DESCRIPTION OF *ClimGen*, A WEATHER GENERATION PROGRAM

Introduction

Long-term series of daily weather data are often required for the analysis of weather-impacted systems (e.g., cropping management systems, hydrologic studies, environmental studies, and others). Weather generators are computer programs that use existing weather records to produce long series of synthetic daily climatic data. The statistical properties of the generated data are expected to be similar to those of the actual data. Weather variables required by many applications include precipitation, maximum and minimum temperature, rainfall, solar radiation, wind speed and some measurement of air water vapor (Acock and Acock, 1991). In some cases, records of such variables may be not available, incomplete, insufficient in length, or only summarized in monthly archives. Weather generators are practical tools to bypass those problems (Johnson et al., 1996).

Several computer programs have been developed that are capable of producing stochastically generated weather data from existing daily data. Examples include WGEN (Richardson and Wright, 1984), WXGEN (Sharpley and Williams, 1990), CLIGEN (Arnold and Elliot, 1996), USCLIMATE (Johnson et al., 1996), CLIMAK (Danuso et al., 1997), and ClimGen (Stöckle et al., 1998).

ClimGen, the focus of this article, is a weather generator that uses similar general principles than WGEN, the first and most widely used weather generator in the US, but with significant modifications and additions. ClimGen generates precipitation, daily maximum and minimum temperature, solar radiation, air humidity, and wind speed. It uses a Weibull distribution to generate precipitation amounts instead of the Gamma distribution used by WGEN. The Weibull distribution is easier to parameterize, describes well the distribution of precipitation amounts, and can be simplified for applications to conditions with minimum data. In ClimGen, all generation parameters are calculated for each site of interest while WGEN used fixed coefficients optimized from a large US weather data base. The advantage is that ClimGen can be applied to any world location with enough information to parameterize the program. WGEN uses truncated Fourier series fits to produce daily values for monthly-calculated quantities of mean weather variables. This arbitrarily chosen functional form can lead to relatively poor fit to the data. 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.

Other features of ClimGen that are not available in WGEN include the generation of vapor pressure deficit (VPD) and wind speed. In addition, alternative approaches allow users to estimate VPD and solar radiation from existing temperature records.

Brief description of ClimGen

ClimGen provides utilities for computing all required generation parameters and statistical summaries from existing daily weather records. The methods for deriving these statistical summaries will not be discussed here. A brief description of the approach used to generate daily weather sequences follows.

Generation of precipitation

The generation of precipitation is based on two assumptions. One is that the rain condition on day *i* is related to the rain condition on day *i-1*, and the other is that the amount of rain on rainy days is described by a suitable distribution function. The first assumption describes a type of model called a Markov chain. Defining P(W/W) as the probability of a wet day on day i given a wet day on day i-1, and P(W/D) as the probability of a wet day on day i given a dry day on day i-1, then P(D/W) = 1- P(W/W) is the probability of a dry day given a wet day on day i-1 and P(D/D)= 1- P(W/D) is the probability of a dry day given a dry day on day i-1. These transition probabilities are calculated for each month at each location of interest. Daily values of these probabilities are interpolated using spline functions.

If we know the state of today's weather (wet or dry), we immediately know the probability of a wet day tomorrow (either

P(W/W) or P(W/D)). ClimGen determines whether a particular day is wet or dry by subtracting P(W/W) or P(W/D) from a random number with a range of 0..1. If the result is greater than zero, the generator assumes no rain on that day. If it is less than or equal to zero, rain is assumed to have occurred, and the amount of rain is determined using a distribution function for rain amounts on wet days. Quadratic spline functions are used for daily interpolation of monthly probabilities of a wet day given a previous wet day and a wet day given a previous dry day.

In the case of a wet day, the amount of precipitation is assumed to follow a Weibull distribution:

$$\mathrm{F}(P) = 1 - \exp\left[-\left(\frac{P}{\beta} \right)^{\alpha} \right] \qquad \text{[Eq. 1]}$$

where F(P) is the cumulative probability of a precipitation amount equal or less than P, and α and β are parameters of the distribution function that are calculated on a monthly basis. This distribution is sampled for each precipitation event using the inverse method.

$$\mathrm{P} = \beta\left[-\ln(r)\right]^{1/\alpha} \qquad \text{[Eq. 2]}$$

where r is a random number between 0 and 1.

Generation of temperature and solar radiation

Daily maximum (Tx) and minimum (Tn) temperature and solar radiation (Rs) values are generated in a single operation. The time series of Tx, Tn and Rs are reduced to a time series of residual elements as follows:

$$\chi_{p,i}(j) = \frac{X_{p,i}(j) - X_i^k(j)}{\sigma_i^k(j)} \qquad \text{[Eq. 3]}$$

where $\chi_{p,i}(j)$ is the residual component for variable j (j=1 for Tx , j=2 for Tn and j=3 for Rs) , year p and day i, $X_{p,i}(j)$ is the daily value of the variable, $\chi_i^k(j)$ and $\sigma_i^k(j)$ the daily mean and standard deviation, with k=0 to indicate dry days and k=1 for wet days. The residual series for each variable are expected to be normally distributed with mean zero and variance of one, and described by a first order linear auto-regressive model. The weakly stationary generating multivariate process proposed by Matalas (1967) is used to generate the residual series as follows:

$$\chi_{p,i}(j) = A\chi_{p,i-1}(j) + B\varepsilon_{p,i}(j) \qquad \text{[Eq. 4]}$$

where $\chi_{p,i}(j)$ and $\chi_{p,i-1}(j)$ are 3x1 matrices for day i and i-1 of year p whose elements are the residuals of Tx, Tn and Rs for day i and i-1 of year p respectively, $\varepsilon_{p,i}(j)$ is a 3x1 matrix of independent random components normally distributed with mean zero and variance one, and A and B are 3x3 matrices whose elements are defined such that the new sequences have the desired serial-correlation and cross-correlation coefficients. The A and B matrices are given by:

$$A = M_1 \cdot M_0^{-1} \qquad \text{[Eq. 5]}$$

$$B \cdot B^T = M_0 - A \cdot M_1^T$$

where superscripts -1 and T denote inverse and transpose respectively and M_0 and M_1 are 3x3 matrices whose elements are $m_0(p,q)$ and $m_1(p,q)$ respectively. The elements $m_0(p,q)$ are the lag-0 cross-correlation coefficients between residuals $\chi_{p,i}(p)$ and $\chi_{p,i}(q)$ and $m_1(p,q)$ are the lag-1 cross-correlation coefficients between $\chi_{p,i}(p)$ and $\chi_{p,i-1}(q)$, where *p* and *q* take on different values of j (i.e., j=1 for Tx , j=2 for Tn and j=3 for Rs).

The A and B matrices are used with Eq.4 to generate new sequences of the residuals of Tx, Tn, and Rs that are serially correlated and cross-correlated. Daily generated values of Tx, Tn and Rs are determined by rearranging terms in Eq.3. to solve for $X_{p,i}(j)$ and using the generated residuals to substitute for $\chi_{p,i}(j)$.The daily mean and standard deviation of Tx, Tn and Rs, conditioned on wet or dry status, are obtained from monthly values using spline functions.

Generation of air humidity

As a first step, daytime dew-point temperature (calculated concurrent with the time of minimum relative humidity and maximum temperature) and night dew-point temperature (calculated concurrent with the time of maximum relative humidity and minimum temperature) are determined from actual weather data as follows.

$$\mathrm{e}^0(T_{dd}) = e^0(T_x)\frac{RH_{\min}}{100} \qquad \text{[Eq. 6]}$$

$$\mathrm{e}^0(T_{dn}) = e^0(T_n)\frac{RH_{\max}}{100} \qquad \text{[Eq. 7]}$$

where $e^0(T)$ is the saturation vapor pressure (kPa) determined at the specified temperature T.

$$e^0(T) = 0.6108\exp\left[\frac{17.27T}{T + 237.3}\right] \qquad \text{[Eq. 8]}$$

Dew-point temperatures are obtained by inverting Eq. 8. A linear regression between daytime and night time dew point temperatures is calculated during the parameter optimization phase of ClimGen.

The second step is to calculate from data the daily maximum vapor pressure deficit (VPD_{max}), which is the maximum difference between saturation vapor pressure and actual vapor pressure. This is normally obtained at the time of minimum relative humidity and maximum temperature.

$$\mathrm{VPD}_{\max} = \mathrm{e}^0(\mathrm{T_x})\left(1 - \frac{\mathrm{RH}_{\min}}{100}\right) \quad \text{[Eq. 9]}$$

The daily maximum vapor pressure deficit can also be estimated from temperature.

$$VPD_{\max} = \frac{e^0(T_x) - e^0(T_n)}{1 - a[e^0(T_x) - e^0(T_n)]} \qquad \text{[Eq. 10]}$$

where a (aridity factor) is a parameter optimized from data by combining Eqs. 9 and 10. After optimization of a, VPD_{max} from Eq.9 and 10 are correlated through a linear regression. Typically two years of humidity and temperature data are sufficient to parameterize ClimGen for air humidity data generation.

During generation, ClimGen calculates VPD_{max} using Eq. 10, the optimized a parameter and linear regression slope and intercept, and generated Tx and Tn values. generates daily minimum relative humidity (RHmin) from generated Tx and Tn. Once VPD_{max} is determined, RH_{min} can be calculated from Eq. 9. Then, daytime dew-point temperature can be calculated from RH_{min} and Eq. 6. Night time dew-point temperature can now be determined using the linear regression with daytime dew-point temperature. Finally, Eq. 7 can be used to calculate RH_{max}.

Generation of wind

This variable is generated without any correlation with other variables. Similarly to precipitation, daily wind speed (U) is represented using a Weibull distribution:

$$\mathrm{F}(U) = 1 - \exp\left[-\left(\frac{U}{\beta}\right)^{\alpha}\right] \qquad \text{[Eq. 11]}$$

where F(U) is the cumulative probability of a wind speed amount equal or less than U, and α and β are parameters of the distribution function that are calculated on a monthly basis. This distribution is sampled for each day of weather generation using the inverse method.

$$\mathrm{U} = \beta\left[-\ln(r)\right]^{1/\alpha} \qquad \text{[Eq. 12]}$$

where r is a random number ranging between 0 and 1.

A Brief corroboration of ClimGen generation capabilities

<u>Methodology</u>

Data from thirteen world locations were available for this study (Table 1). All required parameters for weather generation were determined for each location. Using these parameters, daily series of weather data were generated and compared with actual data. The number of years of daily weather record generated corresponded to the length of the available record at each location. Comparisons for precipitation were only performed when the available record length was at least 25 years. Monthly mean and standard deviation, and the frequency distribution (probability of exceedence) were calculated from the daily series of generated and actual weather data. In addition, the occurrence of extreme values for weekly periods was calculated. To evaluate the agreement between actual and generated data, the following indices were used: Root Mean Square Error (RMSE), the General Standard Deviation ($GSD = RMSE/\overline{O}$, where $\overline{O}$ is the mean of the actual data), and the Willmott (1982) index of agreement (d). The lowest limit of RMSE and GSD is 0, indicating perfect agreement between generated and actual values. The index of agreement ranges between 0 and 1, where a value of 1 indicates perfect agreement. For the interpretation of the performance indices, values of GSD≤0.10 and d≥0.95 were considered indicators of good performance. Values of GSD>0.10 but ≤0.20 and values of d<0.95 but ≥0.90 were considered acceptable. Other values indicated poor performance.

Results and discussion

Results of comparisons between generated and actual daily weather, analyzed for monthly periods, are given in Table 2. Due to space limitations, the results presented are for selected locations, representative of overall performance. In most cases, an excellent agreement was obtained. Exceptions were found for the fraction of wet days and minimum temperature at some locations. However, even in those cases, the agreement was acceptable.

Figure 1 compares, for selected locations, the probability of exceedence of actual and generated daily values for all the weather elements generated. These comparisons are given for illustration purposes, but again they are representative of overall performance. Similar results were obtained for all locations and weather variables (data not shown), showing a close agreement between the actual and generated frequency distributions of daily values.

The comparisons between actual and generated extreme values for weekly periods are shown in Table 3. Large departures were observed in some cases, but all the generated values seemed plausible. Overall, the results of this evaluation show that the generation methods in ClimGen are sound.

Conclusions

Overall, results indicated a good performance of the ClimGen weather generator. In most cases, an excellent agreement between actual and generated weather was found for monthly period comparisons. Frequency distributions of actual and generated daily data were also in good agreement. Tests of extreme values for weekly periods showed that most generated values were plausible, with only a few significant departures between generated and actual values. The agreement between system responses, decision making, and/or interpretations based on actual and generated weather remains to be evaluated.

REFERENCES

Acock, B. and M.C. Acock, 1991. Potential for using long-term field research data to develop and validate crop simulators. Agron. J. 83:56-61.

Arnold, C.D. and W.J. Elliot, 1996. CLIGEN Weather Generator Predictions of Seasonal Wet and Dry Spells in Uganda. Trans. of ASAE 39(3):969-972.

Danuso, F. et al., 1997. CLIMAK reference manual. DPVTA, University of Udine, Italy, 36 p.

Johnson, G.L., C.L. Hanson, S.P. Hardegree and E.B. Ballard, 1996. Stochastic Weather Simulation: Overview and analysis of two commonly used models. Journal of Applied Meteorology 35:1878-1896.

Matalas, N.C., 1967. Mathematical assessment of synthetic hydrology. Water Resources Research 3(4):937-945.

Richardson, C.W., 1982. Dependence structure of daily temperature and solar radiation. Trans. of ASAE 25:735-739.

Richardson, C.W. and D.A. Wright, 1984. WGEN: A model for generating daily weather variables. USDA-ARS, 235 p.

Sharpley, A.N., and J.R. Williams, 1990. EPIC-Erosion/Productivity Impact Calculator: 1. Model Documentation. US Department of Agriculture Technical Bulletin No. 1768, 235 p.

Stockle, CO, P. Steduto, and R.G. Allen. 1998. Estimating daily and daytime mean VPD from daily maximum VPD. 5th Congress of the European Society of Agronomy, Nitra, The Slovak Republic.

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

Willmott, C.J. 1982. Some comments on the evaluation of model performance. Bulletin of American Meteorological Society 63:1309-1313.

Table 1. Locations and number of years of available weather record.

Location			**Length of data record (years)**				
Name	**Lat**	**Long**	**Precip**	**Temp**	**Solar Rad**	**Rel Hum**	**Wind speed**
Akron CO, USA	40.09º N	103.20º W	33	33	15	-	-
Dalby, Australia	27.11º S	151.00º E	29	29	-	-	-
Haarveg, The Netherlands	51.97º N	5.67º E	36	36	36	36*	36
Los Baños, Philippines	14.22º N	122.00º E	14	14	14	14*	14
Katherine, Australia	13.29º S	132.40º E	32	32	-	-	-
Kimberly ID, USA	42.40º N	114.20º W	11	11	-	-	-
Lleida, Spain	41.70º N		6	6	6	6	6
Manhattan KS, USA	39.20º N	96.80º W	32	32	15	-	-
Pisa, Italy	43.40º N	11.00º E	27	27	5	5	5
Prosser WA, USA	46.25º N	119.75º W	10	10	10	10	10
Rodeplaat, S. Africa	25.58º S	28.35º E	13	13	13	-	-
Tel Hadya, Syria	36.01º N	36.93º E	12	12	12	12	12
Versailles, France	48.9º N	2.00º E	34	34	15	-	-

*The daily vapor pressure was estimated based on early morning vapor pressure measurements.

Table 2. Statistics and indicators of agreement between generated and actual daily records summarized for monthly periods.

Location	Statistics		P	f_{wet}	T_{max}	T_{min}	S_t	VPD	U
Akron	Actual	Mean	33.434	0.187	17.204	1.498	16.989	-	-
		Stdev	27.490	0.076	10.670	9.459	6.480	-	-
	Generated	Mean	33.230	0.189	17.253	1.576	16.592	-	-
		Stdev	25.916	0.074	10.573	9.327	6.328	-	-
		RMSE	4.382	0.014	0.376	0.310	0.457	-	-
		GSD	0.020	0.075	0.022	0.206	0.027	-	-
		d	0.999	0.989	0.999	0.999	0.998	-	-
Haarweeg	Actual	Mean	63.185	0.468	13.248	5.290	9.168	0.671	2.412
		Stdev	9.840	0.047	6.890	5.109	6.101	0.396	0.293
	Generated	Mean	65.855	0.480	13.256	5.128	9.279	0.704	2.420
		Stdev	10.599	0.068	6.747	5.178	6.031	0.404	0.285
		RMSE	0.194	0.026	0.273	0.334	0.198	0.049	0.033
		GSD	0.049	0.056	0.021	0.063	0.022	0.074	0.014
		d	0.998	0.939	0.999	0.999	0.999	0.995	0.996
Katherine	Actual	Mean	80.386	0.220	33.912	19.449	-	-	-
		Stdev	98.039	0.226	2.648	5.060	-	-	-
	Generated	Mean	80.733	0.221	33.988	19.425	-	-	-
		Stdev	98.075	0.223	2.509	5.029	-	-	-
		RMSE	4.833	0.016	0.196	0.161	-	-	-
		GSD	0.007	0.075	0.006	0.008	-	-	-
		d	0.999	0.998	0.998	0.999	-	-	-
Pisa	Actual	Mean	73.578	0.263	19.792	9.899	12.466	1.149	1.691
		Stdev	30.518	0.083	6.778	5.641	6.422	0.546	0.207
	Generated	Mean	74.110	0.265	19.804	9.938	12.520	1.181	1.709
		Stdev	32.573	0.093	6.904	5.766	6.463	0.739	0.267
		RMSE	6.479	0.017	0.253	0.252	0.484	0.207	0.082
		GSD	0.015	0.065	0.013	0.025	0.039	0.180	0.048
		d	0.999	0.989	0.999	0.999	0.998	0.969	0.964
Tel Hadya	Actual	Mean	-	-	24.635	10.359	17.309	2.540	2.817
		Stdev	-	-	10.161	7.625	7.566	1.717	1.158
	Generated	Mean	-	-	24.499	10.245	16.982	2.554	2.808
		Stdev	-	-	9.996	7.300	7.363	1.649	1.144
		RMSE	-	-	0.419	0.590	0.489	0.154	0.059
		GSD	-	-	0.017	0.057	0.028	0.061	0.021
		d	-	-	0.999	0.998	0.999	0.997	0.999

Table 3. Comparison of extreme values on a weekly basis.

Location	Maximum rain in a week (mm)			Mean T_{max} of hottest week (ºC)			Mean T_{min} of coldest week (ºC)			Mean S_t of most radiant week (MJ m^{-2} d^{-1})		
	act	**gen**	**E(%)**	**act**	**gen**	**E(%)**	**act**	**gen**	**E(%)**	**act**	**gen**	**E(%)**
Akron	112.776	128.460	13.907	37.383	38.730	3.604	-26.903	-20.523	-23.715	30.853	30.254	-1.941
Haarweg	111.100	127.190	14.482	31.357	29.416	-6.191	-16.714	-9.251	-44.650	26.717	26.570	-0.551
Katherine	235.100	270.600	15.100	41.414	43.304	4.564	3.871	4.913	26.900	-	-	-
Pisa	207.100	183.600	-11.347	36.314	37.411	3.021	-4.000	-4.001	0.036	26.734	26.627	-0.401
Tel Hadya	73.900	63.910	-13.518	42.671	42.784	0.264	-7.671	-2.960	-61.415	30.543	30.499	-0.145

Location	Mean S_t of less radiant week (MJ m^{-2} d^{-1})			Mean VPD_{max} of driest week (kPa)			Mean VPD_{max} of most humid week (kPa)			Mean wind speed of most windy week (m s^{-1})		
	act	**gen**	**E(%)**	**act**	**gen**	**E(%)**	**act**	**gen**	**E(%)**	**act**	**gen**	**E(%)**
Akron	4.439	4.074	-8.213	-	-	-	-	-	-	-	-	-
Haarweg	0.592	0.229	-61.409	3.386	2.704	-20.152	0.051	0.015	-71.463	6.900	5.490	-20.435
Katherine	-	-	-	-	-	-	-	-	-	-	-	-
Pisa	1.835	2.109	14.920	3.092	3.133	1.324	0.085	0.094	11.139	4.136	3.550	-14.174
Tel Hadya	3.614	4.093	13.241	7.311	6.001	-17.918	0.217	0.244	12.436	8.009	7.226	-9.783

$$* \; E(\%) = \frac{\text{gen.} - \text{act.}}{\text{act.}} \cdot 100$$

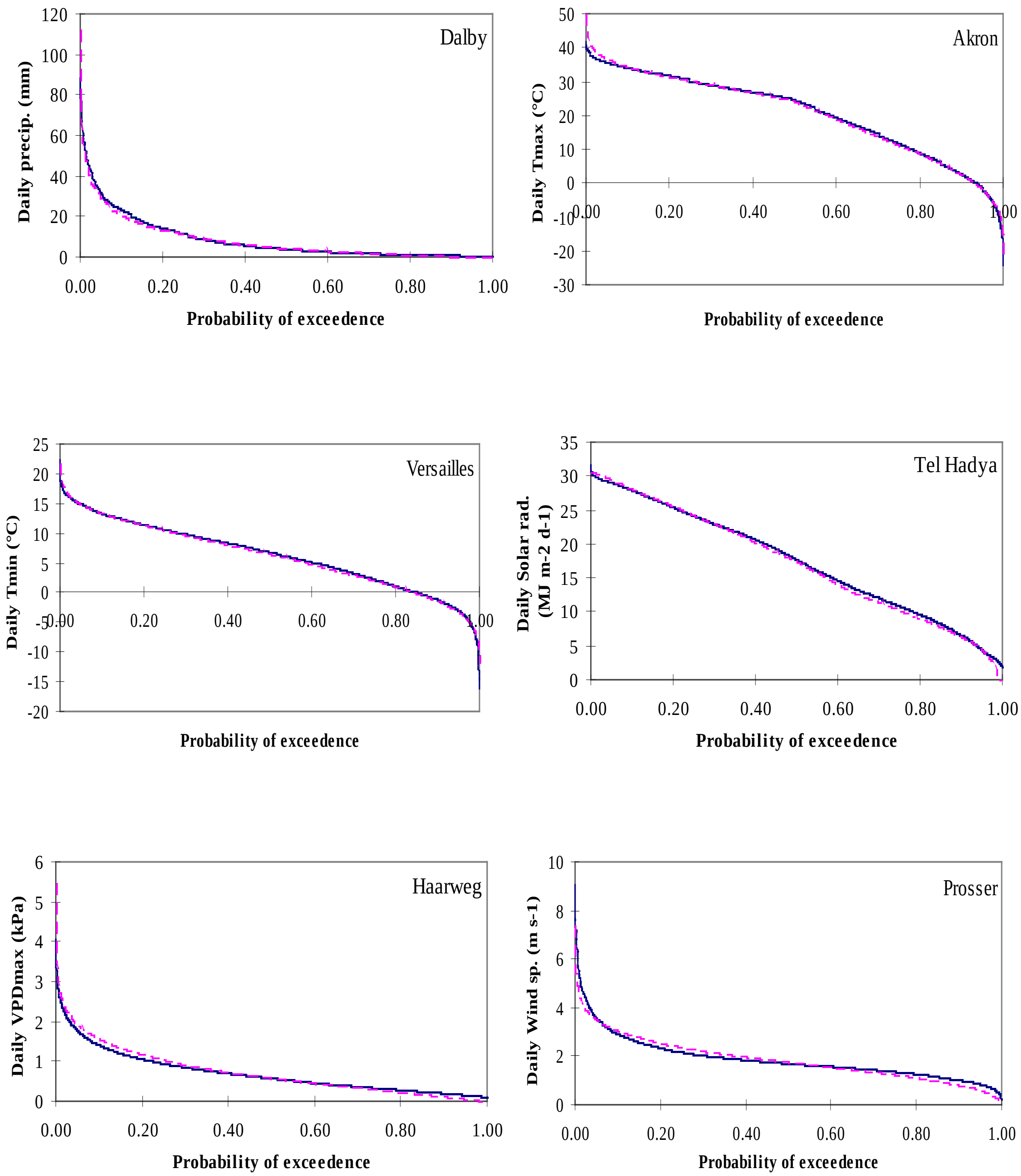

Figure 1. Probability of exceedence of actual (solid line) and generated (dotted line) daily weather variables at selected sites.