African Journal of Environmental Science and Technology Vol. 1 (4), pp. 086-092, November, 2007
Available online at http://www.academicjournals.org/AJest
ISSN 1996-0786 © 2007 Academic Journals

Full Length Research Paper

Assessment of the ClimGen stochastic weather generator at Cameroon sites

Munang Tingem[1]*, Mike Rivington[2], Sayed Azam-Ali[1] and Jeremy Colls[1]

[1]Agriculture and Environmental Science Division, School of Bioscience, University of Nottingham, NG7 2RD, UK.
[2]Macaulay Institute, Craigiebuckler, Aberdeen, AB15 8QH, Scotland UK.

Accepted 2 October, 2007

Simulation of agricultural risk assessment and environmental management requires long series of daily weather data for the area being modelled. Acquiring and formatting this data can be very complex and time-consuming. This has led to the development of weather generation procedures and tools. Weather generators can produce time series of synthetic weather data of any length, interpolating observed data to produce synthetic weather data at new sites. Any generator must be tested to ensure that the data that it produces is satisfactory for the purposes for which it is to be used. The aim of this paper is to test a commonly used weather generator, ClimGen (version 4.1.05) at eight sites with contrasting climates in Cameroon. Statistical test were conducted, including *t*-test and *F*-test, to compare the differences between generated weather data versus 25 years observed weather data. The results showed that the generated weather series was similar to the observed data for its distribution of monthly precipitation and its variances, monthly means and variance of minimum and maximum air temperatures. Based on the results from this study, it can be concluded that ClimGen performs well in the simulation of weather statistics in Cameroon.

Key words: Weather generators, weather data, Cameroon, climate change.

INTRODUCTION

Weather is a key determinant in agricultural production, particularly in rainfed cropping systems commonly found in tropical and arid regions (Tingem et al., 2007). Application of simulation models to represent such agricultural systems normally requires observed long-term daily weather data. These requirements often include observed daily maximum and minimum air temperature and total precipitation (Kuchar, 2004). But the range of observed weather data record at many sites is often insufficient to allow a good estimation of the probability of extreme events. There is therefore a serious limit on the application of agricultural, hydrological and ecosystem simulation models if weather data are not directly available (Hoogenboom, 2000).

Deterministic mathematical models (known as stochastic weather generators) that simulate time series climatic variables have addressed this problem (Richardson and Wright, 1984). These models provide data to augment the existing record at a site or, through interpolation of model parameters, provide climate information where measured data are not available. They have several interconnected components and usually simulate multiple variables using observed historical weather data as inputs and generating synthetic weather data, which are statistically similar to the observed historical weather records (Hoogenboom and Soltani, 2003). Observed time series represents just one realisation of the climate, whereas a weather generator can simulate many realisations, thus providing a wider range of feasible situations.

Weather generators are now widely used by researchers from many different backgrounds in conjunction with their impact models and are becoming a standard component of decision support systems in agriculture, environmental management and hydrology. Generators may be used as supplied, which is dangerous, that is without sufficient validation being carried out for the sites at which they are applied. Testing and validation for loca-

*Corresponding author. E-mail: plxmrt@nottingham.ac.uk.
Tel: +44(0)1158466585. Fax: +44(0)1159516261.

Table 1. Geo-references of the eight study sites.

Location	Latitude (degree)	Longitude (degree)	Elevation (m)	Annual rainfall (mm)
Bamenda	6.05°	10.1°	1239	2378
Batouri	4.47°	14.37°	656	1499
Garoua	9.33°	13.38°	244	1090
Kribi	2.95°	9.89°	16	2634
Maroua	10.44°	14.25°	422	834
Ngaoundere	7.34°	13.57°	1104	1514
Tiko	4.08°	9.37°	52	3198
Yaounde	3.83°	11.51°	760	1655

tions other than those for which they were developed and validated is necessary.

The objective of this study is to test the weather generator, ClimGen (version 4.1.05) (Stöckle et al., 2001) for a range of climates in Cameroon. Validation of this model at Cameroon sites will offer the opportunity to evaluate long-term effects of weather on many environmental processes, such as crop yields, hydrological dynamics and ecosystem functions, which are impossible to evaluate with a limited observed record of historical data.

Description of ClimGen stochastic weather generator

ClimGen is a weather generator that uses principles similar to those in WGEN (Richardson and Wright, 1984), but with significant modification and additions[1].

ClimGen is a daily time step stochastic model that generates daily precipitation, minimum and maximum temperature, solar radiation, humidity and wind speed data series with similar statistics to that of the historical weather data. The model requires inputs of daily series of these weather variables to calculate parameters used in the generation process for any length of period at a location of interest. Further information on ClimGen is well documented elsewhere (Castellvi and Stockle, 2001, Stockle and Nelson, 2003, Stöckle et al., 2001).

Data and choice of sites used in study

Cameroon is a tropical country located in the sub-Saharan region of central-west Africa. The country displays highly contrasting physical and biogeographical features. The climate, reflecting the topography and latitudinal range is very diverse. It comprises two principal climate zones: the equatorial zone and the tropical zone.

The equatorial zone stretches from 2 to 6° N covering the southern and the mountainous western part of the country. Its climate corresponds to the classical Guinean region, with the following subtypes: (1) The seaboard, e.g. Kribi and Tiko with abundant rainfall (2634 and 3198 mm yr^{-1} respectively). (2) The inland areas, e.g. Yaounde with total rainfall < 1660 mm. This climate subtype prevails over the southern part of the south Cameroon plateau, extending into the east of the country around Batouri. (3) North of 6° N, the Sudanese-Sahelian subtype differs from the 'inland' with total rainfall decreasing from 1513 mm a year to 834 mm northward near Lake Chad. The mean temperature ranges from 22° to 29°C increasing from south to north and from the coast to the hinterland. Table 1 show the sites used for the study and identifies the mean differences in the performance of the generator. Appendix A provides further information on the regional climatic differences within Cameroon.

Daily observed values of maximum and minimum temperatures, and rainfall were obtained for 1979-2003 from the University Cooperation for Atmospheric Research (UCAR) (http://dss.ucar.edu/datasets/) for each of the eight sites used in the study. For each region, the data from one of the major weather stations was chosen as representative of the climate of that region.

Comparison tests

The aim in designing weather generators is to produce synthetic weather data which is statistically similar to the observed. The weather generator investigated here was run at eight selected sites and a number of statistical tests (see below), comparing the synthetic and observed data were carried out.

Initially the observed data for a particular site were run through the generator to produce a site universal environmental database (UED) parameter file format. The generator used the file parameter to produce a time series of synthetic data of any length. For each of the eight sites, 25 yr of daily data were generated using ClimGen. Such a long series of data was used so that the statistical properties of the synthetic data would be close to the true distribution seen within the observed data. Monthly means for precipitation, minimum and maximum temperatures were compared using the *t*-test. Var- were compared iances of the monthly mean values for the different

[1]ClimGen uses a Weibull distribution to generate precipitation amounts instead of the Gamma distribution used by WGEN. Selker and Haith (1990) showed the Weibull distribution to be superior to other probability distribution of daily precipitation amount. WGEN uses truncated Fourier series fits to produce daily values for monthly-calculated quantities of mean weather variables. ClimGen uses quadratic spline functions chosen to ensure that the average of the daily values are continuous across month boundaries, and that the first derivative of the function is continuous across month boundaries. Additive features of ClimGen include the generation of vapor pressure deficit (VPD) and wind speed. A temperature-based approach, developed by Bristow and Campbell (1984) is embedded in ClimGen allowing users to estimate solar radiation from existing temperature records in areas where solar radiation input data are not readily available.

Table 2. Results of the statistical tests showing the comparison of the observed precipitation (mm) and maximum temperature (Tmax) (ºC) means and variances with those of 25 year synthetic data generated by ClimGen at eight sites. Probability levels (p-value) calculated by the **t**-test and **F**-test for the monthly means and variance, and percent difference (negative values show model under-estimation) are shown. A probability of 0.05 or lower indicates a departure from the observation that is significant at 5% level.

	Bamenda	Batouri	Garoua	Kribi	Maroua	Ngaoundere	Tiko	Yaounde
Precipitation								
Obs.mean	195.8	123.283	83.1	219.5	65.9	124.7	266.5	135.7
Obs.variance	25185.0	5375.9	8425.6	22020.0	6842.8	12638.0	38782.3	8083.0
Gen.mean	176.1	134.7	88.3	72.3	72.3	126.1	260.3	153.6
Gen.variance	11202.4	8506.1	7600.7	7646.7	7646.4	11874.9	41799.1	7593.5
% difference	-10.1	9.3	6.3	-67.1	9.7	1.1	-2.3	13.2
p-value for ***t***-test	0.724	0.74	0.888	0.00838	0.854	0.976	0.94	0.624
p-value for ***F***-test	0.097	0.229	0.434	0.417	0.429	0.46	0.452	0.46
Tmax								
Obs.mean	23.8	29.6	34.9	30	34.5	29	30	28.4
Obs.variance	2.93	2.5	9.27	2.05	8.81	4	3.25	2.47
Gen.mean	24.7	29.1	34.3	30	33.5	28.1	30.1	27.8
Gen.variance	2.96	3.35	8.89	2	8.08	3.83	3.3	2.33
% difference	3.8	-1.7	-1.7	0.0	-2.9	-3.1	0.3	-2.1
p-value for ***t***-test	0.188	0.457	0.617	0.933	0.447	0.287	0.937	0.329
p-value for ***F***-test	0.493	0.318	0.473	0.484	0.445	0.472	0.49	0.463

A probability of 0.05 or lower indicates a departure from the observation that is significant at 5% level.

years using the *F*-test. The variance value measures the inter-annual variability in the monthly means. These tests are based on the assumption that the observed and synthetic weather data are both random samples from existing distributions and they test the null hypothesis that the two distributions are the same. In the case of observed weather data, such a distribution represents the true climate at the site which would, in the absence of any changes in climate, be the

Distribution of observed data over a very long period. Each test produces a p-value which measures the probability that both sets of data comes from the same distribution. Hence, a low p-value means that the synthetic climate is unlikely to be the same as the true climate and so the generator at that location is probably behaving poorly. A large p-value indicates that the differences are small enough that there is insufficient evidence to reject the null hypothesis.

The percent difference (*E*) between observed (*Obs*) and simulated (*Gen*) mean monthly data was also calculated:

$$E\ (\%) = \frac{Gen - Obs}{Obs} 100$$

Observed and simulated weather data were further compared at three sites (two for precipitation and one for temperature) using exceedence probability (*Pe*, %) distributions, following Weibull (1961) :

$$Pe = \frac{m}{n+1} 100$$

where *m* is the rank order of each weather variable estimate, with *m* = 1 as the largest and *m* = *n* for the lowest, with *n* being the number of observations.

RESULTS

Precipitation and temperature were tested in two ways by comparing: (1) the monthly means using the *t*-test and (2) the monthly variance using the *F*-test. Table 2 summarises the outcome of the series of statistical comparisons for all the test sites. ClimGen performed well in simulating the range of monthly mean precipitation and temperature values at the test sites. The results show that ClimGen was able to reproduce the annual means well for both precipitation and Tmax. For precipitation, only at Kribi did the simulated data show any substantial error. Otherwise, the model produced both over- and under-estimates, but with a mean percent difference of only 3.9% (excluding Kribi), ranging from -10.1 to +13.2%. For Tmax, the results were better, with a mean percent difference of only -0.9%, with an over- and under-estimation range of +3.8 to -3.1%. The model performed particularly well for Tmax at Kribi.

The *t*-test (5% level of significance) indicated there is no significant difference between generated values and observed data apart for monthly means of precipitation at Kribi. The under-estimated monthly mean precipitation at Kribi indicates great care should be exercised in interpreting impact assessment responses obtained in this site from using such weather data as it may have uncertainties pertaining to the above statistics. Possible reasons for this substantial error include issues of the quality of the original (UCAR) precipitation data for Kribi, and the possibility of errors occurring during the UED parameterisation process. Mckague et al. (2003) also noted that

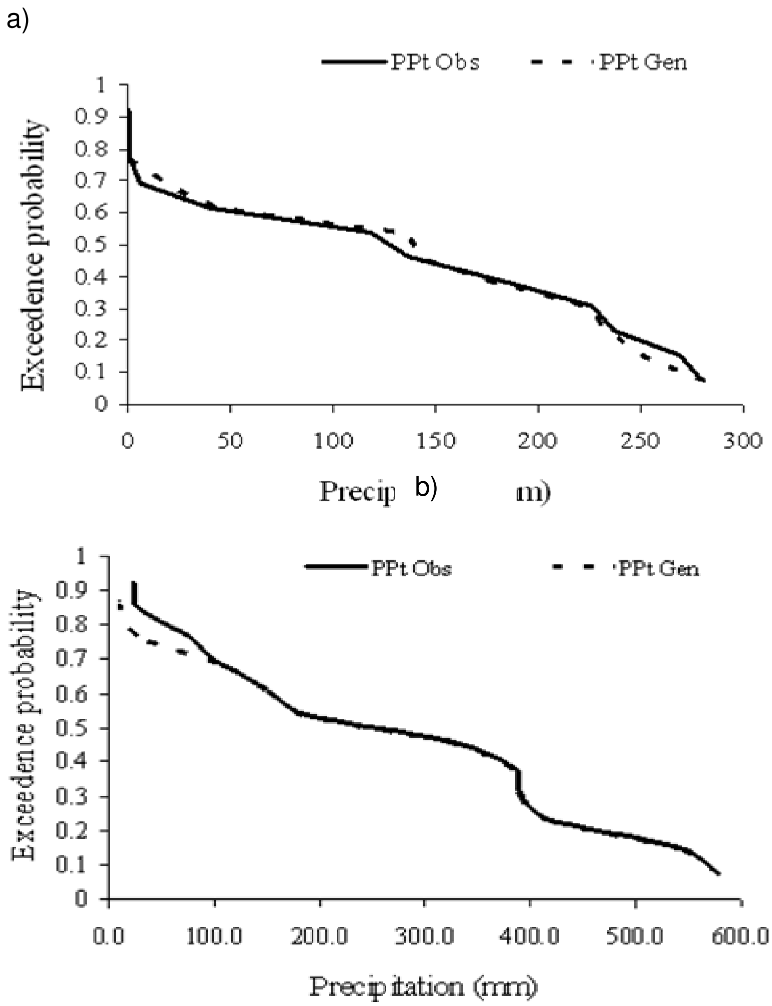


Figure 1. Exceedence probability functions for the distribution of monthly mean precipitation for the observed and generated data by ClimGen (a) monthly mean precipitation in Ngaoundere (b) monthly mean precipitation in Tiko.

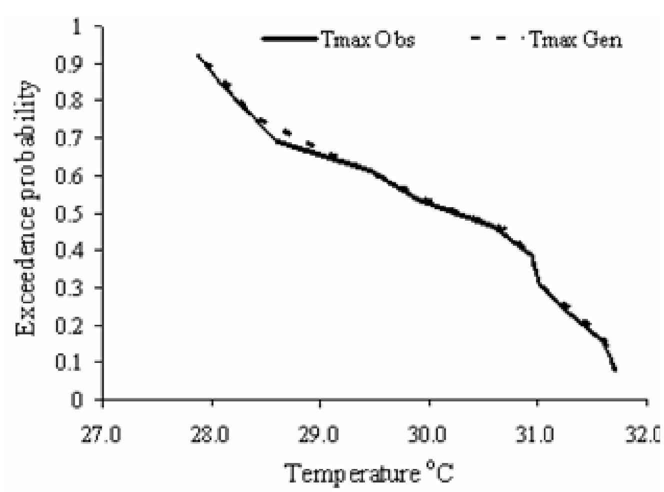
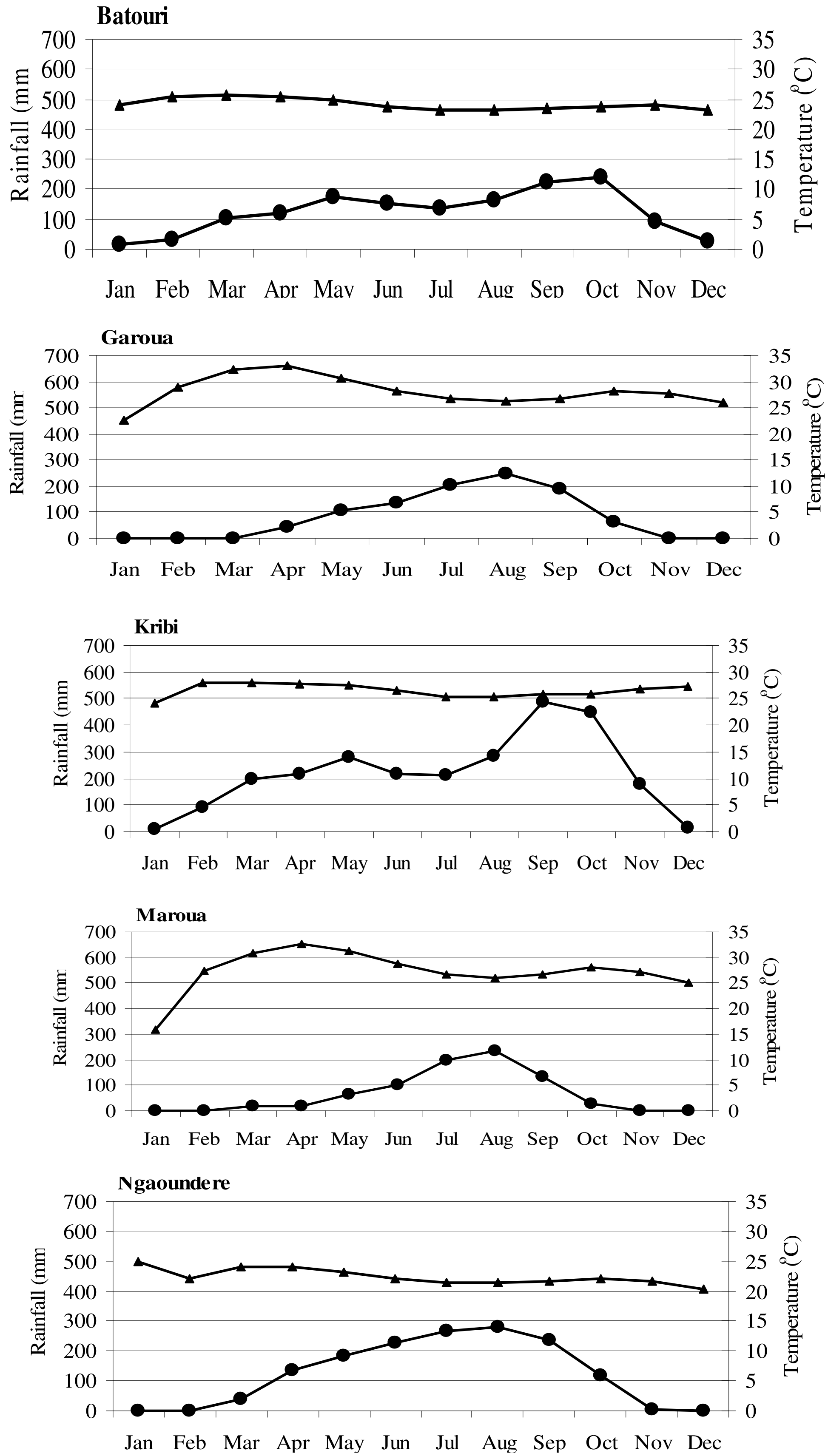


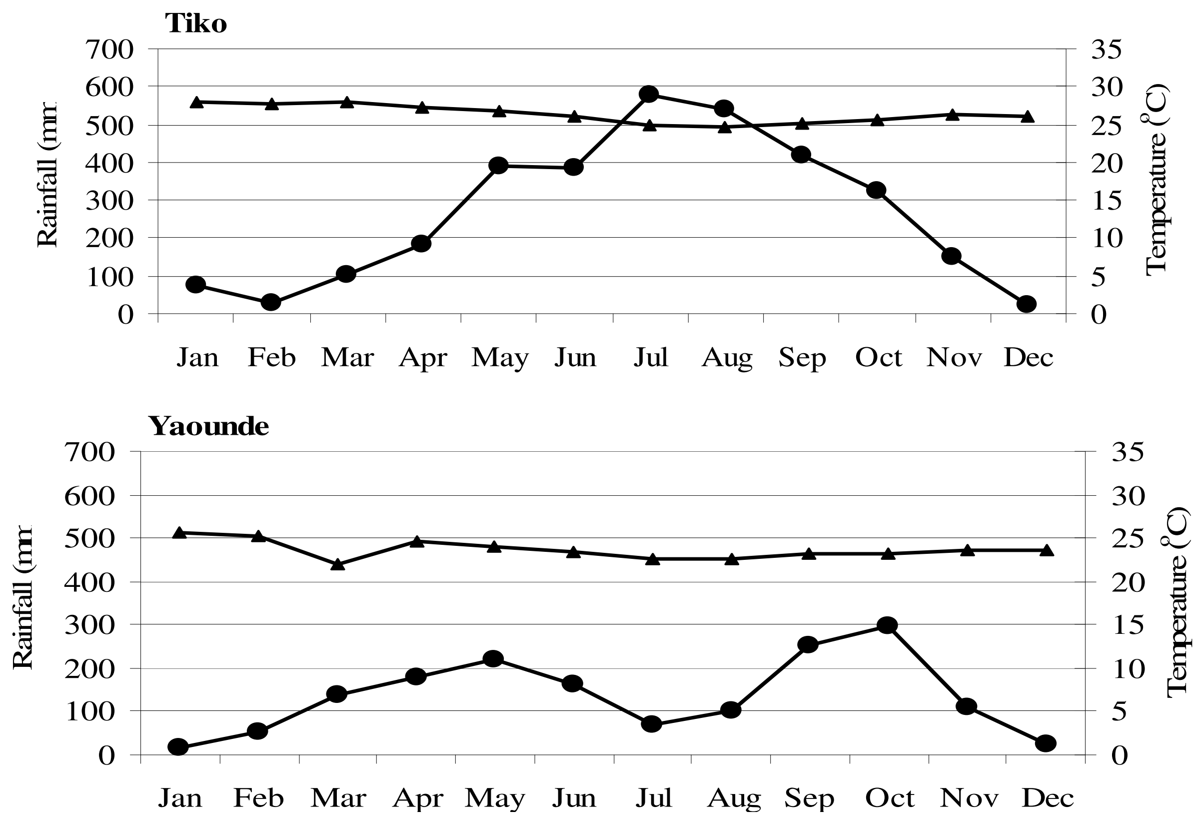
Figure 2. Exceedence probability functions of monthly Tmax for observed data and synthetic generated in Kribi.

errors can occur during the generation period affecting particularly the precipitation patterns, but with the use of Weibull distribution and spline function in ClimGen, the fluctuation among seasons is minimized.

The variance of the generated means of monthly precipitation and maximum temperature was not significantly different from the observed for all the sites indicating that the monthly variation was well reproduced by ClimGen. Further evidence of good model performance comes when one compares observed and generated monthly precipitation and maximum temperatures in Ngaoundere, Tiko and Kribi using exceedence probability graphs (Figure 1 and 2). For precipitation at Ngaoundere, there is a very close match between observed and simulated data across the whole range of monthly means. The same is true for Tiko, except for the drier months (those with < 100 mm). These probability of exceedence plots are an important indicator of model performance, in that they



Appendix Figure A1: Mean monthly rainfall and temperature distribution at studied sites.



Appendix Figure A1 contd. Mean monthly rainfall and temperature distribution at studied sites.

show when the model is able to produce estimates that reproduce the probability of event magnitudes occurring, in this case the mean monthly precipitation and Tmax.

DISCUSSION

The findings show that ClimGen is capable of producing good quality estimates, but also ones that can contain substantial errors. Hence care is required to assess the data quality in advance of it being used for crop, hydrology or ecological modelling. Comparisons between observed and simulated data are a vital part of overall validation, but also provides valuable information on the behaviour (or characteristics) of the data, such as when and where it is able to perform well or not, that is, under-estimating drier mean monthly precipitation at Tiko (Figure 1b) but producing good estimates of Tmax. Knowing the scale of errors (that is, percent difference and variance) is important when interpreting outputs from simulation models that have used to synthetic data as input.

ClimGen appears to perform well across the range of spatial scales and climatic zones found within Cameroon. Further investigation is required to determine why the model performed so poorly for precipitation at Kribi, but this serves as a reminder of the need for assessment prior to data use.

Conclusions

In conclusion, ClimGen showed a good performance, indicating that representative long-term weather data of precipitation and temperatures could in general, be generated from historical weather data. This finding has particular relevance for agricultural modelling applications in Cameroon where there is a limited observed record, making it difficult to evaluate long-term effects of weather on crop yields, hydrological balances and ecosystem responses.

ACKNOWLEDGEMENT

We acknowledge the help and assistance provided by Claudio O. Stöckle, Roger L. Nelson (Biological Systems Engineering Department, Pullman WA, USA) and Gianni Bellocchi (Agrichiana Farming - Montepulciano, Italy) in using ClimGen.

REFERENCES

Bristow KI, Campbell GS (1984). On the relationship between incoming solar Radiation and daily maximum temperature. Agric. and Forest Meteorol. 31: 159-166.

Castellvi F, Stöckle CO (2001). Comparing the performance of WGEN and ClimGen in the generation of temperature and solar radiation. Transactions of Am. Soc. of Agric.l and Biol. Engineers (ASABE), 44: 1683-1687.

Hoogenboom G (2000). Contribution of agro-meteorology to the simulation of crop production and its applications. Agric. and Forest Meteorol. 103: 137-157.
Hoogenboom G, Soltani A (2003). Minimum data requirements for parameter estimation of stochastic weather generators. Climate Res., 25: 109-119.
Kuchar L (2004). Using WGENK to generate synthetic daily weather data for modelling of agricultural processes. Math. and Computers in Simul., 65: 69-75.
Mckague K, Rudra R (2003). ClimGen: A Weather Generation tool for Canadian stations. CSAE/SCGR Meeting, Montreal, Quebec, July 6-9.
Richardson CW, Wright DA (1984) WGEN: A model for generating weather variables. US Department of Agriculture, Agricultural Research Service, ARS-8.
Selker JS, Haith DA (1990). Development and testing of single-parameter precipitation distribution. Water Resour. Res. 26: 2733-2740.
Stöckle CO, Nelson R (2003). Cropping system simulation model user's manual. Washing State University, Pullman, Washington, USA. p. 235.
Stöckle CO, Nelson R, Donatelli M, Castellvì F (2001) ClimGen: a flexible weather generation program. 2nd International Symposium Modelling Cropping Systems. Florence, Italy, 16-18 July.
Tingem M, Rivington M, Azam-Ali SN, Colls JJ (2007). Climate variability and maize production in Cameroon: simulating the effects of extreme dry and wet years. Singap. J. of Trop.l Geography, In Press.
Weibull W (1961). Fatigue testing and analysis of results. Pergamon Press, Oxford, United Kingdom. p. 225.