Climak: A stochastic model for weather data generation

Article *in* Italian Journal of Agronomy · January 2002

Francesco Danuso
University of Udine

Ital. J. Agron., 6, 1, 57-71

Climak: a Stochastic Model For Weather Data Generation

F. DANUSO

Dipartimento di Produzione Vegetale e Tecnologie Agrarie, Università di Udine, Italy

Corresponding Author: F. Danuso, Dipartimento di Produzione Vegetale e Tecnologie Agrarie, Università di Udine, via delle Scienze 208, 33100 Udine, Italy. Tel.: +39 0432 558614, Fax: +39 0432 558603, E-mail: danuso@dpvta.uniud.it

Received: 18 February 2002. Accepted: 28 June 2002.

ABSTRACT

BACKGROUND. Many human activities and ecological processes are affected by climatic conditions. Despite the difficulties in long term meteorological forecasting, a statistical description of climate is possible and used for planning purposes and strategic decisions. With this aim, stochastic models for the generation of daily weather data have been developed (Climak). The generated meteorological data can be used: a) to perform Monte Carlo simulations with deterministic models (e.g., crop growth and ecological models, models for climate risk assessment); b) to better classify the climates; c) to assess environmental scenarios for the effects of climatic changes by "what if" procedures; d) to spatially interpolate the climate parameters, so obtaining data for location not covered by meteorological stations. These data can be used as input for agro-ecological models, mainly when a probabilistic evaluation of climatic uncertainty and risk is of interest. Climak has been developed to directly take into account the among-years variability.

METHODS. Climak generates, as first, the occurrence of rain and the rainfall amount for rainy days. After rainfall generation, minimum and maximum air temperatures are generated, with different parameters for rainy and dry days. Solar radiation is obtained from the astronomical photoperiod and from the daily thermal excursion. Evapotranspiration (reference or potential) is generated from solar radiation, if available; if not, it is obtained from photoperiod and daily maximum temperature.

RESULTS. The model was evaluated using meteorological data coming from different locations of northern Italy and comparing its capability to reproduce the climatic properties with that of another weather generator (Wxgen). The behaviour of Climak was quite satisfactorily even if some minor problems have been highlighted. The minimum number of years of data to correctly estimate the climatic parameters was also determined.

CONCLUSIONS. Climak was shown to be satisfactorily accurate in generating meteorological data representing the climate of a site, even when compared with a well known weather generator. At present the model still has some limitations and it has not been tested in climates other then the temperate ones. These drawbacks will be improved in the next versions of Climak.

Key-words: climate models, weather generators, software, meteorological data, planning.

INTRODUCTION

The climate of a site is the resultant of variations in some relevant climatic factors such as air temperature, precipitation, solar radiation and evaporative demand of the atmosphere. Over time, the weather factors show periodical trends (daily, annual, multi-annual) and stochastic variability. The randomness level of climatic variables can be very different; for example, the rain event can be considered as a purely stochastic process while the daylight duration (photoperiod) can be considered as a completely systematic factor. The weather factors also show correlation and dependence on the previous values of the same and other variables. Many human activities and ecological processes are controlled by the climatic conditions, and so great efforts have been devoted to weather forecasting. Intrinsic limits prevent the long term forecasting but a statistical description of the climate is possible. This knowledge allows climatic stochastic models for the generation of weather data to be developed (Jones et al., 1970; Richardson, 1981; Larsen and Pense, 1982; Wgen, Richardson and Wright, 1984; Shu Geng et al., 1985; Wxgen, Richardson and Nicks, 1990;

USClimate, Hanson et al., 1994; MTCLIM, Thornton et al., 1997; ClimGen, Stöckle et al., 2001; Skills and Richardson, 1998; Parlange and Katz, 2000; GEM, Johnson et al., 1996). The generated meteorological data can be used to perform Monte Carlo simulations with deterministic models (e.g., models of crop growth and of ecological processes, models for the climate risk assessment). The climatic models also permit the classification of different climates, the generation of meteorological data for sites not covered by meteorological stations and the study of the effects of climatic changes by "what if" methods (Wilks, 1992).
A model (Climak; Danuso and Della Mea, 1994) for the generation of daily values for rainfall (Prec, mm/d), minimum air temperature (Tn, °C), maximum air temperature (Tx, °C), integral of solar radiation (Rad, MJ m^{-2} d^{-1}) and evapotranspirative demand of the atmosphere (ETr, mm d^{-1}) is described. The weather generation procedure involves the estimation of climatic parameters from the historical meteorological recordings and the weather generation from deterministic models and residuals sampled from probability distributions. The scheme of the whole generation procedure is reported in Figure 1.
Climak was applied to the regional scale (Danuso et al., 1997) by creating a grid of climatic parameters for the plains of Friuli-Venezia Giulia to be used to perform Monte Carlo simulations with agro-ecological models.

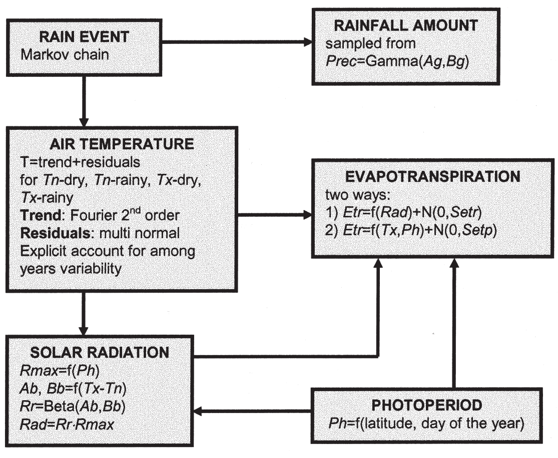

Figure 1. Structure and generation procedure of Climak.

MODEL
Rainfall

Rain is generated first because it is required for the generation of the other variables. Two aspects are considered: the occurrence of the rain event (if a day is rainy or not) and, for the rainy days, the amount of rainfall.

Rain occurrence
Model
The rain event is assumed to be a stochastic process, represented by a first order Markov chain (Larsen and Pense, 1982; Richardson, 1985). The state of each day (rainy or dry) is obtained from the state of the previous day and from the dry to dry (*Pdd*) and rainy to dry (*Prd*) transition probabilities. The complementary dry to rainy (*Pdr*) and rainy to rainy (*Prr*) probabilities are computed as:

$$Pdr = 1 - Pdd \quad \text{and} \quad Prr = 1 - Prd$$

The transition probabilities are considered on a monthly base and then the model requires 24 parameters for the rain event generation (12 for *Pdd* and 12 for *Prd*).
Parameter estimation
The transition probabilities are calculated on a monthly base and are estimated, for a location, on all the available recordings in the data set as:

$$Pdd=Ndd/Nd \quad \text{and} \quad Prd=Nrd/Nr$$

where:
Ndd number of dry days in the month, preceded by a dry day;
Nd total number of dry days in the data set, for the month;
Nrd number of dry days in the month preceded by a rainy day;
Nr total number of rainy days in the data set, for the month.

Rainfall amount
Model
The rainfall amount distribution typically follows a negative exponential-like distribution, with parameters varying during the course of the year. All these distributions are limited to positive values and allow also the generation of rare intense precipitation. The most used probability density function (pdf) for the rainfall amount generation is the Gamma distribution

(Jones et al., 1970; Larsen and Pense, 1982) but exponential and mixed-exponential distributions are also used.
In Climak the precipitation amount is generated by sampling from a two parameters Gamma pdf:

$$Gamma(x) = \frac{Bg^{-Ag} \cdot x^{Ag-1} \cdot e^{-x/Bg}}{\Gamma(Ag)}$$

if $x > 0$ and $Ag > 0$ and $Bg > 0$; otherwise

$$Gamma(x) = 0$$

where $\Gamma(Ag)$ is the gamma function. Ag and Bg are specific parameters for each month. Moreover, because the rain gages cannot detect rainfall below a sensitivity threshold (*Sthr*, about 0.2 mm), the rainfall data are corrected for this threshold. The total number of parameters needed to describe the rainfall amount is 25 (*Sthr*; Ag and Bg for each month).
Parameter estimation
Before estimating the parameters of the Gamma distribution, the sensitivity threshold of the instrument (*Sthr*) is subtracted from rainfall data, in order to obtain a distribution starting from 0. Then the Ag and Bg parameters are estimated, on a monthly base, by the method of the moments ($Ag = M^2/V$; $Bg = V/M$, where M is the mean and V the variance of the daily rainfall amounts). Often, for the estimation of the Gamma pdf parameters, the polynomial approximation of maximum likelihood from Greenwood and Durand (1960) has been used (Larsen and Pense, 1982; Shu et al., 1985). However, the method of the moments was here adopted because, for our data, it showed slightly better in reproducing the original climate statisctics.

Air temperature
Model
Air temperature has a strong dependence on the rainfall status of the day and then it is normally generated with different parameters for rainy and dry days. Moreover the yearly trend of minimum and maximum temperatures show, in high latitudes, a quite regular sinusoidal shape. The approach to calculate daily minimum and maximum temperatures uses the following steps: 1) minimum and maximum air temperature, both in rainy and dry days, are considered separately; 2) depending on the rain status of the day, the mean value of minimum temperature for that day, is calculated from the proper trend; 3) a random residual is added to the value of 2); this residual is sampled from a bivariate normal distribution, autocorrelated with the minimum temperature residuals of the previous day (first-order autoregressive process); 4) depending on the rain status of the day, the mean value of maximum temperature for that day is calculated, from the proper trend;
5) a random residual is added to the value of 4); this residual is sampled from a bivariate normal distribution of minimum and maximum temperature residuals, taking into account the correlation between the residuals of minimum and maximum temperature for the same day.
The minimum (Tn) and maximum (Tx) air temperatures are obtained as:

$Tn=Tnd+Rn$ and $Tx=Txd+Rx$ if the day is dry,

$Tn=Tnr+Rn$ and $Tx=Txr+Rx$ if the day is rainy,

where Tnd is the average daily minimum temperature for the dry days, obtained as a function of the date, by an interpolating second order Fourier series of the type:

$$Tnd=A+B\cdot\sin[(Date-C)\cdot 2\pi/365] + D\cdot\sin[(Date-E)\cdot 2\pi/182.5]$$

where:
- A annual average minimum temperature of the dry days (°C);
- B semi-amplitude of the first term (°C);
- C phase shift for the first term (days), assumed as a constant for a location;
- D semi-amplitude of the second term (°C);
- E phase shift for the second term (days), assumed as a constant for a location;
- $Date$ date as day of the year (from 1 to 365).

A, B and D are obtained by sampling, at the beginning of each year, from the normal distributions $N(MA,SA)$, $N(MB,SB)$, $N(MD,SD)$, so determining the specific trend for each year. Since statistics of measured data do not show correlations among parameters, independent samplings are used.
Analogous calculations, with the specific parameters, are made for the trends of minimum temperatures for rainy days (Tnr), the maximum temperatures for dry days (Txd) and maximum temperature for rainy days (Txr).
Rn is a random residual sampled from a bivariate normal distribution $N2$ (0, $\mathbf{\Omega n}$) with 0

vector of means and **Ωn** covariance matrix of residuals from the minimum temperature of the previous day and the same of the present day. The dispersion matrix corresponds to:

$$\mathbf{\Omega n} = \begin{bmatrix} SRn^2 & RRnn \cdot SRn^2 \\ RRnn \cdot SRn^2 & SRn^2 \end{bmatrix}$$

where:

SRn standard deviation of the residuals from the trends of the minimum temperature, both in dry and rainy days;

RRnn autocorrelation coefficient for minimum temperature residuals, with time lag of 1 day.

Rx is a random residual sampled from a bivariate normal distribution N2(0, **Ωx**), where the dispersion matrix **Ωx** is:

$$\mathbf{\Omega x} = \begin{bmatrix} SRn^2 & RRnx \cdot SRn \cdot SRx \\ RRnx \cdot SRn \cdot SRx & SRx^2 \end{bmatrix}$$

where:

SRx standard deviation of the residuals from the trends of the maximum temperature, both in dry and rainy days;

RRnx correlation coefficient between minimum temperature residuals and maximum temperature residuals of the same day.

The model requires 80 parameters to describe the air temperature. Of these, 32 parameters are used for the Fourier series (*C*, *E*, *MA*, *MB*, *MD*, *SA*, *SB* and *SD* of the four different temperature trends *Tnd*, *Tnr*, *Txd*, *Txr*), while 48 parameters are used to characterise the monthly distributions of the residuals (*SRn*, *SRx*, *RRnn*, *RRnx*).

Parameter estimation

The yearly trends of the four combinations of temperature are obtained by the following procedure:

1) C and *E* are estimated with the iterative non-linear regression algorithm implemented in Climak (modified Gauss-Newton), on the whole measured data set by using the Fourier series. *C* and *E* are assumed to be constant among years, for a specific site;

2) year by year, the values of *A*, *B* and *D* are estimated using linear regression and assuming *C* and *E* as constants;

3) the parameters for the normal distributions of *A*, *B* and *D* are then calculated. The mean values for *A*, *B* and *D* are obtained as:

$MA=\Sigma Aj/n$, $MB=\Sigma Bj/n$ and $MD=\Sigma Dj/n$

where:

j = index of the year: j=1,..., n

n = number of years in the data set

The standard deviations for *A* (*SA*), *B* (*SB*) and *D* (*SD*) are obtained as:

$$SA = \sqrt{\sum (Aj - MA)^2 / n}$$

$$SB = \sqrt{\sum (Bj - MB)^2 / n}$$

$$SD = \sqrt{\sum (Dj - MD)^2 / n}$$

4) by using the values of *A*, *B* and *D* from linear regression on the data set of each year, the trends of the temperatures are obtained and then the residuals from minimum (*Rn*) and maximum temperature (*Rx*) are calculated. While four temperature trends are calculated (depending on the rain status of the day) for the residual distributions no distinction is made between rainy and dry days;

5) subsequently the residuals are evaluated on a monthly base. It is assumed that there is a normal distribution of residuals and no difference between the residual distributions of dry and rainy days. For each month, the procedure computes the standard deviation of *Rn* and *Rx* (*SRn*, *SRx*), the autocorrelation of *Rn* residuals with those of the previous day (*RRnn*) and the correlation between *Rn* and *Rx* (*RRnx*). This specific correlation, instead of others (e.g., autocorrelation of *Tx* or cross-correlation) was selected because it showed the highest values.

Given that the residuals are obtained from the specific trend of each year, the among year variability of the *Rn* and *Rx* distributions has not been considered. The variability of the trends is taken into account by calculating the variability among years of *A*, *B* and *D*.

Solar radiation

Model

The maximum (potential) daily integral of solar radiation (*Rmax*) strictly depends on photoperiod (*Ph*), while the actual radiation at soil level (*Rad*) also depends on the cloudiness of the day. Cloud data are not, in general, available; to evaluate the atmosphere transmittance (related to cloudiness) the effect of cloudiness on daily thermal excursion ($Es = Tx\text{-}Tn$) is considered. Normally, on cloudy days the maximum temperature is low because sun radiation is strongly intercepted and the minimum temperature is high because the long wave irradiation from the soil during the night is reduced by the moisture of the atmosphere, so reducing *Es*.

Radiation is modelled in two steps: the first calculates the annual trend of the maximum daily radiation (*Rmax*). This was found to be well linearly related to the duration of the photoperiod (Figure 2); this relation is considered to be the same for all the years of each location. The second step evaluates the distribution of the ratio between the daily radiation and *Rmax*. This ratio ($Rr = Rad/Rmax$), ranges from 0 to 1 and is linearly related to the daily thermal excursion (*Es*). Moreover, the distribution of *Rr* values for the different *Es* classes was found to be a Beta pdf (Figure 3). The equation for this distribution is:

$$Beta(x) = \frac{\Gamma(Ab + Bb)}{\Gamma(Ab)\cdot\Gamma(Bb)} \cdot x^{Ab-1} \cdot (1-x)^{Bb-1}$$

if $0 \leq x \leq 1$ and $Ab > 0$ and $B\ b > 0$; otherwise

$$Beta(x) = 0$$

The Beta pdf, being defined for real values ranging from 0 to 1, fits well the different shapes of the *Rr* distributions at varying *Es*. *Ab* and *Bb* are parameters, specific for each of the five Es classes adopted.

The model needs 12 parameters for the generation of the solar radiation: 2 for the calculation of *Rmax* from the photoperiod and 10 for the Beta parameters (2 x 5 excursion classes).

Parameter estimation

A preliminary step for the estimation of the radiation parameters is the calculation of the daily astronomical photoperiod depending on latitude and the day of the year. This is performed with the method described in Keisling (1982). The parameters of the linear relation between *Rmax* and *Ph* are obtained by selecting only the maximum values of the solar radiation in ten-day periods of the year:

$$Rmax = b_1 \cdot Ph + b_0$$

Subsequently the program calculates the ratio between the daily radiation *Rad* and the estimated function for *Rmax* (*Rr*). The radiation data are then divided into five classes, depending on five classes of daily air temperature excursion. For each class and from the ratio *Rr* the two parameters of the Beta distribution are evaluated, using the moments:

$$Ab = M^2 \cdot (1\text{-}M)/V - M$$

$$Bb = a \cdot (1\text{-}M)/M$$

where *M* and *V* are the mean and variance of *Rr*, for each excursion class.

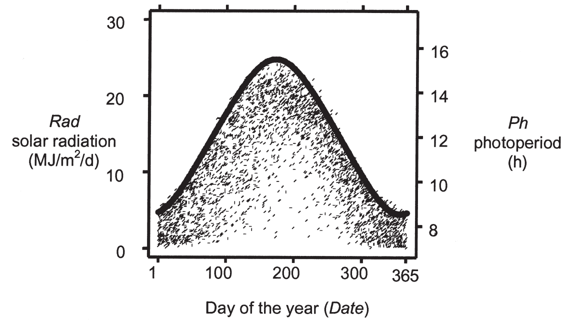

Figure 2. Distribution over the year of daily integrals of solar radiation (*Rad*) recorded at Padua from 1981 to 1988 (represented as dots). The astronomical calculated photoperiod (*Ph*) is also reported (solid line); it was rescaled in order to show its adequacy in representing the maximum solar radiation for each day of the year (*Rmax*).

Evapotranspiration

Model

The reference evapotranspiration always shows a good linear relation with the radiation

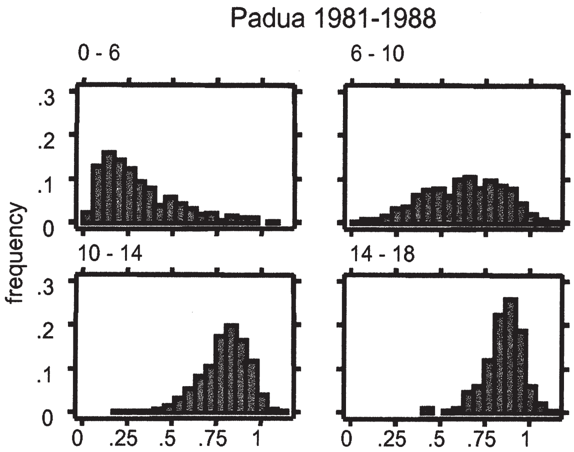

Figure 3. Frequency distribution of the ratio *Rr=Rad/Rmax* for each of four classes of daily temperature excursion (0-6, 6-10 10, 10-14 and 14-18 °C). The distributions can be adequately represented by a Beta pdf. Data were recorded at Padua from 1981 to 1988.

(Doorembos and Pruitt, 1977); the dependence on maximum air temperature and photoperiod is less good. Given that radiation data are often not available in the historical meteorological series, two different approaches are adopted, depending on availability of radiation data:

1) if solar radiation data are available, the daily reference evapotranspiration is obtained as a linear function of the daily radiation, plus a residual from a normal distribution (unique for all the months) with standard deviation *Setr*:

$$ETr = a_1 \cdot Rad + a_0 + N(0, Setr)$$

2) if radiation data are not available, *ETr* is modelled as a function of maximum air temperature and photoperiod. The following relation was found to give a good fit:

$$ETr = c_0 + c_1 \cdot Tx \cdot Ph^2 / 1000 + N(0, Setp)$$

where *Setp* is the standard deviation of the residuals, related to the photoperiod by a linear function: $Setp = d_1 \cdot Ph + d_0$. In this case the standard deviation of the residuals is not considered as a constant, but linearly depending on the photoperiod.

Three parameters are then needed for the generation of the reference evapotranspiration if radiation data are available and 4 in the other case.

Parameter estimation

The estimation of the parameters a_1 and a_0 is simply made by linear regression of *ETr* vs. *Rad*. Then, the standard deviation of the residuals (*Setr*), obtained subtracting the predicted data from the measured ones, is evaluated.

With no available radiation data, c_1 and c_0 come from a linear regression of *ETr* vs. $Tx \cdot Ph^2$. The residuals of this regression show a distribution with a variance increasing with the photoperiod. Then the standard deviation of residuals is calculated separately for each of 12 photoperiod classes. These values are then regressed against photoperiod to give the d_0 and d_1 parameters.

It should be noted that the procedure for evapotranspiration can be applied to simulate every variable representing the evapotranspirative demand of the atmosphere (e.g., without limitation due to water shortage). So, data of "reference evapotranspiration", "potential evapotranspiration", "evapotranspiration from class A pan" and others can be generated, depending on the data set used to estimate the parameters.

PROCEDURE FOR WEATHER DATA GENERATION

Climak generates first the occurrence of rainy or dry day and the rainfall amount if the day is rainy (Figure 1). After rainfall generation, minimum and maximum air temperature are generated, separately, for rainy and dry days. Solar radiation is obtained from the astronomical photoperiod and from the daily thermal excursion. If radiation data are available, the evapotranspiration is generated from solar radiation; if not, it is obtained from photoperiod and maximum temperature.

The *rain event* is generated by sampling, for each day to be generated, a random value from the uniform distribution *U*(0,1). If the current day is dry, the *Pdd* probability is used; if the sampled value *U*(0,1) is less than *Pdd* the following day is set to "dry", otherwise it is set to "rainy". The same procedure is adopted with the *Prd* transition probability if the present day is rainy. The considered probability is that specific for each month.

On the rainy days, the *rainfall amount* (*Prec*) is obtained by sampling a value from a *G*(*Ag*,*Bg*) pdf and adding the threshold value for the instrumental sensitivity:

$$Prec = G(Ag, Bg) + Sthr$$

where the parameters *Ag* and *Bg* are specific for each month.

The sampling routine for the Gamma distribution was implemented merging two algorithms: the first allows the generation from a Gamma pdf when the parameter *Ag* is less than 1 (RGS algorithm; Ahrens et al., 1972; Tadikamalla, 1981; Bratley et al. 1984); the second is used when Ag is greater than or equal to 1 (G3A algorithm; Fishman, 1978).

In general, the rainfall amount distribution has the form of a Gamma pdf with $Ag < 1$ (negative exponential-like); the complete routine has been implemented because it is used also for the generation of the solar radiation.

The generating method for the *air temperature* values is strictly related to the procedure for the estimation of the parameters. After rainfall generation, the minimum and maximum temperatures are generated separately, depending on the status of the day (rainy or dry).

At the beginning of a new year the program samples a value for the parameters A, B and D from the normal probability distributions $N(MA,SA)$, $N(MB,SB)$ and $N(MD,SD)$, for each temperature combination, in order to obtain the annual trends of temperatures by the Fourier series. C and E are considered to be constant for all the years while means and standard deviations of A, B and D are different in relation to the year and for the minimum/maximum and rainy/dry combinations.

Using these parameters, the annual trend of minimum and maximum air temperature on dry and rainy days is generated, for each year; then, month by month, each temperature is obtained by adding the residuals to the trends. The residuals are obtained as follows:

1) the minimum temperature residual (Rn) is sampled from the autocorrelated normal distribution with 0 mean, SRn standard deviation and $RRnn$ autocorrelation:

$$Rn = RRnn \cdot R1n + SRn \cdot \sqrt{1 - RRnn^2} \cdot N(0,1)$$

where:

Rn minimum temperature residual of the current day;

$R1n$ residual of minimum temperature of the previous day, already generated;

$N(0,1)$ value sampled from a normal distribution with 0 mean and 1 standard deviation;

2) the maximum temperature residual (Rx) is sampled from the bivariate normal distribution of parameters 0 mean, SRx standard deviation and $RRnx$ correlation coefficient, depending on the value Rn:

$$Rx = RRnx \cdot SRx \cdot Rn / SRn + SRx \cdot \sqrt{1 - RRnx^2} \cdot N(0,1)$$

The generation of *radiation* data is based on air temperature excursion and so a previous generation of air temperature is required. As in the estimation phase, Climak calculates first the daily photoperiod (Ph). From Ph, the maximum daily radiation $Rmax$ is obtained as $Rmax=b_1 \cdot Ph+b_0$, for each day of the year.

After that, for each day, a value of Rr is generated by sampling from a Beta probability distribution: $Rr=B(Ab,Bb)$, where Ab and Bb are parameters of the Beta pdf, chosen in relation to the temperature excursion class of the day. Five classes, in the range 0 to 20 °C of excursion, are adopted. For the Beta generation, a well-known property of this distribution is used:

$$B(Ab,Bb)=G(Ab,1)/(G(Ab,1)+G(Bb,1))$$

where $B(Ab,Bb)$ is a value extracted from the Beta pdf and $G(Ab,1)$ and $G(Bb,1)$ are values from the Gamma distribution. The generation of a Beta pdf is made by using the algorithm for the Gamma generation twice and setting Ab or Bb to 1.

Finally, the daily solar radiation is calculated as:

$$Rad = Rmax \cdot Rr$$

If radiation parameters are available, the daily reference *evapotranspiration* is obtained, after the generation of radiation, by generating a value from the normal distribution $N(0,Setr)$ as:

$$ETr= a_0 + a_1 \cdot Rad + N(0,Setr)$$

When the radiation parameters are missing, Climak calculates ETr from maximum temperature and photoperiod. The standard deviation of the ETr residuals ($Setp$) is, in this case, obtained as a function of the photoperiod as:

$$Setp = d_1 \cdot Ph+d_0$$

The residuals are then generated from a normal distribution $N(0,Setp)$ and the daily reference evapotranspiration is obtained from maximum air temperature, photoperiod and residuals as:

$$ETr= c_0 + c_1 \cdot Tx \cdot Ph^2/1000 + N(0,Setp)$$

MODEL IMPLEMENTATION

Program use

The current version of Climak (1.4, 24 June 1996) was implemented as a Dos executable program written in TurboPascal 6.0 and requires conventional and EMS memory. Two versions are available: a menu version (**climak.exe**) and a command line version (**climakc.exe**). Both programs are able to estimate parameters (text files with **cmk** name extension) from historical

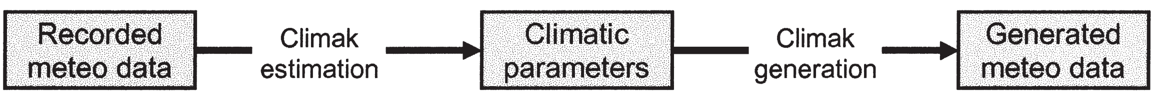

Figure 4. Procedure for the parameter estimation and weather data generation.

data (text files with **met** extension) and to generate meteorological data (text files with **gen** extension) from the parameter file (Figure 4). A version of Climak for the Stata ™ (Stata, 1993) statistical package is also available.

The maximum amount of years of data to be processed depends on the available EMS memory. The same problem exists for the data generation but, in this case, it can be overcome by generating data several times and then joining the different text files of generated data.

The generation of meteorological data is performed by using the climate parameters contained in the parameter files. Only the variables for which the related parameters are in the parameter file are generated: that is, if some values are missing, the corresponding climate variables are not generated. As a minimum, Climak can generates only the rainfall. The generation of temperature requires rainfall; radiation requires temperature and evapotranspiration requires radiation or temperature (Figure 1).

Structure of the meteorological data files

The meteorological data files of Climak are text files with name extension **met** (recorded data) and **gen** (generated data). All the required columns have to be present, even if data are missing. Missing values have to be indicated by a full stop (.). Climak recognises as a comment all the characters following the asterisk symbol (*) in the same row. The meteorological data files have seven columns (variables) which must follow the order: year code, day of the year, precipitation, minimum air temperature, maximum air temperature, solar radiation, reference evapotranspiration. If a variable is not available, its column has to be filled with full stops. The column of year and day of the year cannot contain missing values, while the other variables can.

The minimum set of variables required for the estimation of a parameter set is formed by the year code, the day of the year and the rainfall; in this case only the rainfall parameters can be estimated. To estimate also the temperature parameters, rainfall and both minimum and maximum temperatures have to be supplied. The estimation of the radiation parameters needs the temperature variables. The estimation of the evapotranspiration parameters requires the radiation variable or the temperature variables; when the radiation variable is present, by default, the parameters of *ETr* are obtained from radiation.

Structure of the parameter file

The generation of meteorological data is performed by using the climatic parameters contained in the parameter files. Climatic parameters are estimated by the Climak program from recorded meteorological data.

The parameter files are text files with **cmk** extension and 148 rows of numeric data. Climak recognises as comment each line starting with an asterisk.

The file of estimated parameters has one column of 148 elements. Each parameter has a definite position in the column. Parameters that cannot be estimated are indicated by a full stop (missing value).

A complete set of parameters is of 147/148 items. Of these, 3 are used for the general site description, 49 for the rainfall generation, 80 for the temperatures, 12 for the radiation and 3 or 4 for the reference evapotranspiration (depending on the generation method).

Only the variables for which the related parameters are in the parameter file are generated: that is, if some parameters are missing, the corresponding climate variable is not generated.

Latitude (as degrees and decimals) is used for the photoperiod calculations and then its value has an effect on the generation of radiation and on evapotranspiration. Longitude and altitude serve only for documentation or when a spatial interpolation has to be performed.

The structure of the parameter file is reported in Tables 1, 2, 3 and 4.

MODEL EVALUATION

Number of years needed to estimate Climak parameters

The goodness of the generated weather data depends on the model but also on the quality of the estimate of parameters. This depends, in turn, on the estimation method and on the number of year of data available for the estimation. An evaluation of the minimum sample size to estimate the Climak parameters was carried out. If the meteorological data comes with too short a time period, the estimated values may not correctly represent the statistical properties of the climate of the site and so the generated meteorological data should be used with caution.

The minimum number of years needed to give

Table 1. Structure of the climatic parameter file of Climak. The parameter values estimated on Padua 1981-1988 meteorological data are reported. Each parameter is located in a specific row of the climatic parameter file.
(†) 12 parameters; one for each month. See Table 2.
(‡) 4 parameters; one for each combination: min. temperature-dry day, min. temperature-rainy day, max. temperature-dry day, max. temperature-rainy day. See Table 3.
(¶) 5 parameter; one for each of five thermal excursion classes. See Table 4.
(¥) Parameters a_1, a_0 and *Setr* are used when radiation data are available. If not, the parameters c_1, c_0 and d_1 are estimated. The reported values are for a_1, a_0 and *Setr*.

Rows	Symbol	Values	Parameter description	Unit
General parameters				
1	*Lat*	45	latitude	degrees
2	*Long*	12	longitude	degrees
3	*Alt*	20	altitude above sea level	m
Rainfall parameters				
4	*Sthr*	0.2	rainfall sensitivity threshold	mm/d
5-16	*Pdd*	(†)	transition probability dry to dry day	-
17-28	*Prd*	(†)	transition probability rainy to dry day	-
29-40	*Ag*	(†)	gamma distribution: 1st parameter	-
41-52	*Bg*	(†)	gamma distribution: 2nd parameter	-
Temperature parameters				
53-56	*MA*	(‡)	mean of the *A* parameter of Fourier series	°C
57-60	*MB*	(‡)	mean of the *B* parameter of Fourier series	°C
61-64	*MD*	(‡)	mean of the *D* parameter of Fourier series	°C
65-68	*C*	(‡)	mean of the *C* parameter of Fourier series	date
69-72	*E*	(‡)	mean of the *E* parameter of Fourier series	date
73-76	*SA*	(‡)	standard deviation of *A*	°C
77-80	*SB*	(‡)	standard deviation of *B*	°C
81-84	*SD*	(‡)	standard deviation of *D*	°C
85-96	*SRn*	(†)	standard deviation of *Tn* residuals (*Rn*)	°C
97-108	*SRx*	(†)	standard deviation of *Tx* residuals (*Rx*)	°C
109-120	*RRnn*	(†)	autocorrelation coefficients of *Tn* residuals	-
121-132	*RRnx*	(†)	correlation coefficients between *Tn* and *Tx* residuals	-
Radiation parameters				
133	b_1	2.878	slope of *Rmax* = f(*Ph*)	-
134	b_0	-19.733	intercept of *Rmax* = f(*Ph*)	-
135-139	*Ab*	(¶)	beta distribution: 1*st* parameter	-
140-144	*Bb*	(¶)	beta distribution: 2*nd* parameter	-
Evapotranspiration parameters (¥)				
145	a_1 / c_1	0.286	slope for *ETr*=f(*Rad*) / slope for *ETr*=f(*Tx·Ph2*)	-
146	a_0 / c_0	0.102	constant for *ETr*=f(*Rad*) / const. for *ETr*=f(*Tx·Ph2*)	mm/d
147	*Setr* /d_1	0.917	standard deviation of *ETr* residuals / slope *Setp*=f(*Ph*)	-
148	- / d_0	-	- / constant for *Setp*=f(*Ph*)	mm/d

Table 2. Values for some climatic parameters estimated using the 1981-1988 meteorological data recorded at Legnaro (Padua, Italy) (45° Lat N).

Month	*Pdd* dry to dry transition probability	*Prd* rainy to dry transition probability	*Ag* Gamma pdf parameter	*Bg* Gamma pdf parameter	*SRn* Std.dev. of *Tn* residuals	*SRx* Std.dev. of *Tx* residuals	*RRnn* autocorrelation of *Tn* residuals	*RRnx* correlation *Tn-Tx* residuals
JAN	0.898	0.481	0.600	8.097	2.964	2.881	0.639	0.537
FEB	0.921	0.513	0.602	16.261	2.161	2.963	0.477	0.471
MAR	0.840	0.521	0.549	13.938	2.679	2.657	0.604	0.332
APR	0.785	0.484	0.976	6.133	2.901	3.103	0.742	0.441
MAY	0.697	0.385	0.882	8.040	2.424	2.909	0.667	0.573
JUN	0.790	0.645	0.787	11.592	2.209	3.092	0.602	0.547
JUL	0.855	0.792	1.310	6.225	2.194	2.422	0.572	0.656
AUG	0.846	0.702	1.152	8.644	2.154	2.469	0.594	0.610
SEP	0.881	0.500	0.693	10.326	2.352	2.590	0.532	0.453
OCT	0.864	0.712	0.993	10.953	3.161	2.512	0.687	0.376
NOV	0.888	0.510	0.724	12.663	3.012	2.909	0.679	0.549
DEC	0.893	0.448	1.163	7.720	2.829	2.671	0.610	0.643

Table 3. Values for the annual temperature trend parameters estimated using the 1981-1988 meteorological data recorded at Legnaro (Padua, Italy) (45° Lat N).

Temperature	A		B		C	D		E
type	*MA*	*SA*	*MB*	*SB*		*MD*	*SD*	
Tn-dry	6.83	0.795	10.08	0.595	111.8	-0.545	0.693	130.91
Tn-rainy	8.59	0.559	8.14	0.907	114.6	-0.376	1.335	108.83
Tx-dry	16.71	0.521	12.07	0.461	110.9	-1.817	0.744	128.49
Tx-rainy	15.67	0.328	10.47	0.642	113.4	-0.244	0.649	107.58

a correct estimate of the parameters depends on the degree of variability of the climate itself and on the specific parameters. For example, rainfall values are known to present higher variability than temperature values.

Using the historical meteorological data series of Udine (from 1960 to 1989) the estimation of parameters has been performed, with the specific Climak routine, using a sample of years increasing size from 5 to 30 years. In Figure 5 the estimated values for *Bg* (Gamma parameter for the generation of rainfall amount), for each month, of different series length are reported. For the transition probability which controls the rain events, *Pdd* requires, in the average of months, less year data to obtain a stable estimate than *Prd* (15 and 20 years, respectively). The *Pdd* and *Prd* values are more easily estimated for the winter and spring months (about 10 years are enough). More years are progressively required for the summer and fall months (up to 20 years).

The estimation of the annual temperature trend parameters (*MA*, *MB*, *MD*, *C*, *E*) were good with just 10-15 years, for all the *Tn*-*Tx* and dry-rainy combinations.

Table 4. Values of the Beta pdf parameters (*Ab*, *Bb*) for the five daily temperature excursion classes from 0 to 20 °C. The parameters were estimated using the 1981-1988 meteorological data recorded at Legnaro (Padua, Italy).

Class	Excursion	*Ab*	*Bb*
1	Low	1.092	2.676
2	Medium-low	1.569	1.292
3	Medium	4.476	1.437
4	Medium-high	8.261	1.418
5	High	7.919	0.812

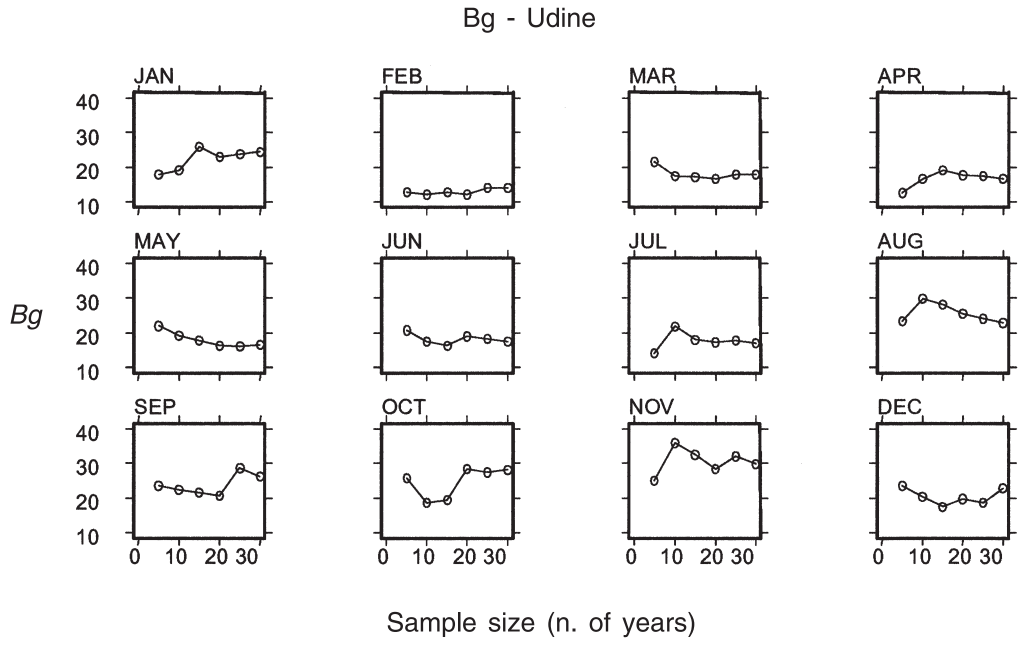

Figure 5. Estimates of the rainfall parameters *Bg*, for each month, using different sample sizes, on meteorological data recorded at Udine from 1960 to 1989.

Performance of Climak and Wxgen

The aim of the weather generators is to generate meteorological sequences having the same statistical properties as the actual ones. Thus, there are two ways to evaluate the goodness of the generated data: the first involves the statistical comparison of parameters of the observed and generated data sets (centrality and dispersion parameters, distributions). Another, pragmatic, way considers the generated data as acceptable when a dynamic simulation model, requiring meteorological data as input, gives the same results with both recorded and generated data.

Climak has already been evaluated and compared with other weather generators also by other authors. Acutis et al. (1998) evaluated Climak as compared with ClimGen (Stöckle et al., 2001) and with Wxgen using long-period series from different localities: Akron (USA), Haarveg (NL), Modena (IT), Tolouse (FR) and Udine (IT). The behaviour of the models was judged satisfactorily.

In the present paper, a comparison between recorded and generated data has been made. The behaviour of Climak seems quite satisfactory, particularly for the generation of temperature and radiation. Table 5 reports a comparison between annual statistics derived from historical data recorded at Padua (North-East Italy) and those generated by Climak and by another weather stochastic generator (Wxgen, Richardson and Nicks, 1990). The climatic parameters for Wxgen have been estimated by the program Wxparm (http://www.brc.tamus.edu/epic/appendixes/wxparm.html). Climak gives quite satisfactory results, particularly for the temperature and radiation generation, even considering the monthly statistics of generated values. Both the models give very similar annual statistics that are not statistically different from those of the original data (Figures 6, 7, 8 and 9).

While the normal rainfalls are well reproduced by both generators (Figure 9), they are not able (Climak less than Wxgen) to reproduce the extreme values of observed rainfalls. So, other generation methods should be applied to generate those values (for example, sampling from a Poisson pdf).

CONCLUSIONS

Climak was found to be sufficiently accurate in generating meteorological data that represents the climate of a site, even when compared with a well known weather generator.

The minimum number of years needed to give a correct estimate of the parameters depends on the degree of variability of the climate itself and on the specific parameters. For example, rainfall required more years of data than temperature in order to be correctly estimated.

In a previous evaluation (Acutis et al., 1999) it was found that Climak did not reproduce the climate as well as Climgen, when using short series to estimate parameters. This is due to the characteristics of Climak, as it was especially designed to cope with the year-to-year variability. In effect, when using long series to estimate the parameters, Climak is better at reproducing the climate itself (Acutis et al., 1998).

The goodness of a weather model basically depends on its model structure, on methods and algorithms for parameter estimation and on algorithms for data generation (sampling from pdf). The algorithms implementation in the computer program plays a major role in model quality, particularly for the sampling phase. So the weather generator was evaluated for the real performance of the software. At present the model has some limitations. It has not been tested in climates other then the temperate ones.

Table 5. Comparison of annual values (totals, averages and related standard deviations) of meteorological variables recorded, from 1981 to 1988, at the Istituto di Agronomia of Padua (North-East Italy) and those generated (30 years) by Wxgen (Richardson and Nicks, 1990) and Climak The significance of the t-test for the differences between the recorded and generated means and total values are also given.

Meteorological variable	Unit	Recorded data	Wxgen generated		Climak generated	
			total/mean-sd	*P* t-test	total/mean-sd	*P* t-test
Rainfall amount	mm/year	750±97	727±119	0.62	718±117	0.48
Rainy days	n/year	85±11.6	86.2±10.8	0.78	81.7±9.8	0.42
Minimum temperature	°C	7.67±0.58	7.57±0.32	0.51	7 .42±0.63	0.32
Maximum temperature	°C	16.83±0.56	16.76±0.40	0.68	16.81±0.67	0.93
Solar radiation	$MJ/m^2/d$	10.18±0.32	10.17±0.28	0.93	10.27±0.44	0.60

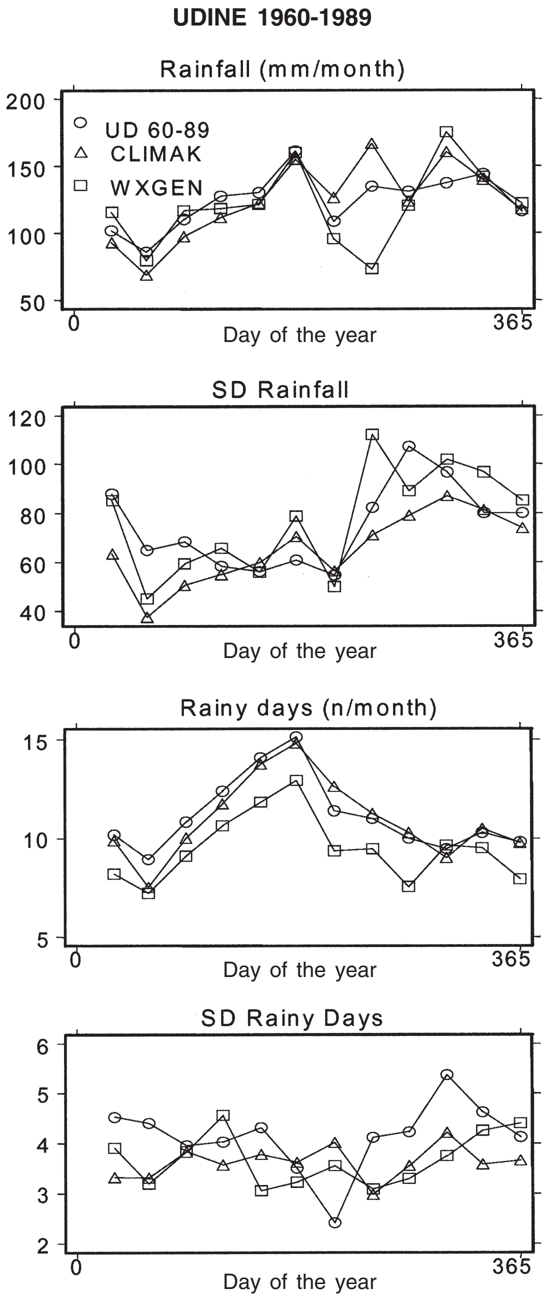

Figure 6. Comparison among and monthly rainfall amount (mm) and monthly number of rainy days and the related standard deviation (*SD*), as recorded at Udine from 1960 to 1989 (circle), generated by Climak (triangle) and by Wxgen (square) after estimating the specific climatic parameter on Udine meteo data.

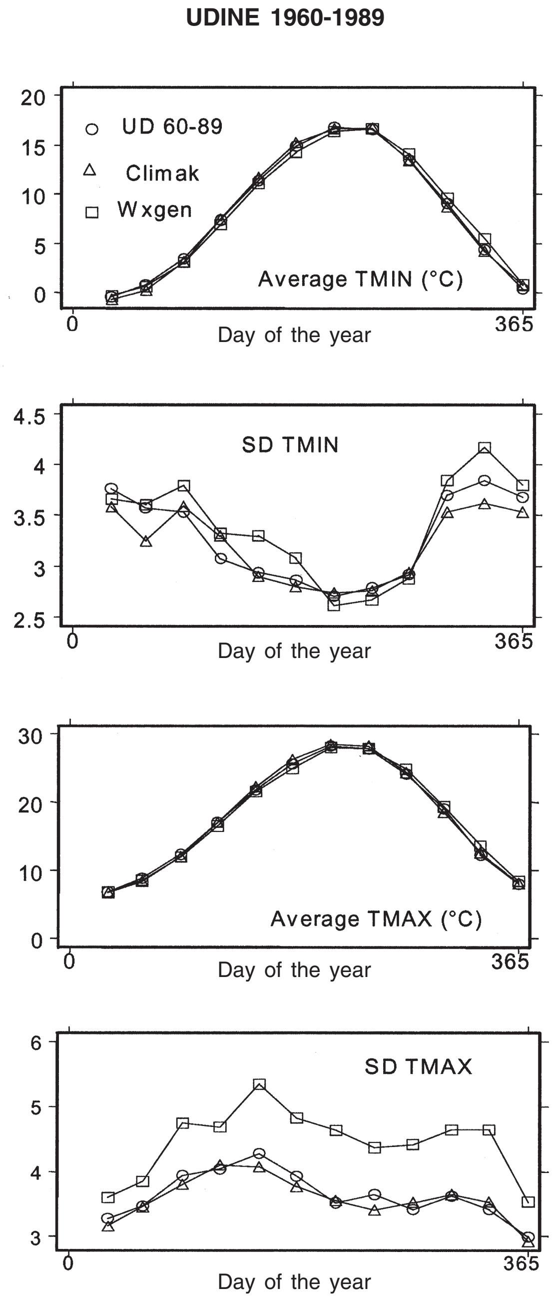

Figure 7. Comparison among minimum and maximum temperature and the related standard deviations (*SD*), as recorded at Udine from 1960 to 1989 (circle), generated by Climak (triangle) and by Wxgen (square) after estimating the specific climatic parameter on Udine meteo data.

So care should be taken when using the weather generator in such situations.

The model is "unit-free" for rainfall, radiation and evapotranspiration but temperature must be in Celsius degrees. This is because absolute values of thermal excursion are used in radiation calculation. Instead, the units or rainfall, radiation and evapotranspiration can be other than those used in this study. This aspect will be fixed in the next versions. Spatial co-ordinates

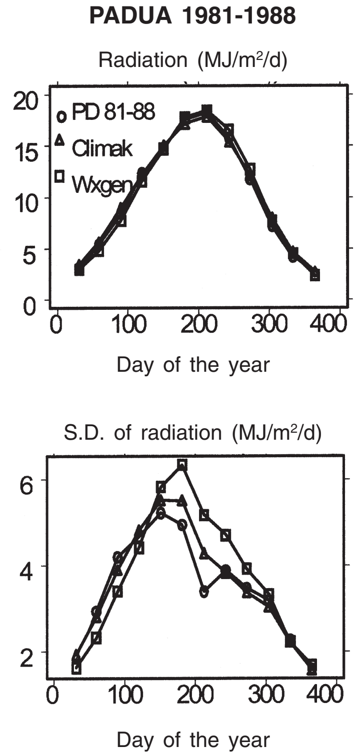

Figure 8. Comparison among daily solar radiation and the related standard deviations (SD), as recorded at Padua from 1981 to 1988 (circle), generated by Climak (triangle) and by Wxgen (square) after estimating the specific climatic parameter.

(latitude and longitude) are also required as degrees and decimals because latitude is used for the photoperiod calculation.

In general, the behaviour of the model is considered satisfactory but some aspects could be improved and inserted into the next versions of Climak: a) generation of other variables such as dewpoint temperature, wind speed, etc.; b) the variables (especially, temperature and rainfall) at a higher time resolution (hourly values, within-storm rainfall intensity, etc.); c) account for spatial correlation in order to generate realistic data on a regional or watershed bases; d) developing methods to spatialise the climatic parameters in hill or mountain conditions; e) a link to global circulation models (GCM) which can supply parameters for the weather generator derived from climate scenario forecasting; f) the use in the model of multi annual trends, related, for example, to the solar spot cycle or to the El Niño Southern Oscillation (ENSO, Woolhiser et al., 1993) to obtain the climatic parameters, thus providing also a sort of climate forecasting; g) the creation of regional georeferenced database of climate parameters of Climak, particularly for hill conditions.

ACKNOWLEDGEMENTS

Work supported by the Italian Ministry of the University and of the Scientific and Technological Research (MURST - 60%). The author wishes to thank Vincenzo Della Mea for his help in developing and programming the first version of Climak. The former Istituto di Agronomia e Coltivazioni Erbacee of the University of Padua and the Ufficio Idrografico del Magistrato alle Acque are also gratefully acknowledged for supplying the meteorological data used for the model evaluation.

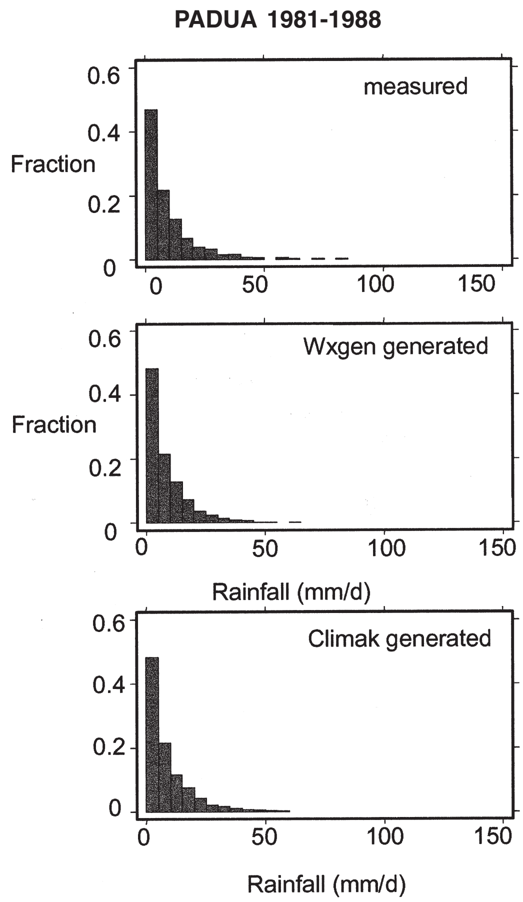

Figure 9. Rainfall amount distribution as recorded at Padua from 1981 to 1988, generated by Climak and Wxgen for the same site.

REFERENCES

Acutis M., Donatelli M., Stöckle C., 1998. Comparing the performance of three weather generators. Proceedings of the 5th ESA Congress, 117-118. Nitra, Slovak Republic.

Acutis M., Donatelli M., Stöckle C., 1999. Performances of two weather generators as a function of the number of available years of measured climatic data. In: Donatelli M., Stöckle C., Villar J.M., Villalobos F. (eds.): Proc. Int. Symposium “Modelling cropping systems”. Lleida, Spain, 21-23 June.

Ahrens J.H, Dieter U., 1972. Computer methods for sampling from gamma, beta, poisson and binomial distributions. Computing, 12, 223-246.

Bratley P., Bennet L.F., Schrage L.E., 1983. A guide to simulation. Springer-Verlag, New York.

Coe R., Stern R.D., 1982. Fitting models to daily rainfall data. J. Appl. Meteorol., 21, 1024-1031.

Danuso F., Della Mea V., 1994. Development and evaluation of a stochastic model for the generation of weather data. III ESA Congress, Abano, Italy, 18-22 September.

Danuso F., Ceschia M., Bortolussi S., Della Mea V., Giovanardi R., 1997. A database of climatic parameters for stochastic weather generation in the region of Friuli-Venezia Giulia. Proc. 14th Int. Congress of Biometeorology, 361-375. Ljubljana, Slovenia, 1-8 September.

Doorembos J., Pruitt W.O., 1977. Guidelines for predicting crop water requirement. FAO Irrigation and Drainage Paper n. 24. FAO, Rome.

Fishman G.S., 1978. Principles of discrete event simulation. J. Wiley & Sons, New York.

Greenwood J.A., Durand D., 1960. Aids for fitting the gamma distribution by maximum likelihood. Technometrics, 2, 55-65.

Hanson C.L., Cumming K.A., Woolhiser D.A., Richardson C.W., 1994. Microcomputer program for daily weather simulation. US Dept. Agric., Agric.Res. Svc. Pub. No. ARS-114, 38 pp.

Johnson G.L., Hanson C.L., Hardegree S.P., Ballard E.B., 1996. Stochastic weather simulation: Overview and analysis of two commonly used models. J. Appl. Meteor., 35, 1878-1896.

Jones J.W., Colwick R.F., Threadgill E.D., 1970. A simulated environmental model of temperature, evaporation, rainfall and soil moisture. ASAE Paper n. 70-404.

Jones P.G., Thornton P.K., 1993. A rainfall generator for agricultural applications in the tropics. Agric. For. Meteorol., 63, 1-19.

Keisling T.C., 1982. Calculation of the length of day. Agron. J., 77, 500-505.

Larsen G.A., Pense R.B., 1982. Stochastic simulation of daily climatic data for agronomic models. Agron. J., 74, 510-514.

Parlange M.B., Katz R.W., 2000. An extended version of the Richardson model for simulating daily weather variables. J. Appl. Meteorol., 39, 610-622.

Richardson C.W., Nicks A.D., 1990. Weather generator descriptor. In: EPIC - Erosion Productivity Impact Calculator. 1. Model documentation. USDA, Tec. Bull. 1768, Temple, TX.

Richardson C.W., Wright D.A., 1984. WGEN: a model for generating daily weather variables. US Dept. Agric., Agric. Res. Svc. Pub. No. ARS-8, 83 pp.

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

Richardson C.W., 1985. Weather simulation for crop management models. Trans. ASAE, 28, 1602-1606.

Shu Geng, Penning de Vries F.W.T., Supit I., 1985. Analysis and simulation of weather variables. Part II: temperature and solar radiation. Simulation Reports CABO-TT n. 5.

Stata Corporation, 1993. Stata Reference Manual: Release 3.1. 6th ed. College Station, TX.

Stöckle C.O., Nelson R.L., Donatelli M., Castellvì F., 2001. ClimGen: a flexible weather generation pogram. Proc. 2nd Int. Symp. Modelling Cropping Systems, 229-230. Florence, Italy, 16-18 July.

Tadikamalla P.R., Johnson M.E., 1981. A complete guide to gamma variate generation. Am. J. Math. Man. Sci., 1, 213-236.

Thornton P.E., Running S.W., White M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. J. Hydrol., 190, 214-251.

Wilks D.S., 1992. Adapting stochastic weather generation algorithms for climate change studies. Clim. Change, 22, 67-84.

Woolhiser D.A, Keefer T.O., Redmond K.T., 1993. Southern oscillation effects on daily precipitation in the southwestern U.S. Water Res. Res., 29, 1287-1295.

CLIMAK: UN MODELLO STOCASTICO DEL CLIMA

RIASSUNTO

SCOPO. Molte attività umane che coinvolgono decisioni strategiche e di progettazione dipendono dalle condizioni climatiche. I modelli stocastici del clima consentono la generazione di dati meteorologici per esegue simulazioni Monte Carlo con modelli di deterministici (ad esempio, modelli colturali), classificare i climi; valutare gli scenari ambientali dovuti ai cambiamenti climatici con procedure "what if" e spazializzare i parametri climatici per aree lontane dalle stazioni meteorologiche. Viene presentato un modello stocastico del clima (Climak), sviluppato per generare sequenze giornaliere di dati meteorologici da impiegare come input per i modelli agro-ecologici, e, in particolare, per le analisi probabilistiche e di rischio.

METODO. Climak ha una struttura simile ad altri "weather generators" ma tiene conto di direttamente della variabilità fra gli anni. Climak genera per primo l'evento piovoso (giorno secco o piovoso) e la quantità di pioggia per i giorni piovosi. Successivamente vengono generate la temperatura minima e massima dell'aria, con parametri diversi per i giorni piovosi o secchi. La radiazione solare è ottenuta dal fotoperiodo astronomico e dalla escursione termica giornaliera. L'evapotraspirazione (di riferimento o potenziale) è generata dalla radiazione solare se disponibile; se no, viene ottenuta dal fotoperiodo e dalla temperatura massima giornaliera.

RISULTATI. Il modello è stato valutato sui dati meteorologici ottenuti in alcune località del nord Italia, in confronto con il modello Wxgen, per la capacità di riprodurre le proprietà climatiche. Il comportamento di Climak è risultato soddisfacente. È stata eseguita anche una valutazione del numero minimo di annate meteorologiche necessarie a stimare correttamente i parametri climatici.

CONCLUSIONI. Climak è risultato accurato nella generazione di dati meteorologici di una determinata località, anche se confrontato con un generatore climatico ben noto ed usato da tempo. Al momento il modello presenta alcune limitazioni: per esempio, non è stato sottoposto a test in climi diversi da quelli temperati. Altri aspetti saranno migliorati nelle prossime versioni del programma.

Parole chiave: clima, modelli stocastici, generatore climatico, software, dati meteorologici.