Stochastic Simulation of Non-stationary Meteorological Time-series
Daily Precipitation Indicators, Maximum and Minimum Air Temperature Simulation using Latent and Transformed Gaussian Processes

Nina Kargapolova[1,2]
[1]*Institute of Computational Mathematics and Mathematical Geophysics, Pr. Lavrentieva 6, Novosibirsk, Russia*
[2]*Novosibirsk State University, Novosibirsk, Russia*

Keywords: Stochastic Simulation, Non-stationary Random Process, Air Temperature, Daily Precipitation, Extreme Weather Event.

Abstract: In this paper a stochastic parametric simulation model that provides daily values for precipitation indicators, maximum and minimum temperature at a single site on a yearlong time-interval is presented. The model is constructed on the assumption that these weather elements are non-stationary random processes and their one-dimensional distributions vary from day to day. A latent Gaussian process and its threshold transformation are used for simulation of precipitation indicators. Parameters of the model (parameters of one-dimensional distributions, auto- and cross-correlation functions) are chosen for each location on the basis of real data from a weather station situated in this location. Several examples of model applications are given. It is shown that simulated data may be used for estimation of probability of extreme weather events occurrence (e.g. sharp temperature drops, extended periods of high temperature and precipitation absence).

1 INTRODUCTION

For solution of different applied problems in such scientific areas as hydrology, agricultural meteorology and population biology, it is quite often necessary to take into account statistical properties of different meteorological processes. For example, it may be necessary to consider probability of occurrence of meteorological elements combinations contributing to forest fires spread, probability of frost occurrence in spring and summer, average number of dry days, etc. Since real data samples are usually small, real data based statistical investigation of rare and extreme weather events is in most cases unreliable. Therefore, instead of small real data samples it is necessary to use samples of simulated data.

In this regard, in recent decades a lot of scientific groups all over the world work at development of so-called "stochastic weather generator". At its core, "generators" are software packages that allow numerically simulate long sequences of random numbers having statistical properties, repeating the basic properties of real meteorological series. Most often series of surface air temperature, daily minimum and maximum temperatures, precipitation and solar radiation are simulated (Furrer, 2007; Kargapolova, 2012; Richardson, 1981; Richardson, 1984; Semenov, 2002). Not only single-site time series, but also spatial and spatio-temporal meteorological random fields are simulated with the use of "weather generators" (Kleiber, 2012; Ogorodnikov, 2013; Kargapolova, 2016). It should be noted that practically all “weather generators” possess same drawback: a model that describes well main properties of a weather process over some region or at several locations may be totally unsuitable over another region (with different physiographic characteristics). At the same time, models that reproduce well characteristics of a weather process on a relatively short time-interval (a week, a month) may not be applicable for longer periods of time (season, year) and vice versa. It means that for each specific applied problem solution it is always a good idea to try several “weather generators” and then to choose the one that “works” better.

In this paper a stochastic parametric simulation model that provides daily values for precipitation indicators, maximum and minimum temperature at a single site on a yearlong time-interval is presented. The model is constructed on the assumption that

Kargapolova, N.
Stochastic Simulation of Non-stationary Meteorological Time-series - Daily Precipitation Indicators, Maximum and Minimum Air Temperature Simulation using Latent and Transformed Gaussian Processes.
DOI: 10.5220/0006358801730179
In *Proceedings of the 7th International Conference on Simulation and Modeling Methodologies, Technologies and Applications (SIMULTECH 2017)*, pages 173-179
ISBN: 978-989-758-265-3

these weather elements are non-stationary random processes and their one-dimensional distributions vary from day to day. A latent Gaussian process and its non-linear transformation (so called threshold transformation) are used for simulation of precipitation indicators. Parameters of the model are chosen for each location on the basis of real data from a weather station situated in this location. Several examples of model applications are given. It is shown that simulated data may be used for estimation of probability of extreme weather events occurrence.

2 MODEL DESPRIPTION

In this section a formal theoretical description of a considered stochastic model is given. Assumptions about properties of a real weather proses that were used for model construction are specified.

A model is constructed for simulation of joint time-series on twelve-month time interval. It is supposed that one-dimensional distribution of daily maximum and minimum temperature are Gaussian. This assumption is in good agreement (in sense of χ^2-criteria) with long-term observation data from weather stations. Parameters of these Gaussian distribution vary from day to day. Figure 1 illustrates variation of daily maximum and minimum temperature sample average on a yearlong interval.

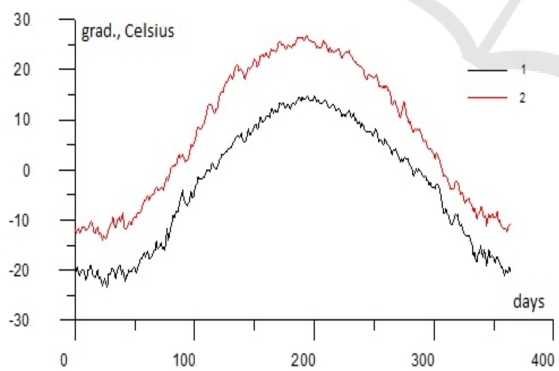

Figure 1: Sample average of daily minimum (1) and maximum (2) temperature. Years of observation: 1976 – 2009. Novosibirsk, Russia.

Daily precipitation indicator in a day number j, $j=\overline{1,N}$ is defined as 1 if amount of precipitation during this day in more of equal than 0.1 mm and as 0 otherwise. It is supposed that $N=365$ (for convenience data for February 29 is not taken into consideration). This means that daily precipitation indicator is a binary random process. Joint time-series of mentioned above weather elements are assumed to be non-stationary on twelve-month time interval.

Each simulated model trajectory is a matrix $M=\left(\vec{I}^T, \vec{A}^T, \vec{E}^T\right)$, where column-vector $\vec{I}^T=\left(I_1, I_2, \ldots, I_N\right)^T$ is a vector whose component I_j is daily minimum air temperature in a day number j, $\vec{A}^T=\left(A_1, A_2, \ldots, A_N\right)^T$ is a vector of daily maximum temperatures and column-vector $\vec{E}^T=\left(E_1, E_2, \ldots, E_N\right)^T$ is a vector of daily precipitation indicators.

Elements of a joint time-series M are calculated with the help of transformations

$$
\begin{aligned}
&I_j=\sigma_j^I \xi_j^I+\mu_j^I, \\
&A_j=\sigma_j^A \xi_j^A+\mu_j^A, \\
&E_j=\begin{cases} 1, & \xi_j^E \le c_j, \\ 0, & \xi_j^E > c_j, \end{cases}
\end{aligned} \tag{2.1}
$$

where vectors

$$
\begin{pmatrix} \mu_1^I \\ \mu_2^I \\ \vdots \\ \mu_N^I \end{pmatrix}, \begin{pmatrix} \mu_1^A \\ \mu_2^A \\ \vdots \\ \mu_N^A \end{pmatrix}, \begin{pmatrix} \sigma_1^I \\ \sigma_2^I \\ \vdots \\ \sigma_N^I \end{pmatrix}, \begin{pmatrix} \sigma_1^A \\ \sigma_2^A \\ \vdots \\ \sigma_N^A \end{pmatrix}
$$

are mean and standard deviation vectors, $j=\overline{1,N}$. Threshold values c_j are defined from equations

$$
P\left(E_j=1\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{c_j}\exp\left(-\frac{t^2}{2}\right)dt=p_j,\ j=\overline{1,N}.
$$

Values of $\mu_j^I, \mu_j^A, \sigma_j^I, \sigma_j^A, p_j$ are estimated on a basis of real data from a weather station. It is obvious that such way of model parameters definition make it possible to take into account seasonal variations of real weather processes It should be noted that for all $j=\overline{1,N}$ equality $c_j=0$ is true if and only if $p_j=0.5$, inequalities $c_j>0$ and $p_j>0.5$ are equivalent. Hereafter it is supposed that $p_i \neq 0,\ p_i \neq 1,\ i = \overline{1,N}$. Variables $\left(\xi_j^I, \xi_j^A, \xi_j^E\right)$ are components of a joint Gaussian process $\Omega = \left(\left(\vec{\xi}^I\right)^T, \left(\vec{\xi}^A\right)^T, \left(\vec{\xi}^E\right)^T\right)$ with zero mean and specific correlation matrix

$$G = \begin{pmatrix} G_{II} & G_{IA} & G_{IE} \\ G_{AI} & G_{AA} & G_{AE} \\ G_{EI} & G_{EA} & G_{EE} \end{pmatrix}.$$

Matrix G must be such that a process $\left(\vec{I}^T, \vec{A}^T, \vec{E}^T\right)$ after transformation (2.1) has a correlation matrix

$$R = \begin{pmatrix} R_{II} & R_{IA} & R_{IE} \\ R_{AI} & R_{AA} & R_{AE} \\ R_{EI} & R_{EA} & R_{EE} \end{pmatrix},$$

that is equal to sample correlation matrix. Method of matrix G calculation is described below. Dimension of matrixes G and R is 1095×1095 ($3N \times 3N$). Element $r_{XY}(i,j)$ of a matrix block R_{XY} is a correlation coefficient between X_i and Y_j ($X, Y \in \{I, A, E\},\ i, j \in \{1, 2, \ldots, N\}$). Element $g_{XY}(i,j)$ is corresponding to $r_{XY}(i,j)$ correlation coefficient of a Gaussian process.

Let's take a closer look at the matrix G and find equations that define this matrix when the matrix R is given. In (Ogorodnikov, 2009) a special case of such equations was considered. Normalisation of two correlated Gaussian random variables doesn't change a correlation coefficient between them, which implies

$$\begin{aligned} &G_{II} = R_{II},\ G_{AA} = R_{AA}, \\ &G_{IA} = R_{IA},\ G_{AI} = R_{AI}. \end{aligned} \tag{2.2}$$

Definition of a correlation coefficient leads to equations

$$r_{IE}(i,j) = \frac{EI_iE_j - EI_iEE_j}{\sqrt{DI_i}\sqrt{DE_j}} = \tag{2.3}$$

$$= -\frac{g_{IE}(i,j)}{\sqrt{2\pi p_j\left(1-p_j\right)}} \exp\left(-c_j^2/2\right),\quad i, j = \overline{1,N}.$$

These equations fully define matrix G_{IE}. Matrixes G_{EI}, G_{AE}, G_{EA} are defined in a similar way. Since

$$r_{EE}(i,j) = \frac{P\left(E_i = 1, E_j = 1\right) - p_ip_j}{\sqrt{p_i\left(1-p_i\right)}\sqrt{p_j\left(1-p_j\right)}},$$

$$P\left(E_i = 1, E_j = 1\right) = P\left(\xi_i^E \le c_i, \xi_j^E \le c_j\right),\quad i, j = \overline{1,N}$$

following equalities hold for $i, j = \overline{1,N}$ & $i \neq j$,

$$r_{EE}(i,j) = \frac{F\left(c_i, c_j, g_{EE}(i,j)\right) - p_ip_j}{\sqrt{p_i\left(1-p_i\right)}\sqrt{p_j\left(1-p_j\right)}}, \tag{2.4}$$

where

$$F(h,k,\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \times$$

$$\times \int_{-\infty}^{h}\int_{-\infty}^{k} \exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left(x^2 - 2\rho xy + y^2\right)\right) dxdy.$$

Obviosly,

$$r_{EE}(i,i) = g_{EE}(i,i) = 1,\ i = \overline{1,N}.$$

It should be noted that equations (2.4) don't have any analytical solutions, but it is possible to solve them numerically. So, equations (2.2) – (2.4) define matrix G and, finally, we may formulate the simulation algorithm.

<u>Algorithm</u>:

Step 1. Estimate $\mu_j^I, \mu_j^A, \sigma_j^I, \sigma_j^A, p_j,\ j = \overline{1,N}$ on a basis of real data.

Step 2. Solving equations (2.2) – (2.4) define matrix G.

Step 3. Simulate required number of trajectories of a joint Gaussian process Ω with zero mean and correlation matrix G.

Step 4. Using equalities (2.1) transform trajectories of a Gaussian processes into trajectories of a non-Gaussian process $M = \left(\vec{I}^T, \vec{A}^T, \vec{E}^T\right)$.

If verification of obtained trajectories gives satisfying result, these trajectories may be used for study of rare / extreme events.

<u>Remark 1</u>. Due to a physical sense of daily minimum and maximum temperatures, an inequality $I_j \le A_j$ must be true for all $j = \overline{1,N}$. But transformation (2.1) doesn't guarantee it. This means that one must eliminate from consideration all trajectories in which this inequality violates. In

practice, it is typical that $\mu_j^I \ll \mu_j^A$ and σ_j^I, σ_j^A are relatively small, so usually there are few trajectories with $I_j > A_j$.

Remark 2. Equations (2.4) are solved numerically, so some computational errors may appear. These errors influence on the matrix G and it may happen that obtained matrix G is not positively-defined. In this case before a Gaussian process simulation a normalisation of the matrix G must be done (see, Ogorodnikov, 1996). There are a lot of algorithms for simulation of a Gaussian process with given correlation matrix. The most common are algorithms based on LU – decomposition of the correlation matrix and on its spectral representation.

Remark 3. Numerical solution of equations (2.4) is a time-consuming problem. There is a way to reduce computational time. So-called Owen's formulas (Owen, 1956) give a representation of function $F(h,k,\rho)$ via one-dimensional integrals:

$$F(c_i, c_j, g_{EE}(i,j)) = \frac{1}{2}\Phi(c_i) + \frac{1}{2}\Phi(c_j) - \\ -T(c_i, a_1) - T(c_j, a_2) + \frac{1}{2},$$

if $c_i \ge 0, c_j \ge 0$ or $c_i < 0, c_j < 0$, and

$$F(c_i, c_j, g_{EE}(i,j)) = \frac{1}{2}\Phi(c_i) + \frac{1}{2}\Phi(c_j) - \\ -T(c_i, a_1) - T(c_j, a_2),$$

if $c_i < 0, c_j \ge 0$ or $c_i \ge 0, c_j < 0$, where

$$\Phi(c_i) = \frac{1}{\sqrt{2\pi}} \int_0^{c_i} \exp\left(-t^2/2\right) dt,$$

$$T(c,a) = \frac{1}{2\pi} \int_0^a \exp\left(-\frac{c^2(1+t^2)}{2}\right) \frac{dt}{1+t^2},$$

$$a_1 = \frac{c_j - c_i g_{EE}(i,j)}{c_i \sqrt{1-g_{EE}^2(i,j)}},\ a_2 = \frac{c_i - c_j g_{EE}(i,j)}{c_j \sqrt{1-g_{EE}^2(i,j)}}.$$

This representation together with the fact that

$$p_i = \frac{1}{2} + \Phi(c_i),\ i = \overline{1,N}$$

let to replace equations (2.4) with equations

$$r_{EE}(i,j) = \frac{\frac{1}{2}p_i + \frac{1}{2}p_j - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} - \\ -\frac{T(c_i, a_1) + T(c_j, a_2)}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}}, \tag{2.5a}$$

if $c_i c_j > 0$,

$$r_{EE}(i,j) = \frac{\frac{1}{2}p_i + \frac{1}{2}p_j - \frac{1}{2} - p_i p_j}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} \\ -\frac{T(c_i, a_1) + T(c_j, a_2)}{\sqrt{p_i(1-p_i)}\sqrt{p_j(1-p_j)}} \tag{2.5b}$$

if $c_i c_j < 0$,

$$r_{EE}(i,j) = -\frac{2T\left(c_j, -\frac{g_{EE}(i,j)}{\sqrt{1-g_{EE}^2(i,j)}}\right)}{\sqrt{p_j(1-p_j)}}, \tag{2.5c}$$

if $c_i = 0, c_j \ne 0$,

$$r_{EE}(i,j) = -\frac{2T\left(c_i, -\frac{g_{EE}(i,j)}{\sqrt{1-g_{EE}^2(i,j)}}\right)}{\sqrt{p_i(1-p_i)}} \tag{2.5d}$$

if $c_i \ne 0, c_j = 0$ and

$$r_{EE}(i,j) = \frac{2\arcsin g_{EE}(i,j)}{\pi} \tag{2.5e}$$

if $c_i = c_j = 0$.

Numerical experiments show that computational time required for solution of equations (2.5a) – (2.5e) is approximately 4 times less than computational time required for solution of equations (2.4). This is due to the fact that computation of a one-dimensional integral is much simpler than computational of a bivariate integral.

Remark 5. In some numerical experiments, obtained matrix G was ill-conditioned and didn't let accurate simulation of a Gaussian process. It calls for further investigations to find out conditions when matrix G is ill-conditioned. Ways of matrix correction are also have to be found.

Remark 6. Since correlation coefficients $g_{EE}(i,j)$ may be found from equations (2.4) (or (2.5a) – (2.5e)) independently from each other and trajectories of the Gaussian process are also simulated independently, parallel computing technologies may be easily applied for simulation of the process $M = \left(\vec{I}^T, \vec{A}^T, \vec{E}^T\right)$.

3 NUMERICAL EXPERIMENTS

Described above stochastic model was used for simulation of joint meteorological non-stationary time-series on more than 50 weather stations situated in different climatic zones in Russia. Verification of the model shows that the model gives satisfactory results for most of the stations. Here is an example of a process characteristic that was used for the model verification. Average numbers of days in a month, when minimum temperature is below 0 ºC and maximum temperature is above 0 ºC ($I_j < 0, A_j > 0$), estimated on basis of real and simulated data, were compared. This characteristic is not the model input parameter, so it can be used for verification. Table 1 presents values of this characteristic. It can be seen from Table 1, that the model reproduces this characteristic accurately (up to a statistical mistake).

Table 1: Average number of days with $I_j < 0, A_j > 0$. St. Petersburg, Russia.

Month	Average number of days	
	Real data	Simulated data
October	4.7	4.9
November	8.9	9.1
December	5.7	5.3
January	9.3	9.4
February	7.7	7.4
March	16.3	16.8

Since the model is adequate to real weather processes, it may be used for study of rare / extreme events. Here are several examples. Hereafter all estimations on basis of real data were done for years of observation from 1976 to 2009 and estimations based on simulated data were done over 10^6 trajectories.

First considered characteristic is a probability of low temperatures and light frosts in spring and summer. These weather events may negatively influence on open-ground planted crop species. Since different species have different resistance to frost, it is necessary to take probability of low temperatures and light frosts into account when choosing a varieties or species of plants. Formally, considered characteristic may be written as $P\left(I_j < \alpha\right)$ when j varies from 121 (May 1) to 243 (August 31). Here α ºC (deg. Celsius) is a given temperature level. Table 2 presents estimations of $P\left(I_j < \alpha\right)$ obtained on basis of real and simulated data. For real and simulated data estimations two and three, respectively, fraction digits are significant. During years of observation, there were no days in considered period with temperature below -6 ºC (this is subminimum temperature for most of plants species), but it doesn't mean that such temperature drop is impossible. Simulated data provides an estimation of probability of this rare and severe weather condition.

Table 2: Estimations of $P\left(I_j < \alpha\right)$ obtained on basis of real and simulated data. Novosibirsk, Russia.

α ºC	$P\left(I_j < \alpha\right),\ j = \overline{121, 234}$	
	Real data	Simulated data
2	0.081	0.085
0	0.030	0.031
-2	0.014	0.013
-6	0.000	0.001

Table 3: Average number of summer time-intervals lasting k days, with absence of precipitation and daily minimum temperature above 20ºC. Astrakhan, Russia.

Period length, days	Average number of time-intervals	
	Real data	Simulated data
k=1	5.3	5.72
k=2	2.0	2.20
k=4	0.7	0.76
k=5	0.6	0.59
k=6	0.0	0.53
k=8	0.0	0.37
k=9	0.3	0.32
k=10	0.1	0.09

Another weather event that may be dangerous both to individuals and to agricultural industry is long-term combination of high air temperature and absence of precipitation. Such combination may negatively influence on individuals' health and may cause soil drying up. Table 3 presents average number of time-intervals lasting k days, when daily minimum temperature was above 20ºC and there were no precipitations (only significant digits are given). Averaging was done over summer months. Once again, described in the paper model reproduces this characteristic for short time-intervals satisfactory, so model results for longer time-intervals may be considered as reasonable.

Finally, let's consider such unpleasant weather event as sharp temperature drop or rise during one day (formally, $\left|A_j - I_j\right| > \beta$, where β ºC is given level.). Numerical analysis shows that this characteristic is reproduced well for $\beta \in [5,14]$. For $\beta > 14$ ºC real data estimations are unreliable. This means that for applied problems solutions it is better to use simulated data estimations. Table 4 presents seasonal probabilities of such temperature variation with $\beta = 20$ ºC.

Table 4: Seasonal probabilities of $\left|A_j - I_j\right| > 20$ ºC. Ulan-Ude, Russia.

Season	Seasonal average number of days	
	Real data	Simulated data
Winter	0.011	0.009
Spring	0.120	0.124
Summer	0.027	0.025
Autumn	0.017	0.021

4 CONCLUSIONS

In this paper a model for simulation of meteorological time-series was considered. It was also shown that simulated trajectories may be used for study of rare / extreme events.

There are several ways of the model improvement. For example, instead of Gaussian one-dimensional distribution of daily minimum and maximum temperatures a mixture of 2 Gaussian distributions may be used. This will make computation of the matrix G much more complex, because it will require usage of the inverse distribution function method, but it will give a chance to reproduce temperature behavior more precisely. Simulation of precipitation indicators $\vec{E}^T$ may be replaced also by simulation of daily precipitation amount $\vec{D}^T = (D_1, D_2, \ldots, D_N)^T$ in a form of a multiplicative process

$$D_j = E_j C_j, \quad j = \overline{1,N},$$

where $\vec{C}^T = (C_1, C_2, \ldots, C_N)^T$ is a conditioned non-Gaussian random process describing amount of daily precipitation on the assumption of their presence. Such process representation is used in a well-known "weather generator" WGEN (Richardson, 1984), but indicator process in WGEN-model differs fundamentally from process $\vec{E}^T$ considered in this paper. These two model's modifications are subject of further research.

ACKNOWLEDGEMENTS

This work was supported by the Russian Foundation for Basis Research (grants No 15-01-01458-a, 16-31-00123-mol-a, 16-31-00038-mol-a) and the President of the Russian Federation (grant No MK-659.2017.1).

REFERENCES

Furrer, E.M., Katz, R.W., 2007. Generalized linear modeling approach to stochastic weather generators. In *Clim. Res., Vol. 34, No 2. P. 129 – 144.*

Kargapolova, N.A., 2016. *Stochastic models of meteorological processes*, NSU CPC, Novosibirsk, 1st edition (in Russian).

Kargapolova, N.A., Ogorodnikov, V.A., 2012. Inhomogeneous Markov chains with periodic matrices of transition probabilities and their application to simulation of meteorological processes. In *Russ. J. Num. Anal. Math. Modelling, Vol. 27, No 3. P. 213 – 228.*

Kleiber, W., Katz, R.W., Rajagopalan, B., 2012. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. In *Water Resour. Res., Vol. 48, No 1. P. 1 – 17.*

Ogorodnikov, V.A., Kargopolova, N.A., Seresseva, O.V., 2013. Numerical stochastic model of spatial fields of daily sums of liquid precipitation. In *Russ. J. Num. Anal. Math. Modelling, Vol. 28, No 2. P. 187 – 200.*

Ogorodnikov, V.A., Khlebnikova, E.I., Kosyak, S.S., 2009. Numerical stochastic simulation of joint non-Gaussian meteorological series. In *Russ. J. Num. Anal. Math. Modelling, Vol. 24, No 5. P. 467 – 480.*

Ogorodnikov, V.A., Prigarin, S.M., 1996. *Numerical Modelling of Random Processes and Fields: Algorithms and Applications*, VSP. Utrecht, 1st edition.

Owen, D.B., 1956. Table for computing bivariate normal probabilities. In *Ann. Math. Statist., Vol. 27, No 4. P. 1075 – 2000.*

Richardson, C.W., 1981. Stochastic simulation of daily precipitation, temperature and solar radiation. In Water Resour. Res., Vol. 17, No 1. P. 182–190.

Richardson, C.W., Wright, D.A., 1984. *WGEN: A Model for Generating Daily Weather Variables*, U. S.

Department of Agriculture, Agricultural Research Service, ARS-8, 1st edition.

Semenov, M.A., Barrow, E.M. 2002. LARS-WG: A Stochastic Weather Generator for Use in Climate Impact Studies. Version 3.0. User Manual. 1st edition.