Implementation and validation of Climak 3 weather generator

Alvaro Rocca¹, Oxana Bashanova¹, Fabrizio Ginaldi², Francesco Danuso¹*

Abstract: Weather generators (WG) are stochastic models, which generates series of weather data of indefinite length with statistical properties similar to those of the original series. WG have been extensively used in different biophysical models, providing them with meteorological input data. Climak 3 is a new version of Climak (Danuso, 2002), capable to generate daily meteorological data of precipitation, minimum and maximum air temperatures, solar radiation, reference evapotranspiration and wind speed. The performance of Climak 3 was tested using meteorological datasets coming from different locations over the world. The results for Italy, Bulgaria and Argentina are presented and discussed.

Keywords: weather generators, stochastic models, validation, climate change.

INTRODUCTION

Climate is one of the main factors which affect human activities and different ecological processes. Great efforts have been devoted to weather forecasting investigations. The study of the climate statistical properties has allowed the development of climatic stochastic models (weather generators) for the generation of weather data (Jones et al., 1970; Richardson, 1981; Larsen and Pense, 1982; Shu Geng et al., 1985; Richardson and Nicks, 1990; Semenov et al., 1998; Donatelli et al., 2005; Donatelli et al., 2009; Birt et al., 2010). Weather generators (WG) are stochastic models, which produce meteorological data of indefinite length, on the base of climatic parameters estimated from historic meteorological data series. Application of weather generators allows Monte Carlo simulations to obtain probability distributions of agro-ecological variables related to climate, spatial interpolation of the climate parameters (thus obtaining data for locations not covered by meteorological stations) and assessment of environmental scenarios depending on climatic changes.

In this paper the third version of the Climak (Danuso and Della Mea, 1994; Danuso et al., 2011) weather generator is presented. Climak was developed in the early 90s and provided significant results (Acutis et al., 1999; Danuso, 2002). Initially Climak generated daily data of precipitation, maximum and minimum temperatures, solar radiation and evapotranspiration. The new version (Climak 3), developed jointly with the weather generator CLIMA (Donatelli et al., 2005; Donatelli et al., 2009), allows also the generation of wind speed data. This version has been developed and implemented using the SEMoLa language (Danuso, 2003). To validate Climak 3, generated meteorological data series were compared with the historical ones.

MATERIALS AND METHODS

Model description

Climak 3 has a structure similar to other weather generators (Jones et al., 1970; Richardson, 1981; Larsen and Pense, 1982; Shu Geng et al., 1985; Richardson and Nicks, 1990; Semenov et al., 1998; Donatelli et al., 2005; Donatelli et al., 2009; Birt et al., 2010). It generates daily total precipitation (Prec), daily minimum and maximum air temperatures (Tmin, Tmax), daily integral of solar radiation (Rg), evapotranspiration (ETr) and daily wind speed (Winds) (Tab. 1). For the evapotranspiration, this could be generated from real measured evapotranspiration or from...
calculated potential or reference evapotranspiration (Allen et al., 1998), depending on which type of evapotranspiration parameters have been estimated. The weather generation procedure consists of 1) estimation of climatic parameters from historical meteorological data, and 2) data generation based on the statistical parameters obtained (Fig. 1). In Climak 3 precipitation are not distinguished between solid (hail, snow) and liquid (rain) precipitation. As a first step, Climak 3 generates the occurrence of rainy or dry day and the rainfall amount, if the day is rainy. After rainfall generation, minimum and maximum air temperatures are generated, separately, for rainy and dry days. Solar radiation is obtained from the astronomical photoperiod (Ph) and from the daily thermal excursion. The evapotranspiration is generated from the solar radiation data; if data of solar radiation are not available, evapotranspiration is obtained from photoperiod and maximum temperature. In the end, wind speed values are generated (Fig. 2).

The state of the day (rainy or dry), being a stochastic process, is represented by a first order Markov chain according to the dry to dry (Pdd), rainy to dry (Prd), dry to rainy (Pdr) and rainy to rainy (Prr) transition probabilities. The transition probabilities parameters were previously estimated from historical data, for each month, as:

\[
Pdd = \frac{Ndd}{Nd} \quad Prd = 1 - Pdd \quad Pdr = Nrd/Nr \quad Prr = 1 - Prd
\]

where:

Ndd number of dry days in the month preceded by a dry day.

<table>
<thead>
<tr>
<th>Meteorological variable</th>
<th>Abbreviation</th>
<th>Unit</th>
<th>Model parameters*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>Prec</td>
<td>mm</td>
<td>Pdd, Prd, Ag, Bg,</td>
</tr>
<tr>
<td>Minimum temperature</td>
<td>Tmin</td>
<td>°C</td>
<td>A, B, C, D, E, Rn, SRn, RRnn,</td>
</tr>
<tr>
<td>Maximum temperature</td>
<td>Tmax</td>
<td>°C</td>
<td>A, B, C, D, E, Rx, SRx, RRx,</td>
</tr>
<tr>
<td>Solar radiation</td>
<td>Rg</td>
<td>MJ/m²·d</td>
<td>b_0, b_1, Ab_0, Bb_1,</td>
</tr>
<tr>
<td>Evapotranspiration</td>
<td>Etr</td>
<td>mm</td>
<td>a_0, a_1, S_w, c_w, c_1, S_an, d_0, d_1,</td>
</tr>
<tr>
<td>Wind speed</td>
<td>Winds</td>
<td>m/s</td>
<td>bw_0, bw_1, bw_2, bw_3, bw_4, bw_5, Rw</td>
</tr>
</tbody>
</table>

*– month; see text for the meaning of the symbols

Tab. 1 - Meteorological variables considered by Climak 3.
Tab. 1 - Variabili meteorologiche considerati in Climak 3.

Fig. 1 - Application of the Climak 3 weather generator.
Fig. 1 - Applicazione del generatore climatico Climak 3.

Fig. 2 - Procedure of generation of meteorological variables.
Fig. 2 - Procedura di generazione delle variabili meteorologiche.
Nd total number of dry days in the month
Nrd number of dry days in the month preceded by a rainy day
Nr total number of rainy days in the month

The rain event is generated by sampling, for each day, 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 probabilities are specific for each month.

If the day is at a rainy status the rainfall amount ($Prec$) is sampled from a Gamma probability density function, considering the threshold value of the instrumental sensitivity ($Sthr$, usually 0.2 mm):

$$Prec = \Gamma(A, B)$$

where $A$ and $B$ are the parameters, specific for each month estimated from historical date.

After rainfall generation, the minimum and maximum temperatures are generated separately, considering the status of the day (rainy or dry):

$$\text{Temperature} = \text{Trend} + \text{Residue}$$

where $\text{Trend}$ is an average daily minimum/maximum temperature for the dry/rainy days, obtained as a function of the date, by interpolating a second order Fourier series:

$$\text{Trend} = A + B \cdot \sin \left( \frac{\text{Doy} - C}{365} \cdot \frac{2 \pi}{365} \right) + D \cdot \sin \left( \frac{\text{Doy} - E}{365} \cdot \frac{4 \pi}{365} \right)$$

where $A$ average annual minimum temperature (°C); $B$ semi-amplitude of the first term (°C); $C$ phase shift for the first term (days); $D$ semi-amplitude of the second term (°C); $E$ phase shift for the second term (days); $\text{Doy}$ day of the year (from 1 to 365 or 366).

Parameters $C$ and $E$, estimated from historical data, are considered constant for all years because of the small variability observed, while means and standard deviations of $A$, $B$ and $D$ parameters are different in relation to the year and for the minimum/maximum and rainy/dry temperature combinations ($Tmin$ trend for dry days, $Tmax$ trend for dry days, $Tmin$ trend for rainy days, $Tmax$ trend for rainy days). These parameters were estimated by linear regression of the trend function (after linearization) of the observed temperatures vs. day of the year. Thus, for each year the annual trends of minimum and maximum air temperature on dry and rainy days are calculated.

During generation these parameter were used for sampling from the normal probability distributions $N(MA, SA)$, $N(MB, SB)$ and $N(MD, SD)$ (where $MA$, $MB$ and $MD$ are the mean values of $A$, $B$ and $D$; $SA$, $SB$ and $SD$ are the standard deviations) at the beginning of each new year.

After the generation of the trend, month by month, temperature is generated by adding the Residue to the trends obtained. Residues for minimum temperature ($Rn$), specific for each month, are sampled from the autocorrelated normal distribution with mean zero and standard deviation $SRn$:

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

where $RRmn$ is the autocorrelation coefficient, $R1n$ the residue of minimum temperature of the previous day, already generated and $N(0,1)$ the value sampled from a normal distribution with 0 mean and 1 standard deviation.

Residues for maximum temperature ($Rx$), also specific for each month, are sampled from the bivariate normal distribution with mean 0, standard deviation $SRx$ and correlation coefficient $RRnx$, depending on the value of the minimum temperature residue $Rn$:

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

$SRn$, $RRmn$, $SRx$ and $RRnx$ parameters were estimated from the historical data.

Daily solar radiation is calculated on the base of the air temperature excursion as:

$$Rg = Rmax \cdot Rr$$

where $Rmax$ is the annual trend of the maximum daily radiation, linearly related to the duration of
the photoperiod ($Ph$) and considered constant for each day of the year. This is performed with the method described in Keisling (1982). The parameters of the linear relation between $R_{max}$ and $Ph$ are obtained by selecting only the maximum values of the solar radiation in ten-day periods of the year.

$$R_{max} = b_1 \cdot Ph + b_0$$

The ratio of the daily radiation and maximum radiation ($R_r = \frac{R_g}{R_{max}}$) is the atmosphere transmittance, which varies from 0 to 1. This ratio is then divided into five air temperature excursion classes, within which it is found to be distributed according to the Beta probability distribution function (pdf), with parameters $A_b$ and $B_b$, estimated from historical datasets. For each class and from the ratio $R_r$ the two parameters of the Beta distribution are estimated, using the moments:

$$A_b = M^2 \cdot \frac{1-M}{V} - M$$
$$B_b = A_b \cdot \frac{1-M}{M}$$

where $M$ and $V$ are the mean and variance of $R_r$, for each excursion class. After rainfall generation, the minimum and maximum temperatures are generated separately, considering the status of the day (rainy or dry).

The evapotranspiration shows the well-known good linear relation with the radiation (Doorembos and Pruitt, 1977); less good is that one with maximum air temperature and photoperiod. Since radiation data are often not available in the historical meteorological datasets, two different approaches for the evapotranspiration generation are adopted. The first one if radiation is available with a more precise generation and the second using temperature and photoperiod:

1) With solar radiation data available: daily evapotranspiration is obtained as a linear function of the daily radiation ($R_g$) plus a residue obtained from a normal distribution (unique for all the months) with standard deviation $Setr$:

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

2) If radiation data are not available: daily evapotranspiration is generated as a function of maximum air temperature ($T_{max}$) and photoperiod ($Ph$):

$$ETr = c_1 \cdot T_{max} \cdot Ph^2 + c_0 + N(0, Setp)$$

where $Setp$ is the standard deviation of the residues, related to the photoperiod by a linear function, $Setp = d_1 \cdot Ph + d_0$.

Wind speed variability has been often expressed using Weibull density function (Takle and Brown, 1977; Weisser and Foxon, 2003; Aksoy et al., 2004; Bhattacharya, 2010). This approach is applied widely and considered as a standard. Despite this, Weibull probability density function is not able to properly represent the minimum values of wind speed (Weisser and Foxon, 2005). Moreover, this approach is not able to correctly describe the annual trend and the monthly distribution of the historical data.

To address these limitations, in Climak 3, a new approach for generating wind speed data was implemented. Daily data of average wind speed ($Winds$) are generated considering four aspects defined through the analysis of meteorological series: wind speed data have an asymmetric distribution (of a logarithmic type), the historical records show the presence of an annual trend, the residues distribution vary from month to month and resulted to be auto-correlated with those of previous days. Thus, the model of wind generation was developed using logarithmically transformed data interpolating the trend with a third-degree polynomial function:

$$LWs = bw_0 + bw_1 \cdot Doy + bw_2 \cdot Doy^2 + bw_3 \cdot Doy^3$$

where $bw_0, \ldots, bw_3$ are the parameters, estimated based on historical data. Then, the wind speed data are obtained as:

$$Winds = \exp(LWs + Rw)$$

$Rw$ is the residue from trend, obtained from the bivariate normal distribution of residues, autocorrelated with the residue of the previous day. In the weather generator, different types of pseudo-random number generators can be used: a simple but faster LCG (linear congruential generator) or the Mersenne Twister series (32, 53 or 64 bit) having a longer period (Matsumoto and Nishimura, 1998). The 32 bit Mersenne Twister pseudo-random number generator has a period of $2^{19937} - 1$. The procedures for parameter
estimation and model validation are implemented as scripts of the SEMoLa framework. Both are completely open, easy to modify and freely available.

**Case study**

The performance of Climak 3 was evaluated using meteorological datasets from different locations of Europe and South-America. Relatively long records of daily weather variables (minimum and maximum air temperature, precipitation, solar radiation, evapotranspiration) were provided by Joint Research Center (EU) and Regional Meteorological Service of the Friuli Venezia Giulia region (OSMER). Wind speed data were available only in datasets from Italy. This data allowed to test Climak 3 performance in different climatic conditions. In this paper, results for mediterranean, moderately continental and subtropical austral climates are presented. The datasets from meteorological stations of Italy, Bulgaria and Argentina were used (Fig. 3). The chosen sites are characterized by different meteorological conditions (Tab. 2). The number of years of the data ranged from 10 for Italy to 20 for Argentina and 35 for Bulgaria.

**Validation**

The goodness of generated weather data depends on the model itself and the quality of the parameters estimation. This, in turn, depends on the calculation methods and on the number of years of historical data available. Weather generators are supposed to generate synthetic weather series with statistical properties similar to the observed ones (Semenov et al., 1998; Donatelli et al., 2005). This means that: (i) means and variances of daily synthetic weather data are to be not significantly different from those of observed series, (ii) means and variances of monthly values of the weather variables from synthetic and observed series are to be comparable, (iii) synthetic weather series are to follow the probability distribution statistically not different from the historical ones.

To test the performance of Climak 3 a validation procedure has been developed. The procedure consists of different graphical analysis evaluating the correspondence of historical and generated data (Semenov et al., 1998; Hayhoe, 2000; Birt et al., 2010), comparing monthly means and monthly standard deviations of all meteorological variables. Rainfall was also compared using relative frequency histograms. Cumulative probability functions were used for graphical representation of the correspondence between historical and generated data distribution.

Validation was performed using generated datasets of 100 years. A long synthetic series provided stable statistical properties thus ensuring that any significant difference between the observed series and the synthetic series is not a result of sampling, as the observed series is only a part of the ‘real’ stochastic process (Qian et al., 2004). Quantile-quantile (QQ) plots are also used to demonstrate visually how well the generated series followed the probability distribution of the observed

### Table 2 - Geographical coordinates and elevation of the study areas.

<table>
<thead>
<tr>
<th>State/Region</th>
<th>Latitude (°)</th>
<th>Longitude (°)</th>
<th>Elevation, m.a.s.l.</th>
<th>Ecoregion division</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy, Friuli V.G.</td>
<td>46.00</td>
<td>13.00</td>
<td>100</td>
<td>Mediterranean</td>
</tr>
<tr>
<td>Bulgaria, Vratsa</td>
<td>43.23</td>
<td>23.52</td>
<td>115</td>
<td>Moderately continental</td>
</tr>
<tr>
<td>Argentina, La Pampa</td>
<td>-36.25</td>
<td>-63.50</td>
<td>119</td>
<td>Subtropical</td>
</tr>
</tbody>
</table>

**Fig. 3 - Study areas geographical position.**

**Fig. 3 - Posizione geografica delle aree di studio.**
Fig. 4 - Daily maximum air temperature (°C) for historical and generated data for Argentina, Bulgaria and Italy.

Fig. 4 - Temperatura massima giornaliera (°C) storica e generata per Argentina, Bulgaria e Italia.

Fig. 5 - Daily solar radiation (MJ/m²/d) for historical and generated data for Argentina, Bulgaria and Italy.

Fig. 5 - Radiazione solare giornaliera (MJ/m²/d) storica e generata per Argentina, Bulgaria e Italia.
series (Qian et al., 2004), for daily minimum and maximum temperature, radiation, wind speed and precipitation. Rain data include only the values between the 10th and the 90th quantile of the distribution to not consider extreme values.

RESULTS AND DISCUSSION
Using Climak 3, meteorological datasets for different locations were generated. It was expected that the software implementation of Climak 3 would provide realistic output data for different climatic conditions. In Figg. 4, 5 and 6 the results for Italy, Bulgaria and Argentina are presented. In Fig. 7 historical and generated data of wind speed (m/s) are presented only for Italy. From these figures of daily data comparison it is possible to notice that Climak 3 gives satisfactory results. In general, the annual trend of all
Fig. 8 - Monthly mean and standard deviation of maximum and minimum temperatures (°C) for historical and generated data for Argentina, Bulgaria and Italy.

Fig. 9 - Monthly mean and standard deviation of precipitation (mm) of historical and generated data for Argentina, Bulgaria and Italy.
variables is followed. However in the generated data there is a slight overestimation of the evapotranspiration standard deviations, which can be neglected since the difference between values does not exceed 1 mm.

Results of monthly means and standard deviations comparison of meteorological variables are presented in Figg. 8, 9, 10, 11, 12. These figures confirm a good fitness of the synthetic and historical data, thus proving a good performance of Climak 3.
CONCLUSIONS

The goodness of a weather models basically depends on the model structure itself, on methods and algorithms applied for parameter estimation and on algorithms for data generation (sampling from probability distribution function). Validation results obtained show that Climak 3 can be considered as sufficiently accurate tool for the generation of meteorological data in temperate and cold climates. In general, the behavior of the model has been satisfactory but some aspects are still to be improved. A new version Climak 4 is now under development that will address these limitations introducing new algorithms for temperatures and radiation. The further works will be focused on the improvement of the estimation and/or generation procedures of evapotranspiration and radiation data, and on a better representation of the $T_{\text{max}}$ and $T_{\text{min}}$ variability. Moreover, it will be necessary to develop issues concerning downscaling of meteorological variables and the
Fig. 14 - Comparison of cumulative distributions of maximum, minimum temperatures and precipitation for historical and generated data (only for the months of January and July) for Argentina, Bulgaria and Italy.

Fig. 14 - Confronto delle distribuzioni cumulate delle temperature massime, minime e delle precipitazioni per dati storici e generati (solo per i mesi di Gennaio e Luglio) per Argentina, Bulgaria e Italia.
Fig. 15 - Quantile–quantile plot of maximum, minimum temperatures, radiation, precipitation and wind for historical and generated data for Argentina, Bulgaria and Italy.

* Precipitation include only the values between the 10th and the 90th quantile of the distribution.
generation of extreme events, especially for precipitation and wind speed. In fact, wind speed model, at present, is not able to represent high speed values, observed in some locations. Furthermore, a stand-alone application tool with easy to use graphical interface (Climak WG) is being developed in order to allow a simpler use of the weather generator. Future developments will include also the generation of hourly data (Fatichi et al., 2011).

Parameter estimation script, generation model and validation procedure are freely available from authors. The executable and source code of Climak 3, the script for the parameters estimation and that for the validation procedure are freely available from the web (http://www.dpvta.uniud.it/~Danuso/docs/Climatica/Climatica_Home.html).

ACKNOWLEDGEMENTS
Authors would like to acknowledge the JRC (Joint Research Center – European Commission) and the Regional Meteorological Services of the FVG region (OSMER) for availability of the meteorological data. Discussion and suggestions from Marcello Donatelli and Roberto Confalonieri have been particularly useful. This research has been funded by the Friuli Venezia Giulia region, in the framework of the project “Filiere agroenergetiche in FVG: valutazione economica energetica e ambientale” - L.R. N. 26 10.11.2005.

REFERENCES