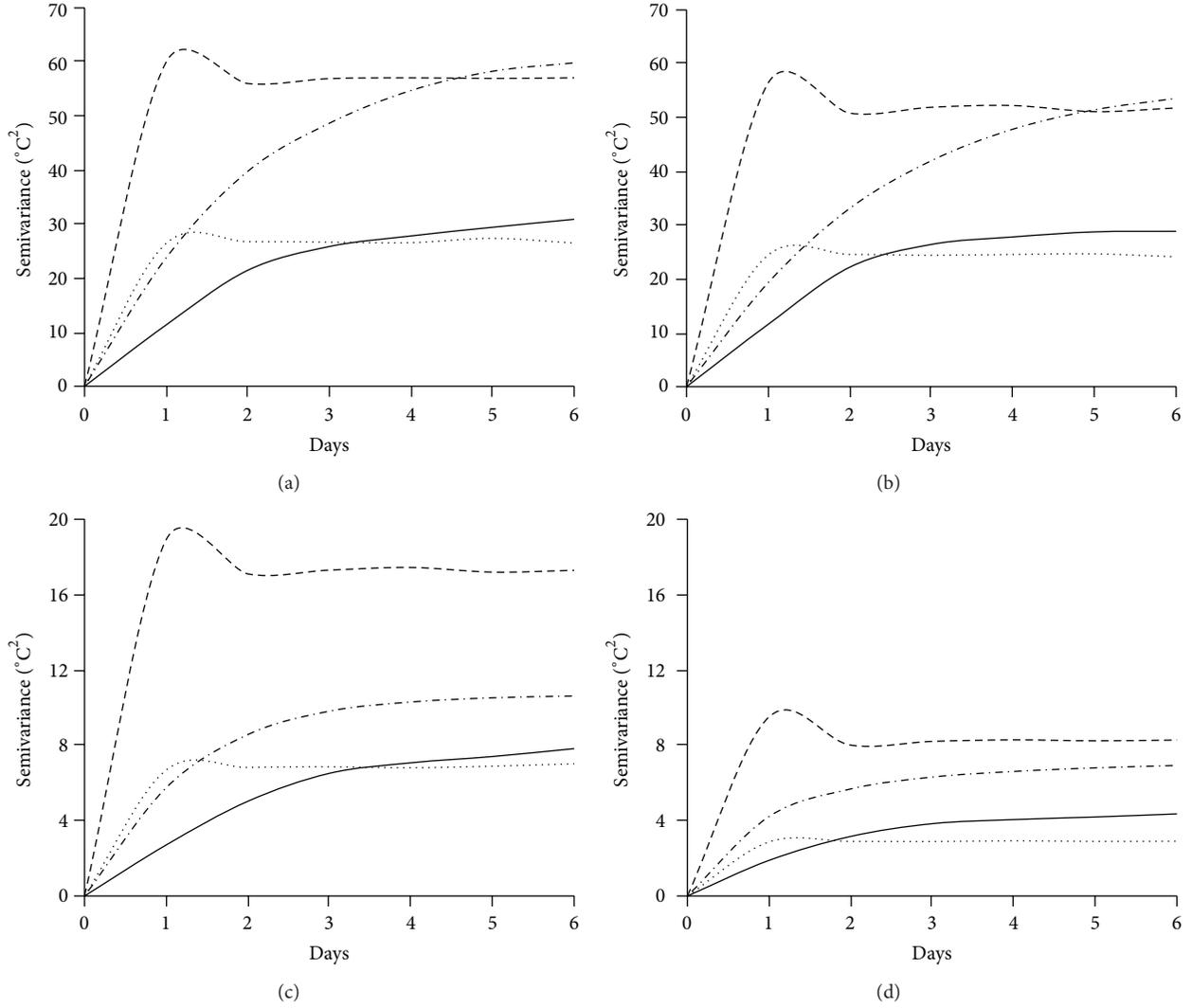


FIGURE 6: Comparison between observed and generated temporal semivariograms of ((a), (c)) maximum and ((b), (d)) minimum temperatures for the months of January ((a), (b)) and July ((c), (d)). Weather station: Asheboro, Randolph, NC. Observed (solid line), generated by first model (dash line), second model (dots line), and third model (dash and dots line).

day. The correlation matrix in this case is composed of two correlation matrices and the cross-correlation matrix (6):

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_X & \mathbf{C}_{XI} \\ \mathbf{C}_{IX} & \mathbf{C}_I \end{bmatrix}. \quad (6)$$

Similar to the second model, a vector of independent standard Gaussian random numbers (\mathbf{R}), with length equal to twice the number of locations, is transformed into a vector (\mathbf{R}^Ψ) whose elements are correlated random numbers (r^Ψ) that follow the observed correlations between stations and cross-correlations between maximum and minimum temperatures within the same station and between stations. To obtain \mathbf{R}^Ψ , a vector of random numbers ($r \sim N[0, 1]$) is matrix-multiplied by the Toeplitz-Cholesky factorization matrix (\mathbf{F}) calculated based on the newly composed correlation matrix produced by (6).

The elements in \mathbf{R}^Ψ are then rescaled according to the Gaussian distributed percentages of the population being

included. Using (3), the elements of this vector are used to finally obtain \tilde{X} and \tilde{I} for all locations.

2.2. Factorization of the Correlation Matrix. To factorize a correlation matrix using Toeplitz-Cholesky factorization or eigendecomposition, the matrix must be a positive-definite matrix. Thus, the resulting upper triangular matrix after applying an LU-decomposition must strictly contain only positive diagonal entries. When this is the case, the factorization is a straightforward process.

When the size of the correlation matrix increases (due to a large number of weather stations or the use of more complex matrices) and/or the correlation values within the matrix are homogeneous (which is the case for temperatures), the correlation matrix may become nonpositive definite. To solve this problem, Rebonato and Jäckel [34] proposed two iterative methods: (a) the hyperspheric decomposition and

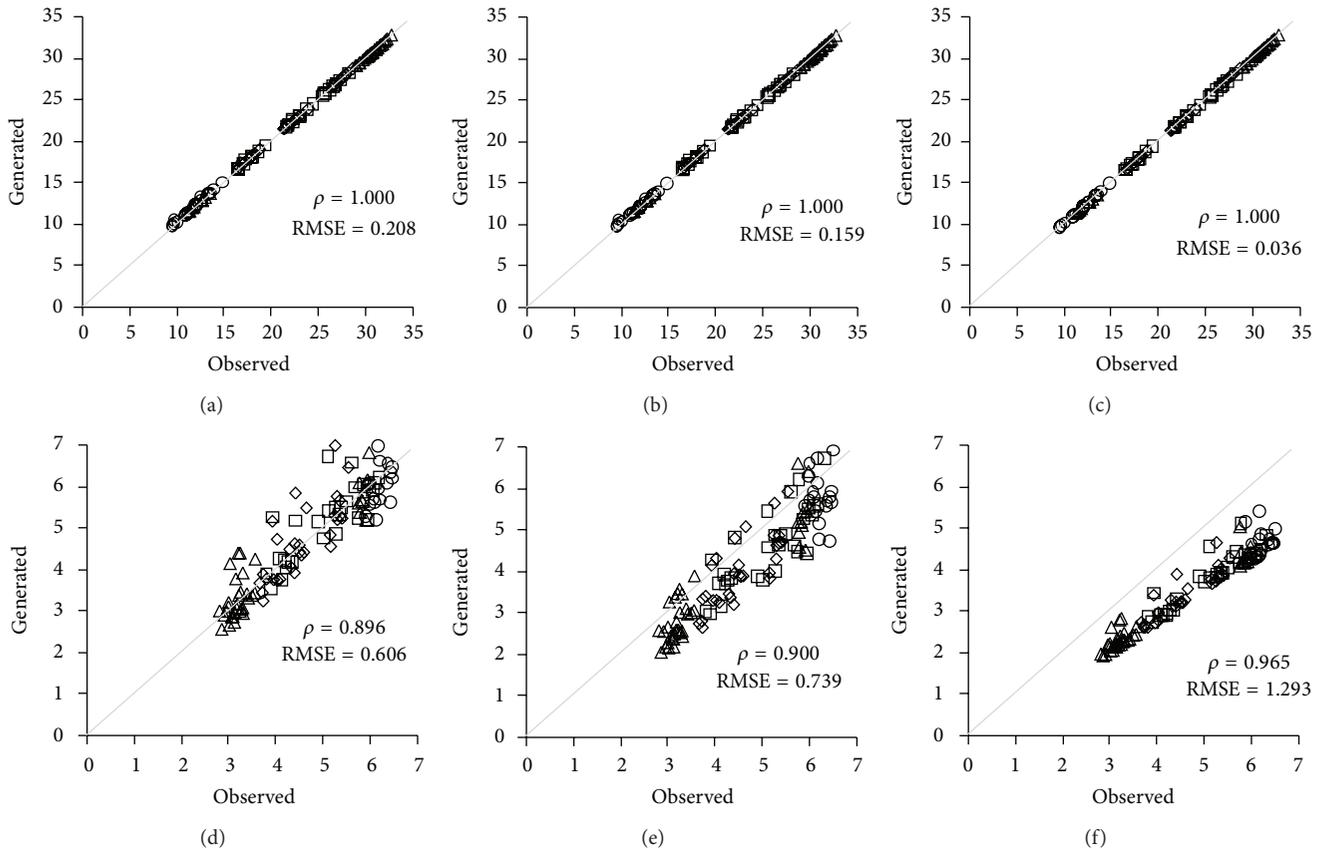


FIGURE 7: Comparison between observed versus generated ((a), (b), (c)) mean maximum temperatures ($^{\circ}\text{C}$) and ((d), (e), (f)) standard deviations of maximum temperatures ($^{\circ}\text{C}$) for each month. ((a), (d)) first model, ((b), (e)) second model, and ((c), (f)) third model. (O) December-January-February, (\square) March-April-May, (Δ) June-July-August, and (\diamond) September-October-November.

(b) the spectral decomposition. Both methods are based on redefining the correlation matrix prior to the factorization to guarantee a positive-definite matrix. These time-consuming methods are adequate for relatively small matrices with heterogeneous correlation values within the matrix. For applications with a large number of variables, these methods are not guaranteed to produce a positive-definite matrix.

The solution proposed and tested in this paper is to apply an LU decomposition to the correlation matrix using an extension of Crout's algorithm [35, 36] (see Figure 9). This algorithm produces a total of n^2 equations for the $(n^2 + n)$ unknown values of the **L**- and **U**-triangular matrices (the diagonal being represented twice). Because the number of unknowns is greater than the number of equations, the standard procedure is to set the n of the unknowns arbitrarily to one and then to try to solve for the others. For this application, the procedure is repeated iteratively by slightly increasing the arbitrary values (steps of 0.01 in our case) until all the diagonal values of the **U**-triangular matrix are positive (see the Appendix). After reaching this condition, the **L**- and **U**-triangular matrices are multiplied to produce the new correlation matrix. This new matrix guarantees the positive-definite matrix necessary for the factorization.

2.3. Case Study

2.3.1. Data. The 10 selected weather stations (Table 3) used in this study are located in the state of North Carolina between $36^{\circ}35'\text{N}$, $84^{\circ}19'\text{W}$ and $33^{\circ}50'\text{N}$, $75^{\circ}28'\text{W}$. The daily observed data of maximum and minimum temperatures and rainfall from these weather stations were obtained from the NOAA/NWS/National Climate Data Center (<http://nndc.noaa.gov/?home.shtml/>). The distance between weather stations in North Carolina ranges from approximately 34 to 164 km. The period from 1974 to 2004 was selected to avoid the effects of temporal climatic shifts detected over time [9, 37].

2.3.2. Evaluation. As stated by Ferraris et al. [38], referring to the evaluation of downscaling models, the stochastic geospatial weather generator models presented in the current paper must be able to reproduce the statistics on which they have been optimized (observed spatial-, temporal- and/or cross-correlations; monthly mean and standard deviations), as well as being able to reproduce other statistics of the data not used in model development (e.g., RMSE, temporal semi-variograms). To satisfy both requirements and to compare the three previously described models, one thousand year-long

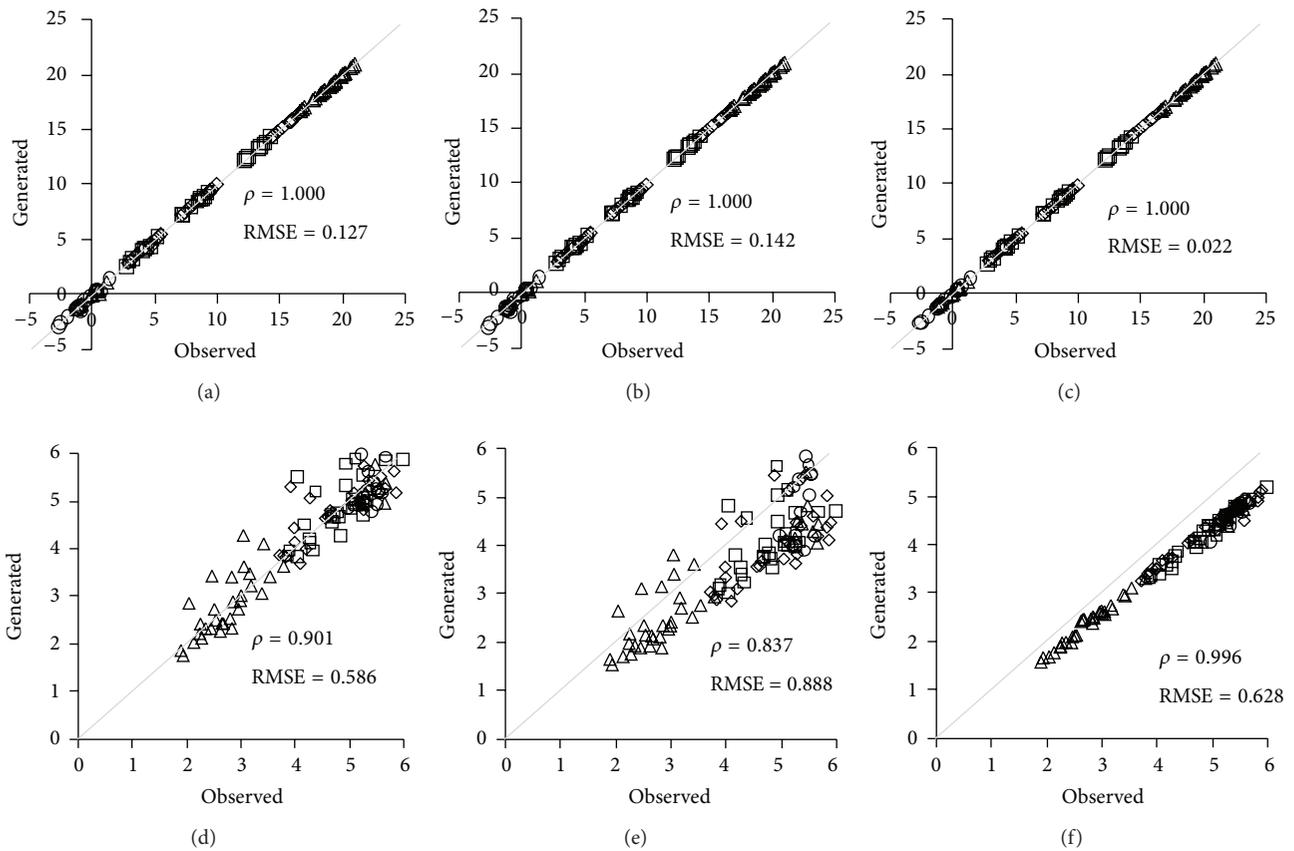


FIGURE 8: Comparison between observed versus generated ((a), (b), (c)) mean minimum temperatures (°C) and ((d), (e), (f)) standard deviations of minimum temperatures (°C) for each month. ((a), (d)) First model, ((b), (e)) second model, and ((c), (f)) third model. (○) December-January-February, (□) March-April-May, (△) June-July-August, and (◇) September-October-November.

replications of daily maximum and minimum temperatures, along with rainfall events and amounts, were generated for the 10 selected weather stations. Correlation coefficients and root mean square errors (RMSEs), between the observed and the generated daily maximum temperature, were compared among the three models. The calculations were performed using the monthly correlations between all pairs of locations. The same comparison was performed for daily minimum temperatures. Correlation coefficients and root mean square errors (RMSEs) calculated based on the monthly correlations between all pairs of locations calculated from the observed and generated daily maximum (minimum) temperatures were compared among the three models. RMSEs were used to represent differences between spatial correlations of the historical records and spatial correlations of the generated temperatures. Cross-correlations between daily values of maximum and minimum temperatures for each month were calculated from the generated data and compared with their observed counterparts. Observed and generated multiannual monthly means of daily maximum and daily minimum temperatures were compared. For every month, a comparison between observed and generated interannual variance of daily maximum and minimum temperatures was performed

by comparing standard deviations. The correlation coefficients and RMSEs were calculated from the mean and the standard deviations of the three models. Temporal semi-variograms calculated from generated daily maximum and minimum temperatures of the three models were compared with the corresponding semi-variograms calculated from the historical record.

3. Results and Discussion

3.1. Generation of Maximum and Minimum Temperatures. Figure 4 shows the comparisons between observed and generated correlations of daily maximum and minimum temperatures among all pairs of locations for each month. The three models reproduced the monthly observed correlations of maximum and minimum temperatures with correlation coefficients ranging from 0.876 to 0.989. For maximum temperatures, RMSE values (as a percentage) relative to the mean observed correlation among all locations were 2%, 10%, and 4% for the three models, respectively; for the minimum temperatures, relative RMSEs were 3%, 11%, and 7%, respectively.

TABLE 2: Variable description.

Symbol	Definition	Units
C_X, C_I	Correlation matrices of maximum and minimum temperatures	Unitless
C_{XI}	Cross-correlation matrix between maximum and minimum temperatures	Unitless
E	Eigen decomposition matrix	Unitless
F	Toeplitz-Cholesky factorization matrix	Unitless
i, j	Locations	Unitless
I, \hat{I}, \bar{I}	Observed, expected and mean daily minimum temperature	$^{\circ}C$
n	Total number of locations	Unitless
η	Total number of pairwise daily observations	Unitless
$\varphi_{ij}, \hat{\varphi}_{ij}$	Cross-correlation between locations i, j based on observed and generated data, respectively	Unitless
r	Random number from a Gaussian $N[0, 1]$ distribution	Unitless
r^{Ψ}	Correlated random number from a Gaussian $N[0, 1]$ distribution	Unitless
R	Vector of random numbers following a Gaussian $N[0, 1]$ distribution	Unitless
R^{Ψ}	Vector of spatially correlated random numbers following a Gaussian $N[0, 1]$ distribution after matrix-multiplying a vector of random numbers by F or E	Unitless
$\rho_{ij}, \hat{\rho}_{ij}$	Pearson's correlation between locations i, j based on observed and generated data, respectively	Unitless
σ_i, σ_j	Standard deviation of daily temperature observations	$^{\circ}C$
$\sigma_{X_i}, \sigma_{I_j}$	Standard deviation of daily observations of maximum and minimum temperatures	$^{\circ}C$
t	Notation of time for a specific day	Days
T_{i_t}, T_{j_t}	Pairwise temperature observations on day t for locations i, j	$^{\circ}C$
T, \hat{T}, \bar{T}	Observed, expected and mean daily temperatures	$^{\circ}C$
X, \hat{X}, \bar{X}	Observed, expected and mean daily maximum temperatures	$^{\circ}C$

TABLE 3: List of weather stations used in this study.

Code	Name	County	Latitude ($^{\circ}N$)	Longitude ($^{\circ}W$)	Altitude (m a.s.l.)
0286	Asheboro	Randolph	$35^{\circ}42'$	$79^{\circ}50'$	265
1677	Chapel Hill	Orange	$35^{\circ}55'$	$79^{\circ}05'$	152
2500	Dunn	Harnett	$35^{\circ}19'$	$78^{\circ}41'$	61
3017	Fayetteville	Cumberland	$35^{\circ}04'$	$78^{\circ}52'$	29
4464	Jackson Springs	Montgomery	$35^{\circ}11'$	$79^{\circ}41'$	223
4860	Laurinburg	Scotland	$34^{\circ}45'$	$79^{\circ}28'$	64
7079	Raleigh State University	Wake	$35^{\circ}48'$	$78^{\circ}42'$	122
7656	Sanford	Lee	$35^{\circ}32'$	$79^{\circ}03'$	80
7924	Siler City	Chatham	$35^{\circ}46'$	$79^{\circ}28'$	186
7994	Smithfield	Johnston	$35^{\circ}31'$	$78^{\circ}21'$	46

As expected, only the third model reproduced the monthly observed cross-correlations between daily maximum and minimum temperatures. The correlation coefficient (relative RMSE) for the third model was 0.920 (10%), whereas the correlation coefficients (relative RMSEs) for the first and second models were 0.765 (57%) and 0.738 (67%), respectively (Figure 5).

Figure 6 shows the temporal structure of the observed and generated daily maximum and minimum temperatures from the weather station at Asheboro. Only one location is shown because all temporal semivariograms followed the same trend. The weather station was randomly selected. There were differences in the observed sill and range parameters of the temporal semivariograms. As expected, the second

model adequately represented the sill parameter; however, the range parameter was 2 days shorter than the observed range. This behavior could be because only a 1-day lag was included in the correlation matrix; future research should explore the extension of the input matrix of the second model to 2-day or even 3-day lag. The third model overestimated the sill parameter but adequately reproduced the range parameter (better in the summer than in the winter). A hybrid model between the second and third models could potentially better represent both the sill and the range parameters of the temporal semivariogram. This should be the focus of future research. The first model produced a range of 2 days shorter than the second model, and it overestimated the sill parameter as did the third model.

Comparisons between observed and generated multianual monthly means of daily maximum temperatures and daily minimum temperatures are shown in Figures 7(a), 7(b), 7(c), 8(a), 8(b), and 8(c), respectively. For both variables and for the three models, correlations were 1.000. Relative RMSEs for maximum temperature were 1%, 0.7%, and 0.2% for models from one to three, respectively. For minimum temperature, relative RMSEs were 1%, 2%, and 0.3%, respectively. The interannual variability of maximum temperatures (Figures 7(d), 7(e), and 7(f)) and minimum temperatures (Figures 8(d), 8(e), and 8(f)) were well reproduced by the three models. For maximum temperatures, correlation coefficients for the three models were 0.896, 0.900, and 0.965, respectively; for minimum temperatures, they were 0.901, 0.837, and 0.996, respectively. For maximum and minimum temperatures, relative RMSEs were 13% and 13% for the first model, 16% and 20% for the second model, and 27% and 14% for the third model, respectively. These comparisons were made between the observed and generated interannual variance of daily maximum and minimum temperatures, and they are directly related to the overdispersion problem (i.e., the underestimation of the interannual variability) reported in other weather generators [39]. All three models presented in this paper were able to reproduce the interannual variability of the observed temperatures; however all models, especially the third model, slightly underestimated the standard deviation.

3.2. Factorization of Nonpositive Definite Correlation Matrices.

For this paper, only 10 weather stations were used to demonstrate the performance of the described models; however, the factorization method was tested with 65 weather stations (results not published). In the third model, the size of the correlation matrix reached 130×130 elements. All twelve monthly correlation matrices were nonpositive definite. To factorize these matrices, the hyperspherical decomposition and the spectral decomposition methods [34] were applied, but without success. The extended Crout's algorithm needed

less than 10 iterations to solve the set of equations. In the current study, the new n -unknown values assigned to solved the Crout's algorithms were less than or equal to 1.10.

4. Conclusions

The three proposed stochastic models for generating spatially and temporally correlated daily maximum and minimum temperature data reproduced the main statistics of the observed historical record of each individual weather station as well as the spatial correlations between pairs of weather stations for each month, the temporal correlations within sites, and/or the cross-correlations between variables. The proposed models did not show the overdispersion problem on the interannual variability of maximum and minimum temperatures. Among the presented models, the final selection will depend on the specific user's necessities regarding the relative importance of taking into account spatial, temporal, and/or cross correlations.

The method for transforming a non-positive definite matrix into a positive definite matrix to satisfy the matrix factorization requirements produced satisfactory results.

The methods proposed in this paper can be used in the same manner in any other region (larger area, different density of the station network, complex topography, etc.). After some simple transformation to fit other non-Gaussian distributions, the methods proposed can be applied to different time step duration (e.g., hourly), other variables [10], or being applied outside the atmospheric sciences to datasets showing spatial-, temporal- and/or cross-correlated structure. Also, the use of semivariograms based on the spatial correlation could help if one wants to apply the model on a region where the length of the dataset is too short to permit a proper calculation of the correlation matrices. To follow the World Meteorological Organization's guidelines, a minimum of 30 years of daily historical records are necessary to calibrate the statistical models; however, to consider the effects of climate change, specific analyses must be performed to determine the best historical record length [9].

The idea of mixing the second (spatial and temporal correlations) and the third (spatial and cross-correlations) models will be carried out in future researches. The expectations of this hybrid model is to surpass its predecessor, generating data that is spatial-, temporal- and cross-correlated, closely matching the observed historical record; however, it is also expected that the necessary computational resources will increase.

Appendix

Extension of Crout's algorithm

$$\mathbf{C} = \begin{bmatrix} \rho_{00} & \cdots & \cdots & \rho_{0(n-1)} \\ \cdot & \cdots & & \cdot \\ \cdot & \cdots & \cdots & \cdot \\ \rho_{(n-1),0} & \cdots & \cdots & \rho_{(n-1)(n-1)} \end{bmatrix}; \quad \mathbf{L} = \begin{bmatrix} l_{00} & 0 & \cdots & 0 \\ \cdot & \cdots & \cdots & \cdot \\ \cdot & \cdots & \cdots & 0 \\ l_{(n-1),0} & \cdots & \cdots & l_{(n-1)(n-1)} \end{bmatrix}; \quad \mathbf{U} = \begin{bmatrix} u_{00} & \cdots & \cdots & u_{0(n-1)} \\ 0 & \cdots & & \cdot \\ \cdot & \cdots & \cdots & \cdot \\ 0 & \cdots & 0 & u_{(n-1)(n-1)} \end{bmatrix}. \quad (\text{A.1})$$

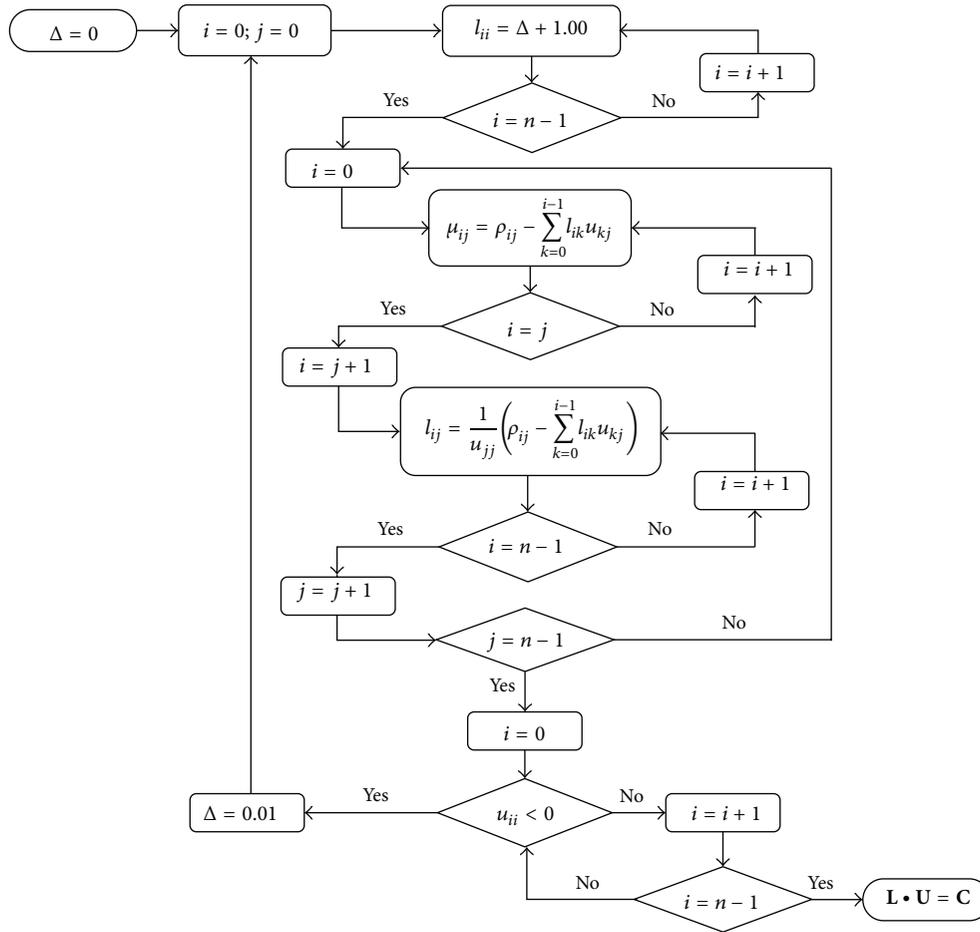


FIGURE 9: Flow diagram detailing the sequence of processes for the extended Crout's algorithm.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

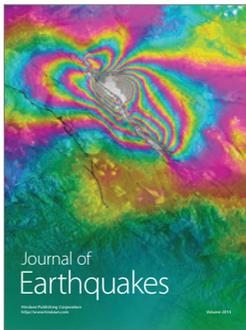
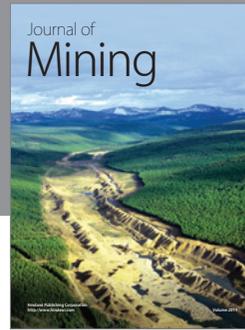
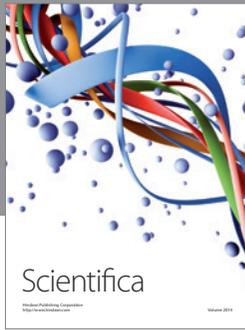
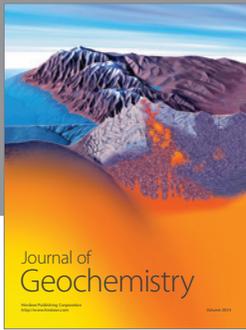
Acknowledgments

This research project was supported in part through grants from: the Agriculture and Good Research Initiative Competitive Grant no. 2011-00828 from the USDA National Institute of Food and Agriculture; the National Atmospheric and Space Administration—Earth Science Division Applied Sciences Program (NNH08ZDA001N-DECISIONS and NNH09ZDA001N-GULF), and it was developed under the auspices of AgMIP: The Agricultural Model Intercomparison and Improvement Project, and the CropClimate Initiative. The views expressed in this paper are those of the author and do not necessarily reflect the views of NASA, USDA, AgMIP, or any of its subagencies. The author wishes to thank two anonymous reviewers for their observations.

References

- [1] D. S. Wilks, "Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain," *Agricultural and Forest Meteorology*, vol. 96, no. 1-3, pp. 85-101, 1999.
- [2] D. S. Wilks, "A gridded multisite weather generator and synchronization to observed weather data," *Water Resources Research*, vol. 45, no. 10, Article ID W10419, 2009.
- [3] M. Khalili, F. Brissette, and R. Leconte, "Stochastic multi-site generation of daily weather data," *Stochastic Environmental Research and Risk Assessment*, vol. 23, no. 6, pp. 837-849, 2009.
- [4] B. Qian, J. Corte-Real, and H. Xu, "Multisite stochastic weather models for impact studies," *International Journal of Climatology*, vol. 22, no. 11, pp. 1377-1397, 2002.
- [5] D. S. Wilks, "Multisite generalization of a daily stochastic precipitation generation model," *Journal of Hydrology*, vol. 210, no. 1-4, pp. 178-191, 1998.
- [6] S. Apipattanavis, G. Podestá, B. Rajagopalan, and R. Katz, "A semiparametric multivariate and multisite weather generator," *Water Resources Research*, vol. 43, no. 11, pp. 1-19, 2007.

- [7] J. J. Beersma and T. A. Buishand, "Multi-site simulation of daily precipitation and temperature conditional on the atmospheric circulation," *Climate Research*, vol. 25, no. 2, pp. 121–133, 2003.
- [8] R. Leander and T. A. Buishand, "A daily weather generator based on a two-stage resampling algorithm," *Journal of Hydrology*, vol. 374, no. 3–4, pp. 185–195, 2009.
- [9] G. A. Baigorria, J. W. Jones, and J. J. O'Brien, "Understanding rainfall spatial variability in southeast USA at different timescales," *International Journal of Climatology*, vol. 27, no. 6, pp. 749–760, 2007.
- [10] A. M. Wootten, *Evaluating weather generator techniques for downscaling precipitation for seasonal forecasting [M.S. thesis]*, North Carolina State University, Raleigh, NC, USA, 2011.
- [11] C. C. Romero, M. D. Dukes, G. A. Baigorria, and R. Cohen, "Comparing theoretical irrigation requirement and actual irrigation for citrus in Florida," *Agricultural Water Management*, vol. 96, no. 3, pp. 473–483, 2009.
- [12] P. Droogers and R. G. Allen, "Estimating reference evapotranspiration under inaccurate data conditions," *Irrigation and Drainage Systems*, vol. 16, no. 1, pp. 33–45, 2002.
- [13] K. Bristow and G. Campbell, "On the relationship between incoming solar radiation and daily maximum and minimum temperature," *Agricultural and Forest Meteorology*, vol. 31, no. 2, pp. 159–166, 1984.
- [14] G. Hargreaves and Z. Samani, "Estimating potential evapotranspiration," *Journal of Irrigation and Drainage Engineering*, vol. 108, no. 3, pp. 225–230, 1982.
- [15] P. Cantelaube and J.-M. Terres, "Seasonal weather forecasts for crop yield modelling in Europe," *Tellus A*, vol. 57, no. 3, pp. 476–487, 2005.
- [16] H. Feddersen and U. Andersen, "A method for statistical downscaling of seasonal ensemble predictions," *Tellus A*, vol. 57, no. 3, pp. 398–408, 2005.
- [17] P. H. Gleick, "Methods for evaluating the regional hydrologic impacts of global climatic changes," *Journal of Hydrology*, vol. 88, no. 1–2, pp. 97–116, 1986.
- [18] B. Efron and R. J. Tibshirani, *An Introduction To the Bootstrap. Number 57 in Monographs on Statistics and Applied Probability*, Chapman and Hall, New York, NY, USA, 1993.
- [19] G. A. Baigorria, J. W. Jones, D.-W. Shin, A. Mishra, and J. J. O'Brien, "Assessing uncertainties in crop model simulations using daily bias-corrected Regional Circulation Model outputs," *Climate Research*, vol. 34, no. 3, pp. 211–222, 2007.
- [20] D. W. Shin, G. A. Baigorria, Y.-K. Lim et al., "Assessing maize and peanut yield simulations with various seasonal climate data in the southeastern United States," *Journal of Applied Meteorology and Climatology*, vol. 49, no. 4, pp. 592–603, 2010.
- [21] C. C. Romero, G. A. Baigorria, and L. Stroosnijder, "Changes of erosive rainfall for El Niño and La Niña years in the northern Andean Highlands of Peru," *Climatic Change*, vol. 85, no. 3–4, pp. 343–356, 2007.
- [22] F. S. Royce, C. W. Fraisse, and G. A. Baigorria, "ENSO classification indices and summer crop yields in the Southeastern USA," *Agricultural and Forest Meteorology*, vol. 151, no. 7, pp. 817–826, 2011.
- [23] G. A. Baigorria and J. W. Jones, "GiST: a stochastic model for generating spatially and temporally correlated daily rainfall data," *Journal of Climate*, vol. 23, no. 22, pp. 5990–6008, 2010.
- [24] C. W. Richardson and D. A. Wright, "WGEN, a model for generating daily weather variables," Tech. Rep., USDA ARS Bulletin No. ARS-8, Washington, DC, USA, Government Printing Office, 1984.
- [25] J. J. Filliben, "The probability plot correlation coefficient test for normality," *Technometrics*, vol. 17, no. 1, pp. 111–117, 1975.
- [26] R. L. Iman and W. J. Conover, "A distribution-free approach to inducing rank correlation among input variables," *Communications in Statistics*, vol. 11, no. 3, pp. 311–334, 1982.
- [27] E. M. Scheuer and D. S. Stoller, "On the generation of normal random vectors," *Technometrics*, vol. 4, no. 2, pp. 278–281, 1962.
- [28] C. Benoît, "Note sur une méthode de résolution des équations normales provenant de l'application de la méthode des moindres carrés a un système d'équations linéaires en nombre inférieur a celui des inconnues—application de la méthode a la résolution d'un système défini d'équations linéaires," *Bulletin Géodésique*, vol. 2, no. 1, pp. 67–77, 1924.
- [29] C. Brezinski, "The life and work of André Cholesky," *Numerical Algorithms*, vol. 43, no. 3, pp. 279–288, 2006.
- [30] O. Toeplitz, *Die Jacobische Transformation der Quadratischen Formen von Unendlich Vielen Veränderlichen*, Nachrichten von der Akademie der Wissenschaften zu Göttingen, Mathematisch—Physikalische Klasse, Göttingen, Germany, 1907.
- [31] O. Taussky and J. Todd, "Cholesky, Toeplitz and the triangular factorization of symmetric matrices," *Numerical Algorithms*, vol. 41, no. 2, pp. 197–202, 2006.
- [32] D. Hilbert, *Grundzüge Einer Allgemeinen Theorie der Linearen Integralgleichungen*, Nachrichten von der Königliche Gesellschaft der Wissenschaften zu Göttingen, Mathematisch—Physikalische Klasse, Göttingen, Germany, 1904.
- [33] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933.
- [34] R. Rebonato and P. Jäckel, "The most general methodology to create a valid correlation matrix for risk management and option pricing purposes," *Journal of Risk*, vol. 2, pp. 17–26, 2000.
- [35] P. D. Crout, "Determinants and solving systems of linear equations with real or complex coefficients," *Transactions of the American Institute of Electrical Engineers*, vol. 60, no. 12, pp. 1235–1240, 1941.
- [36] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New York, NY, USA, 3rd edition, 2007.
- [37] R. P. Boyles and S. Raman, "Analysis of climate trends in North Carolina (1949–1998)," *Environment International*, vol. 29, no. 2–3, pp. 263–275, 2003.
- [38] L. Ferraris, S. Gabellani, N. Reboria, and A. Provenzale, "A comparison of stochastic models for spatial rainfall downscaling," *Water Resources Research*, vol. 39, pp. 1368–1383, 2003.
- [39] D. R. Cox, "Some remarks on overdispersion," *Biometrika*, vol. 70, no. 1, pp. 269–274, 1983.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

