Journal of Agrometeorology 7 (2) : 231 - 240 (Dec 2005)

Comparative evaluation of two weather generators for Pūnjab

GARGI DAS and S. S. RAY
Agro-ecology and Management Division, ARG
Space Applications Centre (ISRO), Ahmedabad – 380 015

ABSTRACT

Weather generators (WG) are computer models that generate synthetic series of daily or sub-daily resolution weather data at a site conditional on the statistical features of the historically observed climate. In the present study, two WGs, i.e. ClimGen and LARSWG were compared to evaluate their suitability for the area of study, i.e. Amritsar weather station of Punjab. Twenty years of daily data of rainfall, maximum and minimum temperature was used as input and six years' data was used for validation. Evaluation was done on the basis of RMSE (Root Mean Square Error) and r^2 value, between generated and actual data. Both models generated temperature data very close to actual with RMSE ranging between 0.69 to 3.15 and r^2 ranging from 0.89 to 0.99. However, error was high for rainfall generation. In most of the cases, LARSWG provided better accuracy than ClimGen in generating weather data.

Key words: Weather generators, ClimGen, LARSWG, Rainfall, Temperature

Meteorological data are essential input parameters used in studies of crop growth, water quality, water availability, soil erosion, climate change, etc. Use of such data are often hindered due to problems like missing data, data format errors, etc. To combat such problems. Weather simulation models (or weather generators) are often used to generate synthetic daily weather data based on stochastic structure of the meteorological processes. A stochastic weather generator (WG) produces artificial time series of weather data of unlimited length for a location based on the statistical characteristics of observed weather at that location. These types of statistical models are generally developed in two steps, with the first step focusing on the modeling of daily precipitation, whilst the second concentrates on the remaining variables of interest, such as maximum and minimum temperature, solar radiation, humidity and wind speed, which are modelled conditional upon precipitation occurrence. For each month different model parameters are used in order to reflect seasonal variations in both the values of the variables themselves and in their cross-correlations. There are two basic types of stochastic WG - referred to as 'Richardson' (Richardson, 1981; Richardson and Wright, 1984) or 'serial' (Racsko *et al.*, 1991; Semenov *et al.*, 1998) types.

In the present study, two popular WGs, i.e. ClimGen (Roger, 2002) and LARSWG (Semenov, 2002) were evaluated for rainfall and temperature (maximum and minimum) generation for a semi-arid climate, such as Punjab. This work was carried out to evaluate the performance of the two WG models for cropping system sustainability modeling. The work is a part of the Cropping System Analysis for Punjab state (Panigrahy *et al.*, 2003).

MATERIALS AND METHODS

Study area

Amritsar district of Punjab state was considered for the study. The weather characteristics are given in Table 1, which is the average of 20 years daily weather data. Amritsar has a sub-tropical, semi-arid and monsoon type of climate. The maximum temperature ranges from 18.5°C during winter to 39.47°C during the summer months. Due to the monsoon type of climate, most of the rainfall is received during the months of July and August.

Model description

In a 'Richardson' type WG (e.g., ClimGen) precipitation occurrence is modelled using a first-order two-state Markov procedure, which describes two precipitation classes, i.e., wet or dry, and takes into account precipitation occurrence on the previous day only. The Markov process gives information on transition probabilities, e.g., on the probability of a wet day following a dry day or on the probability of a wet day following a wet day, calculated from the observed station data. If precipitation occurs, then the amount of precipitation falling on wet days is determined usually by using a predefined frequency distribution i.e. Weibull distribution (Selkar and Haith, 1990). The remaining climate variables, such as temperature and radiation, are then calculated based on their correlations with each other and on the wet or dry status of each day. The ClimGen software is based on the assumption that temperature is a weakly stationary process (Matalas, 1967). It considers maximum and minimum temperature to be a continuous, multivariate stochastic process with daily means and standard deviation conditioned by the precipitation status (wet and dry) of the day (Richardson, 1981). The time series of each variable (i.e. maximum and minimum temperature) is reduced to a time series of residual elements through the removal of the periodic means and scaling by the standard deviations.

LARSWG (Long Ashton Research Station Weather Generator) is a serial type of WG. In "serial approach" the first step is modeling the sequence of dry and wet series of days. The amount of precipitation and the remaining climate variables are then generated dependent on the wet or dry series by using Semi-empirical distributions. This means that every single observation in the daily station data record is used in the modeling process. The daily minimum and maximum temperature are considered as stochastic processes with daily means and daily standard deviations conditioned on the

Table 1: Monthly average weather parameters of the study site

Months	Maximum temperature (°C)	Minimum temperature (°C)	Rainfall (mm)
January	18.5	3.5	23.0
February	21.6	6.1	33.4
March	25.2	10.3	46.7
April	33.8	16.2	29.2
May	36.7	20.1	23.2
June	39.5	24.6	69.8
July	33.4	24.1	211.1
August	34.2	25.1	181.3
September	32.8	21.1	71.4
October	32.3	15.2	17.6
November	25.8	8.3	5.9
December	21.3	4.4	16.6
Average	29.6	14.9	729.2

wet and dry status of the day. The seasonal cycles of means and standard deviations are modelled by Fourier series of order 3 and the residuals are approximated by a normal distribution. The observed residuals, obtained by removing the fitted mean value from the observed data, are used to analyse a time autocorrelation for minimum and maximum temperatures. For ease, both of these are assumed to be constant through the whole year for both dry and wet days with average value from the observed data being used. Minimum and maximum temperature residuals have a pre-set cross-correlation of 0.6 (Semenov, 2002).

Data used

For our study, we used daily meteorological data (rainfall, minimum and maximum temperature) of twenty-six years (1971 to 1996), collected from weather observatory located in Amritsar city (31.63° N, 74.87°E). The first twenty years' (1971-1990) data was used as input to the models for their calibration and last five years' (1991-1996) data was used for evaluation of the generated data.

METHODOLOGY

For ClimGen, the prescribed format of the input daily weather series has to be prepared year-wise, i.e. one data file for each year. Thus 20 years, from 1971 to 1990, input weather series were prepared. Data was generated for twenty-six years i.e. from 1971-1996. The generated data was then compared against the actual data.

Unlike ClimGen, in LARSWG (version 3.0) the data of all the input years should be contained in a single file. For generating data, we need to choose a random seed value, which controls the stochastic component of LARSWG. There are a number of pre-set random seeds available, which is a prime number and ranges from 500 to 2500. A number of different realizations of weather time series can be generated with the help of these random seed values. These realisations will all have the same statistical characteristics, but they will differ on a day-to-day basis. In the present version, there are only five random seed values (577,677,797,877 and 977). So the application was run five times and the generated data for twenty-six years was recorded in different files. All the generated data were then analysed separately against the original weather data.

Statistical evaluation of the generated data

The generated data obtained from both the generators were analyzed in similar pattern. For ClimGen, the weather series generated was from the year 1971 to 1996. The data taken in consideration for comparison was from 1991 to 1996 only. Similarly, though, in LARSWG the generated series does not give any year number, only the last 6 years of data was taken for the comparative study. In the first step the 6 years generated as well as the actual data was averaged to compute monthly mean data. These monthly means were then compared using RMSE values and r^2 values between actual and generated data. For LARSWG, additionally, the results of different random seed numbers were compared among themselves to arrive at the best generated data with respect to the RMSE values. Then the generated data from the best random seed number of LARSWG was further used for comparing with the generated data from ClimGen.

RESULTS AND DISCUSSION

In both the models we had used input for twenty years (for model calibration) and data was generated for twenty-six years starting from same base year as that of input data. The generated data of ClimGen for first twenty years was exactly same as the input data. However for LARSWG, all generated data were stochastically generated data irrespective of the year. This shows that LARSWG takes input data just to examine the statistical characteristics of the particular place and then generates data with the same statistical characteristics regardless of the year. For evaluation only the last six years' generated data from each model was used and compared against the actual data of 1991-1996.

Best random seed number

From the RMSE values it was found that for rainfall and minimum temperature, seed number 979 was most accurate, whereas, for maximum temperature, seed number 797 showed to be more accurate than others. So, for further comparison between the two generators, only these particular random seed numbers for the respective weather parameters were used in LARSWG.

RMSE values

The patterns of actual and generated

Table 2: RMSE of rainfall, maximum and minimum temperature of Amritsar

Year	Maximum temperature(°C)		Maximum temperature(°C)		Rainfall (mm)	
	ClimGen	LARSWG	ClimGen	LARSWG	ClimGen	LARSWG
1991	1.62	1.41	1.05	1.12	85.97	61.75
1992	2.33	1.13	1.66	1.07	46.25	34.48
1993	1.45	1.91	1.66	1.43	92.53	117.49
1994	1.08	1.17	0.92	0.69	45.07	73.92
1995	2.13	1.99	3.15	2.99	97.85	29.26
1996	1.71	1.55	1.41	1.86	102.30	86.89
Average	1.72	1.53	1.64	1.53	78.33	67.30

values of minimum and maximum temperature and rainfall are shown in Fig.1, which represents the data of 1992, as an example. The patterns shows that generated values of maximum and minimum temperature matched well with the actual values for both the models. However, there was comparatively high difference for maximum temperature in the month of May for ClimGen generated values. ClimGen also showed large deviation in rainfall generation, the maximum difference being in the months of May and July. LARSWG generated curves were relatively closer to the actual data pattern. Similar patterns were also seen for other years. Similar to graphical pattern, RMSE values were low for temperature and high for rainfall (Table 2). As can be derived from the figures, both the generators performed reasonably well. It can be observed that LARSWG performs better than ClimGen in all aspects. Range of RMSE values developed by LARSWG for maximum temperature, minimum temperature and rainfall vary from 1.13 to 1.99, 0.69 to 2.99 and 29.26 to 117.49, respectively whereas, ClimGen showed higher RMSE values of 1.08 to 2.33, 0.920 to 3.15 and 45.07 to 102.30. For all the three parameters the average RMSE for LARSWG generated data was lower than the ClimGen generated data.

Regression analysis

The coefficient of determination of the regression equations between generated and actual data are represented in Table 3 and Fig. 2. While the table shows the results of all six years, the figures present the situation in 1992, as an example. Here also, except for the maximum temperature, where both the generators gave comparable results, LARSWG is superior. In case rainfall there is wide difference between the regression lines and 1:1 lines. The generated rainfall was higher than actual value in low rainfall cases and lower than actual value in high rainfall cases. The regression lines for

Table 3: R^2 values for Rainfall, Maximum and Minimum temperature of Amritsar

Year	Maximum temperature		Maximum temperature		Maximum temperature	
	ClimGen	LARSWG	ClimGen	LARSWG	ClimGen	LARSWG
1991	0.94	0.95	0.98	0.98	0.53	0.042
1992	0.94	0.97	0.96	0.98	0.34	0.53
1993	0.95	0.93	0.96	0.97	0.52	0.17
1994	0.97	0.97	0.98	0.99	0.6	0.85
1995	0.92	0.92	0.96	0.98	0.21	0.72
1996	0.95	0.94	0.98	0.97	0.27	0.47
Average	0.95	0.95	0.97	0.98	0.41	0.46

maximum and minimum temperature matched well with the 1:1 line, being closest for minimum temperature. Similarly the average r^2 values for temperature generation ranged from 0.95 to 0.98, whereas for rainfall generation it was between 0.41-0.46. In majority cases LARSWG provided higher r^2 values than ClimGen for all the three parameters.

From the above comparison, it is evident that rainfall registered the maximum RMSE and least regression coefficient than temperature, for both the generators. The reason, as given by Richardson (1981), may be due to high proportion of zero observation in daily rainfall data and also due to skewed distribution of rainfall besides the higher interannual variability of rainfall compared to temperatures. Also, it can be concluded that LARSWG performed better than ClimGen. It could be possible for LARSWG to generate better rainfall results due to its "serial approach" which models the sequence of dry and wet series of days. The simulation of precipitation occurrence is based on the distributions of the length of continuous sequences, or series, of wet and dry days. This is different from the approach suggested by Bailey (1964) and re-used by Richardson (1981), which applies a first-order Markov chain to describe the occurrence of wet and dry days. The main limitation of the 'Markovian' approach is that the Markov chain has a 'limited memory' of rare events and, for example, could fail to simulate accurately long dry series at certain locations (Racsko *et al*, 1991). Similarly, temperature is also better generated by LARSWG. This might be due to the fact that the generation of temperature and other weather variables are dependant on the generation of rainfall.

The probable reasons for any disagreement between the actual and generated values can be attributed to several reasons. As noticed by Harmel *et al.* (2002), the daily maximum and minimum

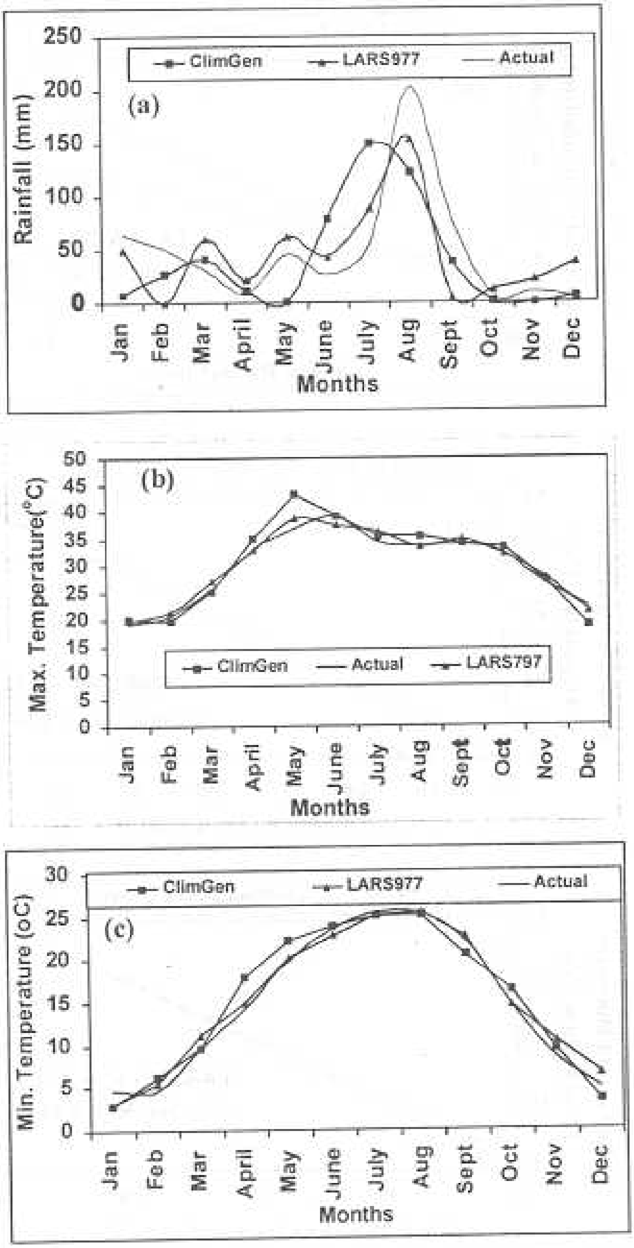

Fig 1: Comparison of weather parameters generated by ClimGen and LARSWG, a) Rainfall b) Maximum temperature c) Minimum temperature

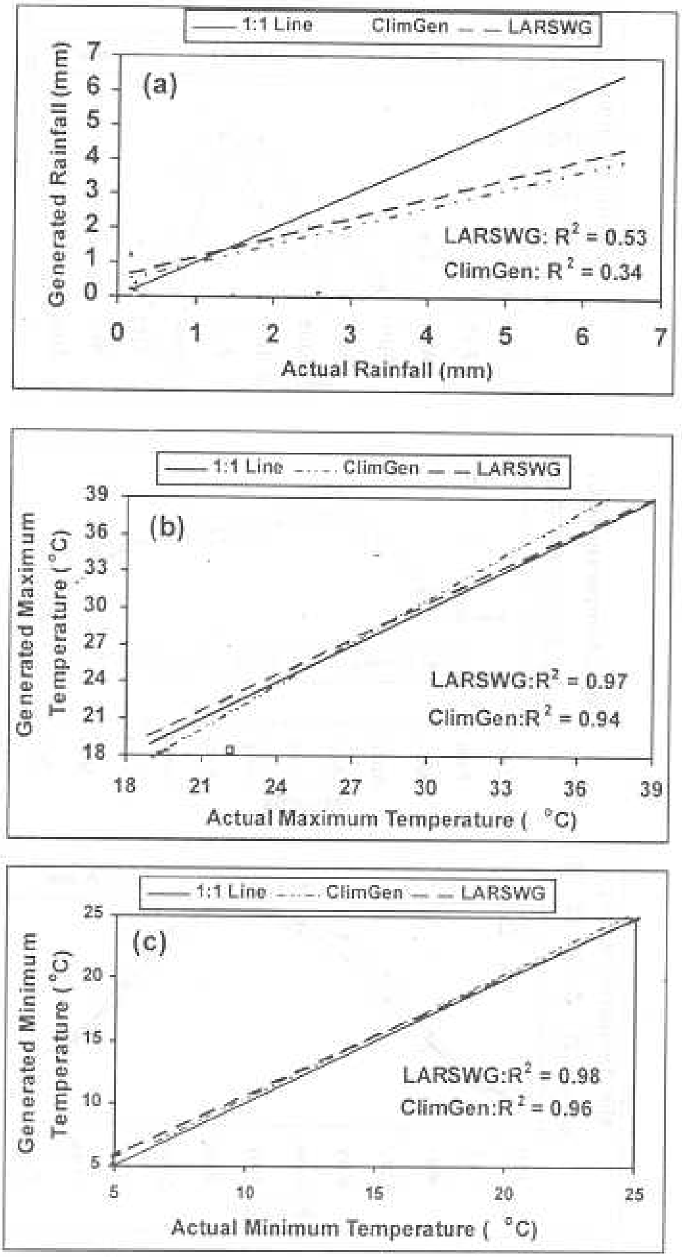

Fig.2: The regression lines between actual and genearted weather parameters by ClimGen and LARSWG, a) Rainfall b) Maximum temperature c) Minimum temperature

temperature are not generally normally distributed in each month but are often slightly skewed which contradicts the assumption of normality used by most WGs. The difference in the generated data of LARSWG can be attributed to the smooth curve fitting of daily mean values for minimum temperature and for maximum temperature (Semenov, 2002), The model has this function to maximize elimination of the random noise in the observed data in order to get closer to the actual climate for the site. Differences are likely to be due to departures of the observed values from the smooth pattern for the data.

ACKNOWLDGEMENT

Authors are grateful to Shri J. S. Parihar, Group Director, Agricultural Resources Group and Dr S. Panigrahy, Head, Agroecology and Management Division, for guidance and support to carry out this work.

REFERENCES

Arnold, C.D. and Elliot, W.J. 1996. CLIMGEN Weather generator predictions of seasonal wet and dry spells in Uganda. *Trans. of ASAE,* 39(3): 969-972.

Goodess, C.M. 2000. The Construction of daily rainfall scenarios for Mediterranean sites using a circulation-type approach to downscaling, Ph.D. Thesis, University of East Anglia, Norwich.

Goodess, C.M. and Palutikof, J.P. 1998. Development of daily rainfall scenarios for southeast Spain using a circulation-type approach to downscaling. *Int. J. Clim.*, 18:1051-1083.

Harmel, R. D., Richardson, C. W., Hanson, C. L. and Johnson, G. L. 2002. Evaluating the adequacy of simulating maximum and minimum daily air temperature with the normal distribution. *J. Appl. Meteorol.*, 41(7) 744–753.

Matalas, N.C. 1967. Mathematical assessment of synthetic hydrology. *Wat. Resour. Res.*, 3(4): 937-945.

Nelson, R. 2002. ClimGen- Climatic data generator: User's Manual.

Panigrahy, S., Ray, S. S., Parihar, J. S., Sood, Anil, Patel, L. B. and Sharma, P. K. (2003) Cropping system Analysis of Punjab State using Remote Sensing and GIS. Scientific Report. RSAM/SAC/CS/SN/04/2003. Space Applications Centre, Ahmedabad.

Racsko, P., Szeidl, L. and Semenov, M. 1991. A serial approach to local stochastic weather models. *Ecol. Modelling,* 57:27-41.

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

Richardson, C.W. and Wright, D.A. 1984. WGEN: a model for generating daily weather variables. *US Department of Agriculture, Agricultural Research Service,* ARS-8, 83pp.

Semenov, M.A. 2002. LARSWG-A Stochastic Weather Generator for Use in Climate Impact Studies. User Manual.

Semenov, M.A., Brooks, R. J., Barrow, E. M. and Richardson, C.W. 1998. Comparison of the WGEN and LARSWG stochastic weather generators for diverse climates. *Clim. Res.*, 10:95-107.

Stöckle, C.O., Campbell, G.S. and Nelson, R.1999. ClimGen manual. Biological Systems Engineering Department, Washington State University, Pullman, WA, 28 p.