Weather Generator École de Technologie Supérieure (WeaGETS)

Version 1.1

User Manual

October 2010

The stochastic weather generator WeaGETS can be downloaded from the Mathworks file exchange website http://www.mathworks.com/matlabcentral/fileexchange/ . It is provided free of charge, with due acknowledgement for academic research purposes only. It cannot be used for any commercial purposes. WeaGETS is programmed in Matlab. Accordingly, the Matlab software package is needed for running WeaGETS. Although WeaGETS has been extensively tested, the routines are provided 'as is' with no guarantee against remaining errors.

For citations purposes, please use:

Chen, J., Brissette, P.F., Leconte, R., 2010. A daily stochastic weather generator for preserving low-frequency of climate variability. Journal of Hydrology 388, 480-490.

Addresses for communications:

François Brissette:

Departement of construction engineering, École de technologie supérieure, Université du Québec

1100, rue Notre-Dame Ouest, Montreal, Qc, Canada. H3C 1K3

Email: Francois.Brissette@etsmtl.ca

Jie Chen:

Département de génie de la contruction, École de technologie supérieure, Université du Québec

1100, rue Notre-Dame Ouest, Montreal, Qc, Canada. H3C 1K3

Email: chj092413@yahoo.com.cn

Foreword:

WeaGETS went through several iterations. Prof. Robert Leconte (now at Sherbrooke University) wrote the original Matlab code based on WGEN (Richardson and Wright, 1984). Prof. François Brissette then modified the code and streamlined it close to its current form. Master student Annie Caron tested several aspects of the code and added higher order Markov Chains for precipitation occurrence (Caron et al., 2008). Finally, PhD student Jie Chen provided several additional options including the correction scheme for the well know problem of the underestimation of inter annual variability (Chen et al., 2010), and the CLIGEN temperature scheme (Chen et al., 2010).

Input data file:

The supplied file *chat.mat* contains all inputs needed to run WeaGETS.

1. Introduction

WeaGETS is a Matlab-based versatile stochastic daily weather generator (WeaGETS) for producing daily precipitation, maximum and minimum temperatures (Tmax and Tmin) series of unlimited length, thus permitting impact studies of rare occurrences of meteorological variables. Furthermore, it can be used in climate change studies as a downscaling tool by perturbing their parameters to account for expected changes in precipitation and temperature. First, second and third-order Markov models are provided to generate precipitation occurrence, and exponential and gamma distributions are available to produce daily precipitation quantity. Precipitation generating parameters have options to be smoothed using Fourier harmonics. Two schemes (unconditional and conditional) are available to simulate Tmax and Tmin. Finally, a spectral correction approach is included to correct the well-known underestimation of monthly and inter-annual variability associated with weather generators.

WeaGETS has the advantage of incorporating the computational schemes of other well-known weather generators, as well as offering unique options, such as correction of the underestimation of inter-annual variability, and the ability to use Markov chains of varying orders. More importantly, the use of Matlab allows for easy modification of the source code to suit the specific needs of users. It would be very easy, for example, to add additional precipitation distribution functions. Finally, Matlab offers an integrated environment to further analyze the data generated by WeaGETS.

2. Model description

WeaGETS provides three options to generate precipitation occurrence, two options to produce precipitation quantity and two options to simulate Tmax and Tmin. There is also an option of smoothing the precipitation parameters with Fourier harmonics following Richardson's approach (1981), and to correct for the low-frequency variability of precipitation and temperature following the spectral correction method of Chen et al. (2010).

The basic input data include an observed weather data filename, a filename to store the subsequently generated data, a precipitation threshold value (minimum rainfall amount in 'mm' for a day to be considered wet) and the number of years of data to generate. Fig. 1 presents the WeaGETS structure chart.

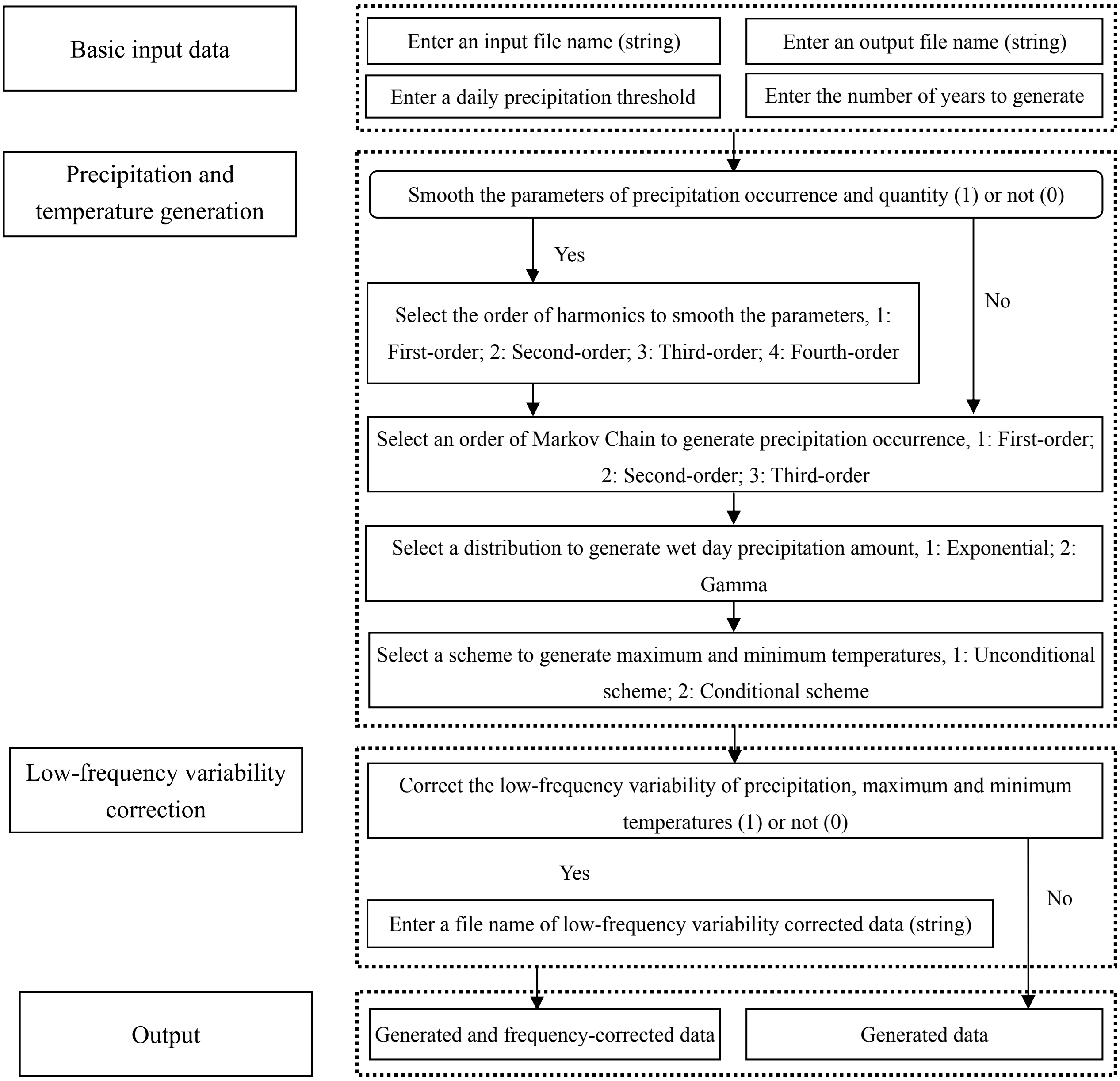

Fig. 1. Structure chart of the WeaGETS stochastic weather generator

2.1.Smoothing Scheme

The precipitation occurrence parameters include the transition probabilities of first, second and third-order Markov chains. For precipitation amounts, there is one parameter for the exponential distribution, and two parameters for the gamma distribution. These parameters are computed on a biweekly basis (26 estimations over the whole year). Because of climate variability and the finite length of the historical records, the variation from one 2-week period to the other will not be smooth, and the true yearly distribution of the parameter value will be party hidden. The user can decide to accept sudden variations (keeping constant parameters

values for the 2-week period) or to smooth the computed distribution to allow for smooth transitions of the parameters on a daily basis. In the latter case, WeaGETS will try to reproduce the precipitation characteristics of the smoothed line and not of the original observed values. In this case, generated precipitation may be slightly different than the observed precipitation. One to four Fourier harmonics can be used to smooth the yearly parameters distribution. The smoothing process eliminates sharp parameter transitions between computing periods that may occur due to outliers, especially for short time series. Fig. 2 presents the P10 parameter smoothed by Fourier harmonics. A first-order Fourier harmonic is clearly inadequate in this case. A higher number of harmonics will better fit the data at the potential expense of reproducing trends that may not exist (as would be the case in Fig. 2d).

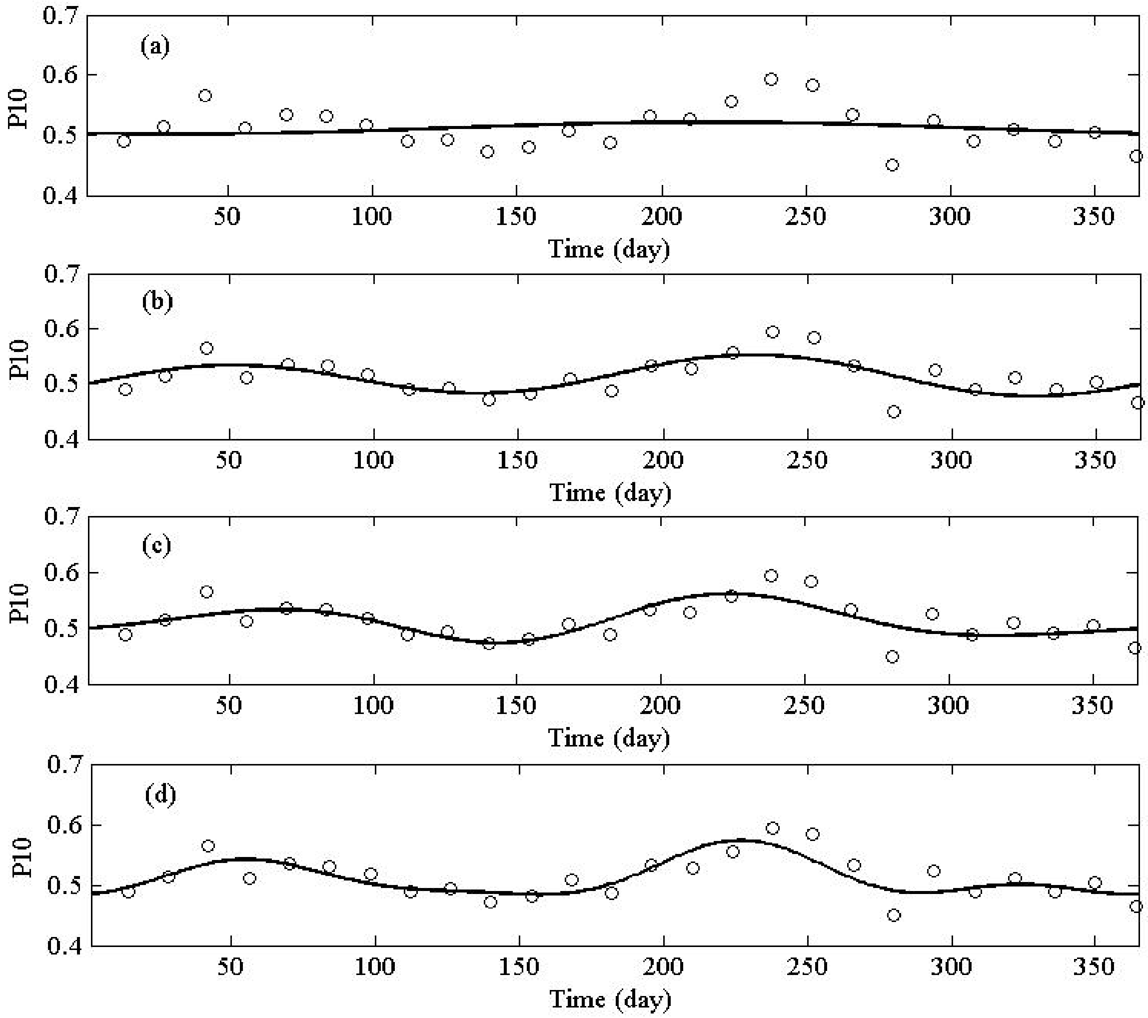

Fig. 2. A dry day following a wet day (P10) calculated at a two-week scale and smoothed by first-order (a), second-order (b), third-order (c) and fourth-order (d) Fourier harmonics.

The choice of smoothing or not, and how much smoothing is needed, is partly a

philosophical debate and will depend on the experience of the modeler. In most cases, the use of two harmonics is adequate for representing seasonal trends in the precipitation-generating parameters, but this depends on local climatology.

2.2. Generation of precipitation occurrence

WeaGETS provides three options including first, second and third-order Markov models to produce precipitation occurrence. The first-order Markov process is the simplest and most widely used. The probability of precipitation on a given day is based on the wet or dry status of the previous day, which can be defined in terms of two transition probabilities, P01 and P11:

P01 = Pr{precipitation on day t | no precipitation on day t-1}   (1a)

P11 = Pr{precipitation on day t | precipitation on day t-1}   (1b)

Since precipitation either occurs or does not occur on a given day, the two complementary transition probabilities are P00 = 1 - P01 and P10 = 1 - P11.

A generalization of the first-order Markov model is to consider higher-order Markov models such as the second and third-order models. Letting $R_t = 0$ if day *t* is dry, and $R_t = 1$ if day *t* is wet, Equations (1a) and (1b) can be extended to the second and third-order Markov chains following equations 2 and 3:

$P_{ijk} = \Pr\{ R_t = k \mid R_t = j \mid R_t = i \}$ (2)

$P_{hijk} = \Pr\{R_t = k \mid R_t = j \mid R_t = i \mid R_t = h\}$ (3)

where *h*, *i*, *j* and *k* =0 or 1, respectively.

The number of parameters required to characterize precipitation occurrence increase exponentially with the order of Markov process. This means that two, four and eight parameters must be estimated for first, second and third-order Markov models, respectively. As mentioned earlier, first-order Markov chains may not be adequate for generating long dry or wet spells. Higher-order Markov models perform better, but more parameters must be determined. Since a minimum number of rainfall events need to be present to adequately estimate transition probabilities, second and third-order parameter estimation requires longer time series of observed precipitation. If the goal is to use WeaGETS as a downscaling tool for climate change studies, the first-order process is usually more practical because it only

requires the perturbation of two parameters.

2.3.Generation of precipitation quantity

For a predicted rainy day, two probability distribution functions are available to produce the daily precipitation quantity. The first is the one-parameter exponential distribution, which has a probability density function given by

$$f(x) = \lambda e^{-\lambda x} \qquad (4)$$

where x is the daily precipitation intensity and λ is the distribution parameter (equal to the inverse of the mean).

The other function is the two-parameter gamma distribution. The probability density function for this distribution is given by

$$f(x) = \frac{(x/\beta)^{\alpha-1}\exp[-x/\beta]}{\beta\Gamma(\alpha)} \qquad (5)$$

where α and β are the two distribution parameters, and $\Gamma(\alpha)$ indicates the gamma function evaluated at α. This method is easy to compute and performs better than the exponential distribution. Therefore, it is widely used to generate daily precipitation quantity. It would be very easy to add other distribution functions, such as the mixed exponential (a three-parameter distribution) that has also been used in the literature.

2.4.Generation of maximum and minimum temperatures

Similarly to WGEN, the WeaGETS uses a first-order linear autoregressive model to generate Tmax and Tmin. The observed time series is first reduced to residual elements by subtracting the daily means and dividing by the standard deviations. The means and standard deviations are conditioned on the wet or dry status. The residual series are then generated by

$$\chi_{p,i}(j) = A\chi_{p,i-1}(j) + B\varepsilon_{p,i}(j) \qquad (6)$$

where $x_{p,i}(j)$ is a (2×1) matrix for day i of year p whose elements are the residuals of Tmax (j=1) and Tmin (j=2); $\varepsilon_{p,i}(j)$ is a (2×1) matrix of independent random components that are normally distributed with a mean of zero and a variance of unity. A and B are (2×2) matrices whose elements are defined such that the new sequences have the desired auto and

cross correlation coefficients. The A and B matrices are determined by

$$A = M_1 M_0^{-1} \tag{7}$$

$$B = M_0 - M_1 M_0^{-1} M_1^T \tag{8}$$

where the superscripts -1 and *T* denote the inverse and transpose of the matrix, respectively, and M_0 and M_1 are the lag 0 and lag 1 covariance matrices.

Two options are available to generate Tmax and Tmin on top of the generated residual series. The first is derived from WGEN or version 5.111 of CLIGEN. The daily values of Tmax and Tmin are found by multiplying the residuals by the standard deviation σ and adding the mean μ (equations 9 and 10). Throughout this paper, this option is referred to as the unconditional scheme.

$$T_{\max} = \mu_{\max} + \sigma_{\max} \times \chi_{p,i} \tag{9}$$

$$T_{\min} = \mu_{\min} + \sigma_{\min} \times \chi_{p,i} \tag{10}$$

Because Tmax and Tmin are generated independently of each other based on equations (9) and (10), there are a number of cases where Tmin is larger than Tmax. Thus, a range check is imposed to force Tmin to be less than Tmax. For example, if Tmin is greater than Tmax, Tmin is set equal to Tmax – 1.

The other option is derived from the latest version of CLIGEN (version 5.22564). The temperature with the smallest standard deviation between Tmax and Tmin is first computed, followed by the others (Chen et al. 2008). This option is referred to as the conditional scheme throughout this paper. If the standard deviation of Tmax is larger than or equal to the standard deviation of Tmin, daily temperatures are generated by equations (11) and (12):

$$T_{\min} = \mu_{\min} + \sigma_{\min} \times \chi_{p,i} \tag{11}$$

$$T_{\max} = T_{\min} + (\mu_{\max} - \mu_{\min}) + \sqrt{\sigma_{\max}{}^2 - \sigma_{\min}{}^2} \times \chi_{p,i} \tag{12}$$

If the standard deviation of Tmax is less than that of Tmin, daily temperatures are generated by equations (13) and (14):

$$T_{\max} = \mu_{\max} + \sigma_{\max} \times \chi_{p,i} \tag{13}$$

$$T_{\min} = T_{\max} - (\mu_{\max} - \mu_{\min}) - \sqrt{\sigma_{\min}{}^2 - \sigma_{\max}{}^2} \times \chi_{p,i} \tag{14}$$

Using this scheme, Tmin is always less than Tmax and no range check is necessary.

2.5. Correction of low-frequency variability

Weather generators underestimate the monthly and inter-annual variance, because they do not take into account the low-frequency component of climate variability. WeaGETS provides an approach to correct for this underestimation, for both precipitation and temperature.

Low-frequency variability is first modeled using a Fast Fourier Transform (FFT) based on the power spectra of the annual time series of precipitation and temperature. Generations of monthly and yearly precipitation and yearly average temperatures data are achieved by assigning random phases for each spectral component, which preserve the power spectrum and variances as well as the autocorrelation function. The link to daily parameters is established through linear functions. Throughout this paper, this is referred to as the spectral correction approach/method. The correction of monthly and inter-annual variability for precipitation follows the approach of Chen et al. (2010). Their results show that this approach performs very well in preserving the low-frequency variability of precipitation and temperatures.

3. Generation process

3.1. Input data

The input data consists of daily precipitation, Tmax and Tmin. The model does not take into account bissextile years. Any significant precipitation occurring on a February 29th should be redistributed equally on February 28th and March 1st. The maximum and minimum temperatures of a February 29th can be simply removed. Missing data should be assigned a -999 value. The input file contains the following matrices and vectors:

(1) P: matrix with dimensions [*nyears**365], where *nyears* is the number of years, containing daily precipitation in mm.

(2) Tmax: matrix with dimensions [*nyears* *365], where *nyears* is the number of years, containing maximum temperature in Celsius.

(3) Tmin: matrix with dimensions [*nyears* *365], where *nyears* is the number of years, containing minimum temperature in Celsius.

(4) yearP: vector of length [*nyears* *1] containing the years covered by the precipitation.

(5) yearT: vector of length [*nyears* *1] containing the years covered by the Tmax and Tmin.

3.2.Output data

The output also consists of daily precipitation, Tmax and Tmin values. It contains the following matrices:

(1) gP: matrix with dimensions [*gnyears**365], where *gnyears* is the number of years of generated precipitation in mm without low-frequency variability correction.

(2) gTmax: matrix with dimensions [*gnyears* *365], where *gnyears* is the number of years of generated Tmax in Celsius without low-frequency variability correction.

(3) gTmin: matrix with dimensions [*gnyears* *365], where *gnyears* is the number of years of generated Tmin in Celsius without low-frequency variability correction.

If the low-frequency variability correction option is chosen, another file will be produced. It also contains three matrices, named corP, corTmax and corTmin, respectively.

(1) corP: matrix with dimensions [*gnyears* *365], where *gnyears* is the number of years of generated precipitation in mm with low-frequency variability correction.

(2) corTmax: matrix with dimensions [*gnyears* *365], where *gnyears* is the number of years of generated Tmax in Celsius with low-frequency variability correction.

(3) corTmin: matrix with dimensions [*gnyears* *365], where *gnyears* is the number of years of generated Tmin in Celsius with low-frequency variability correction.

3.3.Running the program

There are many subprograms in the WeaGETS package, but the user only needs to run the main program *RUN_WeaGETS.m.* All of the options will then be offered in the form of questions, presented as follows:

(1) Basic input

a. Enter an input file name (string):

A name for the observed data shall be entered within single quotes, for instance, '*filename*' for the supplied file.

b. Enter an output file name (string):

A name for the generated data shall be entered within single quotes, for example '*filename_generated*'.

c. Enter a daily precipitation threshold:

   Precipitation threshold is the amount of precipitation used to determine whether a given day is wet or not (0.1mm is the most commonly used value).

d. Enter the number of years to generate:

   The number of years of the generated time series of precipitation and temperatures is entered here.

(2) Precipitation and temperature generation

a. Smooth the parameters of precipitation occurrence and quantity (1) or do not smooth (0).

b. If option 1 is selected, enter the number of harmonics to be used (between 1 and 4).

c. Select an order of Markov Chain to generate precipitation occurrence, 1: First-order; 2: Second-order; 3: Third-order.

d. Select a distribution to generate wet day precipitation amount: 1: Exponential or 2: Gamma.

e. Select a scheme to generate Tmax and Tmin: 1: Unconditional or 2: Conditional.

(3) Low-frequency variability correction.

a. Correct the low-frequency variability of precipitation, Tmax and Tmin (1) or do not correct (0).

   If option 1 is selected, a filename containing the corrected data will need to be entered.

Once weather generation is completed, the first year of generated data without and with the low-frequency variability correction will be plotted.

Acknowledgement:

This work was partially supported by the Natural Science and Engineering Research Council of Canada (NSERC), Hydro-Québec, Manitoba Hydro and the Ouranos Consortium on Regional Climatology and Adaptation to Climate Change.

References

Caron, A., Leconte, R., Brissette, F.P, 2008. Calibration and validation of a stochastic weather generator for climate change studies. Canadian Water Resources Journal. 33(3): 233-256.

Chen J., Brissette, P.F., Leconte, R., 2011. Assessment and improvement of stochastic weather generators in simulating maximum and minimum temperatures. Transactions of the ASABE. (Under review)

Chen J., Zhang, X.C., Liu, W.Z., Li, Z., 2008. Assessment and Improvement of CLIGEN Non-Precipitation Parameters for the Loess Plateau of China. Transactions of the ASABE 51(3), 901-913.

Chen, J., Brissette, P.F., Leconte, R., 2010. A daily stochastic weather generator for preserving low-frequency of climate variability. Journal of Hydrology 388, 480-490.

Chen, J., Brissette, P.F., Leconte, R., Caron, A. 2011. WeaGETS -- a Matlab based daily scale weather generator for generating precipitation and temperature. Environmental modelling & software. (Under review)

Nicks, A.D., Lane, L.J., Gander, G.A., 1995. Weather generator, Ch. 2. In USDA−Water Erosion Prediction Project: Hillslope Profile and Watershed Model Documentation, eds. D. C. Flanagan, and M. A. Nearing. NSERL Report No. 10. West Lafayette, Ind.: USDA−ARS−NSERL.

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

Richardson, C.W., Wright, D.A., 1984. WGEN: A model for generating daily weather variables. U.S. Depart. Agr, Agricultural Research Service. Publ. ARS-8.