

A METHODOLOGY FOR THE SYNTHESIS OF HOURLY WEATHER DATA

K. M. KNIGHT, S. A. KLEIN, and J. A. DUFFIE

Solar Energy Laboratory, University of Wisconsin-Madison, Madison, WI 53706, U.S.A.

Abstract—The ability to generate weather data from limited inputs and location independent correlations would allow simulations of environmentally driven systems to be run at locations for which detailed weather records do not exist. A further improvement would be to generate just one year of data which would yield the same simulation results as those of long-term data, thereby minimizing the computation needed to assess long-term system performance. In this paper, a methodology to calculate such a year of data is described. It is a combination and modification of several previously published generation methods. Statistics obtained from long-term (22 years), Typical Meteorological Year, and synthetically generated weather data are compared at three locations (Albuquerque, NM, Madison, WI, and New York, NY).

1. INTRODUCTION

The performance of environmentally driven systems is dependent upon the solar radiation, ambient temperature, humidity, and windspeed. These variables are neither completely random nor deterministic; they can best be described as irregular functions of time, both on small (e.g., hourly or daily) and large (e.g., seasonally or yearly) time scales. It is the irregular behavior of the weather which complicates the analyses of solar energy systems and makes experimental determinations of system performance time consuming, expensive, and inconclusive with regard to the manner in which they would have performed under other climatic conditions.

Simulations have in common with experiments the problem of accounting for the irregular behavior of weather. A simulation will provide results only for the period over which weather data are provided. Sufficient data (e.g., many years) must be provided in order for the simulation to calculate the long-term average performance of the system. There are two problems associated with long-term simulations. First, the hour-by-hour simulation of a system over many years involves significant computational effort and is therefore slow and/or expensive. Second, hourly records of meteorological variables for extended periods of time do not exist for many locations.

Since simulations must often be performed for locations without detailed recorded data, two methods of providing data are commonly used; extrapolation and synthetic generation. Extrapolation involves using data at one or more neighboring or similar climate locations to infer the weather data at a location for which recorded data are not available. Extrapolation of existing data may result in significant errors[1]; there may also be errors associated with synthetic generation.

Numerous authors have developed models for the generation of radiation series (both daily and/or hourly), for example[2-12]. Temperature and windspeed have also been the object of several studies, for

example[11-15]. Degelman[16,17] developed a weather generator which requires only limited monthly-average values as input and outputs hourly radiation, temperature, humidity, and windspeed values. Validation studies[18] of the Degelman program have revealed areas of weakness, including some location dependence.

In this paper, techniques for the generation of radiation and ambient temperature data are presented, along with suggestions for humidity and windspeed. Required input data consists only of monthly-average values. Representative statistics calculated from the generated, long-term, and "Typical Meteorological Year" (TMY) data[19] are presented for three locations: Albuquerque, NM, Madison, WI, and New York, NY. Albuquerque is classified as a mountain desert climate, Madison as a temperate continental climate, and New York as a temperate oceanic climate according to Köppen's reformed classification[20]. The long-term data are taken from the SOLMET database[21], and consist of 22 to 23 years of hourly radiation, temperature, humidity and windspeed data. The definition of "long-term" is difficult; indeed the weather statistics when viewed on a large enough scale are not stationary. For the purpose of this analysis, however, long-term is defined by the length of available weather records, namely 22 to 23 years.

Cross-correlations are not directly reproduced by the modelling techniques presented here. This is perhaps the most serious shortcoming; further research in this area should include a study of the cross-correlations between the different weather variables, first quantifying the cross-correlations, then evaluating their effect on system performance, and finally, if necessary, determining a method for including them in the model, such as by using multivariate time series.

2. STATISTICS OF WEATHER DATA

Viewed on a short-term basis, meteorological variables seem to be highly random and unpredictable.

However, statistical analyses of these variables indicate that they are not as unpredictable as they appear. Often the distributions and diurnal variations can be expressed in location independent forms. The following is an accounting of the correlations used to produce the results shown in this paper; if other more accurate correlations are available, they can be substituted.

The clearness index is defined as the ratio of the global solar radiation on a horizontal surface to the global extraterrestrial solar radiation on a horizontal surface [22], and can be defined either on a daily (K_t) or an hourly (k_t) basis. The distribution of the daily clearness index was shown by Liu and Jordan [23] to be primarily dependent on the monthly-average daily clearness index, \bar{K}_t . Hollands and Huget [25] developed an implicit expression for the Liu and Jordan distributions. Bendt *et al.* [24] analyzed 20 years of data from 90 U.S. locations and developed the following expression for the daily K_t distribution based on their data which agreed well with the curves of [23] except at high fraction values:

$$f = \frac{\exp(\gamma K_{t,\min}) - \exp(\gamma K_t)}{\exp(\gamma K_{t,\min}) - \exp(\gamma K_{t,\max})} \quad (1)$$

where f is the cumulative fraction of occurrence and γ is found implicitly from the following equation:

$$\bar{K}_t = \frac{\left(K_{t,\min} - \frac{1}{\gamma}\right) \exp(\gamma K_{t,\min}) - \left(K_{t,\max} - \frac{1}{\gamma}\right) \exp(\gamma K_{t,\max})}{\exp(\gamma K_{t,\min}) - \exp(\gamma K_{t,\max})} \quad (2)$$

Alternatively, Herzog [26] gives an explicit relation for γ from a curve fit:

$$\gamma = -1.498 + \frac{1.184\xi - 27.182 \exp(-1.5\xi)}{K_{t,\max} - K_{t,\min}} \quad (3)$$

where

$$\xi = \frac{K_{t,\max} - K_{t,\min}}{K_{t,\max} - \bar{K}_t}$$

$K_{t,\min}$ was recommended by Bendt *et al.* [24] as a constant value of $K_{t,\min} = 0.05$; no expression was provided for $K_{t,\max}$. Hollands and Huget [25] recommend that $K_{t,\max}$ can be estimated from:

$$K_{t,\max} = 0.6313 + 0.267\bar{K}_t - 11.9(\bar{K}_t - 0.75)^8 \quad (4)$$

Recently, several papers have questioned the universality of these K_t distributions, particularly their validity in tropical climates. Saunier *et al.* [27] examined K_t distributions from 5 locations (4 tropical summer rain climates and 1 subtropical winter rain climate [20]) and significant discrepancies with the Bendt

et al. [24] correlation are apparent. An expression for the K_t distribution in tropical climates is presented in [27]. Gordon and Reddy [9] have proposed a new expression for the K_t distribution for all climates, including the variance in the parameters necessary to define the distribution. However, the data they analyzed consists of 7 tropical summer rain climates, 2 subtropical summer rain climates, 1 subtropical winter rain climate, 1 desert climate, and 2 steppe climates [20]. No temperate climate data were included, and as such, the universality is as yet unverified. Stuart and Hollands [28] propose modelling the distribution of the beam transmittance instead of the clearness index, and provide an expression for the distribution, the parameters of which are the air mass and the mean value of the beam transmittance. Again, only limited climate types were examined: 3 temperate continental climates and 1 temperate oceanic climate [20]. The results presented in this paper were obtained by using the Bendt *et al.* correlation (eqn 1) with eqn (3) for estimating γ . This correlation was chosen in part for its simplicity; if, however, the user is dealing with tropical climates, another expression, for example Saunier *et al.* [27], should be substituted.

The long-term monthly-average diurnal variation of hourly total radiation (when divided by the average daily total radiation) has been shown to be primarily a function of the hour angle, ω , and the sunset hour angle, ω_s [23,29,30].

$$r_t = \frac{\bar{I}}{\bar{H}} = (a + b \cos \omega) \frac{I_o}{H_o} \quad (5)$$

$$a = 0.409 + 0.5016 \sin(\omega_s - 60)$$

$$b = 0.6609 + 0.4767 \sin(\omega_s - 60)$$

Equation (5) can be used to calculate the monthly-average radiation for an hour, e.g., 9–10 a.m., when the monthly-average daily radiation is known. Equation (5) does not give any information about the hourly radiation values for each day of the month (other than the mean). It has been suggested [22], however, that eqn (5) be used to obtain the expected average value of the ratio I/H for a given location, day, and hour. Multiplying by the daily total radiation, H , then gives the expected average value of the hourly radiation, I_{LT} , for that location, day, hour and daily radiation value. This radiation value is the "long-term average" value, and correspondingly, a "long-term" average k_t value, k_{tm} , can be approximated as:

$$k_{tm} = \frac{I_{LT}}{I_o} = \frac{\left(\frac{I}{H}\right) H}{I_o} \frac{H_o}{H_o} = \frac{H}{H_o} \left(\frac{I}{H}\right) \frac{H_o}{I_o} \approx K_t r_t \frac{H_o}{I_o} \quad (6)$$

Alternatively, Graham *et al.* [2] have proposed the following relation for estimating the long-term average value of the radiation at an hour:

$$k_{im} = K_i - 1.167K_i^2(1 - K_i) + 0.979(1 - K_i)\exp\left[\frac{-1.141(1 - K_i)}{K_i \cos \theta_z}\right] \quad (7)$$

where k_{im} is defined as the long-term average value of k_i for that hour of the day, month and daily clearness index value, K_i . The average of the deviations of k_i from k_{im} obtained from each expression (eqns (6) and (7)) is small; however, the average of the deviations obtained from the long-term data (22 years) for Albuquerque, Madison, and New York using eqn (6) is less than those reported by Graham [4] for Toronto, Swift Current, and Vancouver using eqn (7), as seen in Fig. 1. In addition, eqn (7) does not always return the correct K_i value. For example, in Madison using a value of $K_i = 0.50$ in eqn (7), the value of K_i obtained from the hourly k_{im} values is 0.46 on January 15; 0.50 on March 15; and 0.53 on June 15. The K_i values obtained from eqn (6), however, agree with the input K_i values to ± 0.005 .

Autocorrelation refers to the dependence of the current value of a meteorological variable on earlier values. It is generally most useful to eliminate deterministic trends in the data set before attempting to calculate autocorrelation values, resulting in a "de-trended" series. Estimates of the autocorrelation are often computed from the following equation [31,32], although this is not the only expression used in the literature to estimate autocorrelation:

$$\hat{r}(p) = \frac{\text{Cov}(p)}{\text{Cov}(0)} = \frac{\frac{1}{N} \sum_{i=1}^{N-p} (y_i - \bar{y})(y_{i+1} - \bar{y})}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} \quad (8)$$

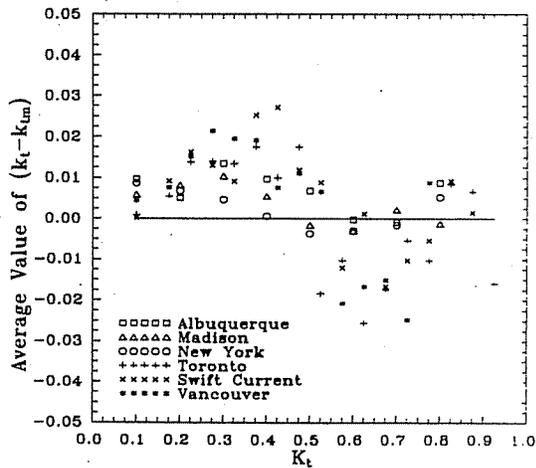


Fig. 1. Average values of the deviations of k_i from k_{im} grouped by daily clearness index. k_{im} is calculated from eqn (6) for Albuquerque, Madison, and New York; from eqn (7) for Toronto, Swift Current, and Vancouver. Data for Toronto, Swift Current, and Vancouver is from [4].

where p indicates the lag. The autocorrelation of daily total radiation has been investigated by various authors, for example [3-9,18,33-35]. The referral to "daily total radiation" implies all of the related de-trended radiation variables used by the different authors in the autocorrelation estimations, for example, K_i and Z , where $Z = [(K_i/\bar{K}_i) - 1]/\sigma_{K_i/\bar{K}_i}$. The lag one autocorrelation of the annual series of daily total radiation (including the related de-trended radiation variables) is generally in the range of 0.15 to 0.35 [3,4,6-8,34,35] and shows no systematic dependence on location or climate type. Part of the variation in the estimates is due to the use of different de-trended radiation variables. An average value of 0.29 is reported by Graham [4] for the K_i series; Klein [35] suggests using an average value of 0.3. The reported estimates of the lag one autocorrelation of the monthly series of daily total radiation cover a wide range: -0.16 to 0.55 [4,5,9,35]. Again, part of this variation is due to the different de-trended radiation variables, but some of the variation is also due to different autocorrelation estimation methods. In addition, the associated standard error of the monthly values tends to be larger since the data set is smaller. It is therefore difficult to draw any conclusions about the lag one autocorrelation values of the monthly series.

Many of these authors [4,6-8,35] have concluded that the autocorrelation in daily solar radiation can be described by a first order autoregressive model. An autoregressive (AR) model [31] is simply a model in which the present value of some variable, Y_t , is regressed on previous values:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_N Y_{t-N} + \epsilon_t \quad (9)$$

The order of the model indicates the number of parameters in the model. For example, for a first order autoregressive model (AR1), N in eqn (9) equals 1; the single parameter of an AR1 model, ϕ_1 , is the lag one autocorrelation coefficient.

The diurnal variation of the hourly monthly-average temperatures, T_h , has a location and month independent shape when standardized by subtracting the mean and dividing by the amplitude. Erbs *et al.* [36] demonstrated that the average normalized diurnal temperature variation can be represented by

$$\begin{aligned} (\bar{T}_h - \bar{T})/A = & 0.4632 \cos(t^* - 3.805) \\ & + 0.0984 \cos(2t^* - 0.360) \\ & + 0.0168 \cos(3t^* - 0.822) \\ & + 0.0138 \cos(4t^* - 3.513) \quad (10) \\ t^* = & 2\pi(t - 1)/24 \end{aligned}$$

where \bar{T} is the monthly-average daily temperature and t is the hour of the day defined such that $t = 1$ at 1 a.m. and $t = 24$ at midnight. The amplitude, A , in °C, while varying considerably with month and location, was shown to be related to \bar{K}_i [36]:

$$A = 25.8\bar{K}_i - 5.21. \quad (11)$$

The distribution of the daily average temperature about the monthly-average daily temperature exhibits wide variations in the mean and standard deviation for different months and locations. Although the distribution is not Gaussian, Erbs *et al.*[36] showed that when expressed in the terms of a normalized variable, h , the cumulative distribution can be represented by the following equation:

$$F_{\text{temp}} = \frac{1}{1 + \exp(-3.396h)} \quad (12)$$

where

$$h = (T - \bar{T})/(\sigma_m \sqrt{N_m/24}). \quad (13)$$

In eqn (13), N_m is the number of hours in the month and σ_m is the standard deviation of \bar{T} about its long-term average value; σ_m can be estimated as a function of \bar{T} and σ_{yr} , the standard deviation of the 12 \bar{T} 's about the yearly-average temperature[36]:

$$\sigma_m = 1.45 - 0.0290\bar{T} + 0.0664\sigma_{yr}. \quad (14)$$

Erbs *et al.*[36] suggest that the expression for the distribution (eqn (12)) is equally valid for the distribution of the hourly temperatures about their hourly monthly-average values if \bar{T}_h is substituted for \bar{T} in eqn (13). The relation for σ_m can be used in the hourly distribution since σ_m is approximately equal to the standard deviation of the hourly temperatures[36].

3. "TYPICAL METEOROLOGICAL YEAR" DATA

"Typical Meteorological Years" of hourly weather data (commonly known as TMY data) have been developed for many U.S. locations[19]. They consist of 12 typical months, where each month was selected from 23 years of recorded data as being representative of the long term. Specifically, the three criteria used to define 'typical' as stated in [19] are (i) distributions should be "close to the long-term distributions"; (ii) sequences of daily variables should be "in some sense like the sequences often registered"; and (iii) cross-correlations should be "like the correlation observed in the meteorological data." Since it is virtually impossible to find one month (e.g., one January) out of the 23 (e.g., 23 Januarys) in which the above are all satisfied exactly for all variables, the actual selection process involved minimizing the difference from long-term distributions, means, and daily persistence (autocorrelation) of a set of 13 weather indices, of which some indices were judged to be more important. As such, the TMY data set is a compromise between using a reduced set of weather data and accurately portraying the statistics represented in long-term data. Although TMY data are commonly used in simulations, no published study of the accuracy of simulation results obtained with these data was found.

4. THE TWO TYPES OF WEATHER DATA GENERATION SCHEMES

There are two different approaches to weather data generation. The first, Type I, is to combine the existing information concerning distributions, autocorrelations, and possibly cross-correlations of the meteorological variables into a format which is driven by a random number generator and produces, over the long-term, weather statistics which are indistinguishable from those of recorded data. Various types of stochastic models usually form the basis of these models. This is the most common approach[2-15]. The disadvantage of these models is that the generated data are subject to the same variability as real data and consequently simulations have to be run for long periods in order to reproduce statistics apparent in long-term weather records. From this synthetic data a year similar in concept to a TMY year could be constructed, but as with actual data, it is difficult to choose one year of data which closely represents the statistics observed in the long-term data. As an alternative, Gordon and Reddy[9] recommend generating years of data until one is generated in which the statistics are similar to the long-term, and storing the random number seed. While this approach may be plausible for generating data for just one weather variable, when generating four different variables, the number of generated years required before one is found in which all of the statistics match the long-term would increase considerably.

The second approach to weather generation, Type II, is to directly generate just one 'typical' year in which all of the statistics match those of the long-term, including short-term randomness. This is the approach taken by Degelman[16,17]. This paper describes a technique which uses parts of a Type I generator to provide realistic short-term variation with the ultimate goal of producing a Type II year.

5. GENERATION OF RADIATION DATA

Hourly radiation values are generated by a two-step process. First, the radiation for each day of the month is obtained from the daily clearness index distribution and ordered according to a predetermined sequence which maintains the approximate autocorrelation of daily radiation. Second, for each day, a series of hourly clearness index values is generated from a first-order autoregressive model.

Daily K_t values are obtained from the daily K_t cumulative distribution function. The cumulative distribution function relates the cumulative fraction of occurrence, F , to K_t . The cumulative fraction of occurrence specifies the fraction of time the K_t variable will be less than a specified value of K_t . As an example, for a 31-day month, there is a value of K_t (call it κ) corresponding to an F value of $\frac{1}{31}$, meaning that over the long-term, only $\frac{1}{31}$ of the time will K_t take on a value less than κ , that is, only 1 out of 31 days will have a K_t value less than κ . The standard procedure for selecting a value is to take the value of K_t at the average

of this F value ($\frac{1}{31}$) and the previous one (0), for example, $F = \frac{1}{62}$. The K_t value corresponding to this F value can then be found from the cumulative distribution function. Similarly, one and only one K_t value will occur between the K_t value associated with $F = \frac{1}{31}$ and $F = \frac{2}{62}$; this is the K_t value corresponding to an average F value of $\frac{1}{62}$. Continuing on in this manner, 31 K_t values can be generated. This method is easily applied to a 30- or 28-day month.

While this technique provides the daily clearness index values for a month, it does not specify the order in which the K_t values occur. They should not occur in either ascending or descending order, yet neither should they be ordered randomly. The lag one autocorrelation is a measure of the appropriate order, and as previously noted, is generally in the range of 0.15 to 0.35, an indication of weak positive correlation. The approach used by Degelman [16,17] capitalizes on the similarity (i.e., all weak positive) in autocorrelation at different locations, and fixes the order in which the daily K_t values occur, yielding a series of clearness indices for which the lag one autocorrelation is approximately equal to the appropriate average value. Specifically, the integers 1 to 31 are assigned to the 31 K_t values obtained from the daily "source" distribution (the daily K_t distribution), with 1 corresponding to the smallest K_t value and 31 to the largest. The integers 1 to 31 are then placed in an order such that when the K_t values corresponding to the numbers are placed in that order, the approximate lag one autocorrelation is reproduced. Figure 2 illustrates this process for a 5-day month. Throughout the rest of this paper, the term "sequence" is used exclusively to refer to one of the sequences listed in Table 1.

Modelling the annual series of daily total radiation is the technique used by various authors [3,5,7,35]. Because the autocorrelation estimates obtained from the annual series are more accurate, in addition to being

Table 1. Sequences for ordering daily weather variables

K_t	$\bar{K}_t \leq 0.45$	24, 28, 11, 19, 18, 3, 2, 4, 9, 20, 14, 23, 8, 16, 21, 26, 15, 10, 22, 17, 5, 1, 6, 29, 12, 7, 31, 30, 27, 13, 25
K_t	$0.45 < \bar{K}_t < 0.55$	34, 27, 11, 19, 18, 3, 2, 4, 9, 20, 14, 23, 8, 16, 21, 7, 22, 10, 28, 6, 5, 1, 26, 29, 12, 17, 31, 30, 15, 13, 25
K_t	$\bar{K}_t \geq 0.55$	24, 27, 11, 4, 18, 3, 2, 19, 9, 25, 14, 23, 8, 16, 21, 26, 22, 10, 15, 17, 5, 1, 6, 29, 12, 7, 31, 20, 28, 13, 30
T		20, 29, 13, 26, 31, 30, 21, 12, 14, 11, 2, 1, 3, 15, 25, 9, 5, 7, 6, 4, 19, 8, 10, 23, 22, 27, 16, 18, 28, 17, 24
T_{max}		24, 29, 14, 21, 31, 30, 23, 5, 12, 11, 2, 1, 7, 16, 25, 10, 8, 3, 4, 9, 18, 6, 13, 26, 20, 22, 15, 17, 27, 19, 28

somewhat universal, the estimate of the lag one autocorrelation of the annual series was chosen to be replicated by the sequence. In particular, Graham's recommended estimate of 0.29 was used as a target for the data presented in this paper. If there is a significant monthly variation in the autocorrelation structure, this approach represents a simplification, and if the user's system is expected to be highly sensitive to the daily radiation autocorrelation, the user is cautioned that this method may not be appropriate; the use of TMY data under these conditions may also be inappropriate.

The autocorrelation values generated from Degelman's sequence were slightly higher than those reported above, ranging from 0.28 to 0.44, while also varying as a function of \bar{K}_t . The variation is due to the different shapes of the K_t distribution curves. Using trial and error, three sequences were developed for three ranges of \bar{K}_t (see Table 1), such that the lag one autocorrelation for all values of \bar{K}_t is within the range of 0.15 to 0.35. The same sequences are always used for ordering the daily K_t values, however, the starting position within the sequence is randomly determined. While the resulting series of K_t values is not the order in which real days always occur, it is a possible order which preserves the proper autocorrelation between successive days.

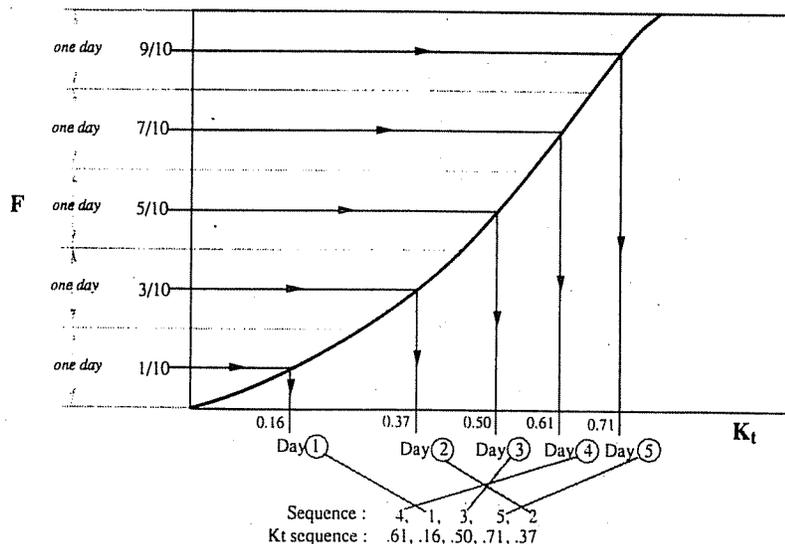


Fig. 2. Obtaining the daily K_t values from the K_t distribution according to a fixed sequence for a 5-day month.

There is nothing special about the particular sequences used; any others which yield the correct autocorrelation could be substituted. Alternatively, if the user specifically knows the lag one daily autocorrelation for the location of interest, by trial-and-error, a different sequence can easily be obtained which would yield that value. If a different K_t distribution (i.e., not the Liu and Jordan distribution or any of the curve fits to it) is used, such as for tropical climates, these sequences should also be modified since the autocorrelation value obtained from the sequences is sensitive to the distribution shape.

Once the daily value of K_t is known, the long-term 'mean' value of clearness index for each hour, k_{tm} , is estimated from eqn (6). Generated days in which the hourly to daily radiation ratio is equal to the long-term value are smooth and symmetric. Such behavior is not typical and generated radiation values with this behavior can lead to significant errors in performance estimates for some systems [37]. The goal is to generate an average month, but the days within this month should be realistic in the sense that the hourly radiation values should display the same variability observed in actual days. This suggests that a Type I generation scheme should be used on an hourly basis. Graham [4] developed a first order autoregressive model for reproducing hourly k_t values; a slight modification of this model was used for the work reported here. Specifically, the long-term average k_t value for an hour, k_{tm} , is estimated from a correlation (such as eqns (6) or (7)); the deviations of the k_t values from the long-term average value are modeled stochastically. The model of Graham [4] was selected for its generality, simplicity, and accuracy, although the recent work of Gordon and Reddy [10] has questioned the universality of this model, particularly for non-temperate climates.

To employ a stochastic model, the series to be modeled must be stationary (i.e., the probability structure must be the same for all time) and Gaussian. The distribution of k_t about k_{tm} satisfies neither of these requirements. The shape of the distribution is dependent on both the hour of the day and the value of the daily clearness index. For a particular value of K_t , the variance of the k_t distribution will be larger for an hour far from noon. Physical constraints require the variance to be different for different daily K_t values; for a day with a high value of K_t , every hour must be clear (and closer to the long-term average value), whereas for a mid-range value of K_t , some hours could be cloudy and some could be clear. Since k_t is bounded by 0 and 1, the distributions will be skewed to the left for low k_{tm} values and skewed to the right for high k_{tm} values. Modeling the k_t 's therefore requires modeling a variable whose probability structure changes each hour and each day.

To eliminate this problem, Graham [4] transforms the k_t values through their cumulative distribution function to a normally distributed variable, χ , with mean 0 and variance 1. This transformed variable can then be represented by a first order autoregressive model (eqn (9)). The parameter ϕ_1 was found by Gra-

ham to be a weak function of K_t , but not statistically different from the mean value of 0.54. This value also represents the lag one autocorrelation of the deviations of k_t from k_{tm} . Since as many as 50 values are needed in a series for a reliable estimate of the lag one autocorrelation [31], and since the k_t series is not continuous (errors would be introduced by catenating each day's hourly series), a pooled estimate of the autocorrelation was used by Graham [4] to estimate the lag one autocorrelation:

$$\hat{r}_1 = \frac{\sum_{i=2}^{\text{days } N-1} (y_i - \bar{y})(y_{i+1} - \bar{y})}{\sum_{i=2}^{\text{days } N-1} (y_i - \bar{y})^2} \quad (15)$$

where N indicates the number of hourly k_t values in a day. Good agreement with Graham's results was found from an analysis of 22 years of data from Albuquerque, Madison, and New York [18].

To generate the k_t values, each hour a χ value is obtained by randomly selecting a value for ϵ_t from a Gaussian distribution and applying eqn (9) with N equal to 1. χ is transformed to the non-Gaussian k_t by equating the cumulative distribution functions. The expression for the normal cumulative distribution with a mean of 0 and variance of 1 is

$$F_{\text{normal}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2} t^2\right) dt \\ = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \quad (16)$$

Both the mean and the shape of the cumulative distribution of the k_t values are observed to be functions of K_t and hour [18]. This functional dependence can be approximated with a single curve when normalized by subtracting the mean and dividing by the standard deviation

$$F_{kt} = \frac{1}{1 + \exp(-1.585h)} \quad (17)$$

where

$$h = (k_t - k_{tm}) / (\sigma_{kt}) \quad (18)$$

This analytic representation for the cumulative distribution function was developed from the Madison, WI and Albuquerque, NM TMY data. While this data is not totally independent of the long-term data, it represents only a small subset of the long-term data. This expression was fit to the data using a nonlinear least-squares regression routine; for the poorest fit, R^2 dropped from 96% to 93% when the general coefficient was substituted in place of the location-hour- K_t specific coefficients. This expression is an approximation only; a more detailed analysis of this distribution is necessary, including examination at a wider range of

climate types. k_{tm} is estimated from eqn (6), while σ_{kt} is found from an expression developed by Graham [4]:

$$\sigma_{kt} = 0.1557 \sin\left(\frac{\pi K_t}{0.933}\right) \quad (19)$$

Equation (19) was developed using Graham's expression for k_{tm} (eqn (7)). however, as shown in Fig. 3, it was found to be adequate for representing the standard deviation when k_{tm} is calculated from eqn (6). It also has the advantage of being developed from independent data. It should be noted that the distribution of k_t about k_{tm} is different from many other k_t distributions, for example, it is *not* the same as the distribution of the k_t values about their monthly-average value at an hour.

Equating the cumulative distribution functions and solving for k_t yields

$$k_t = k_{tm} - \frac{\sigma_{kt}}{1.58} \ln \left[\frac{1}{0.5 \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right] - 1} \right] \quad (20)$$

Since the series of k_t 's is not continuous, a new series of x 's must be generated each day (the last x on one day should not be used as the x_{i-1} for the first hour of the next day). The mean value of x , zero, was used for the initial value of x_{i-1} .

When the hourly radiation values are summed, the daily total of the generated radiation is not necessarily equivalent to the original 'target' value. Over a month, these discrepancies tend to average out, although it is also possible to check each day's total against its target value and scale the hourly values (by multiplying them by the ratio of the 'target' K_t value to the generated K_t value) such that the 'target' K_t value is matched exactly. The effect of this correction on the diurnal variation is insignificant; however, the hourly lag one autocorrelation is affected. An additional disadvantage of this

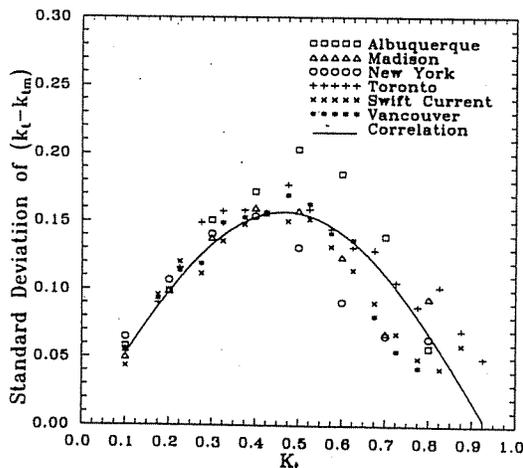


Fig. 3. Standard deviations of the deviations of k_t from k_{tm} grouped by daily clearness index. k_{tm} is calculated from eqn (6) for Albuquerque, Madison, and New York; from eqn (7) for Toronto, Swift Current, and Vancouver. Data for Toronto, Swift Current, and Vancouver is from [4].

Table 2. Comparison of monthly means of daily clearness index and daily average temperature for selected locations

	K_t (Madison)			T (°C) (Albuquerque)			
	LT	G(uc)	TMY	LT	G(c)	G(uc)	TMY
Jan	0.44	0.47	0.44	1.3	1.3	1.8	1.4
Feb	0.50	0.49	0.49	3.9	3.9	5.4	3.4
Mar	0.50	0.48	0.54	8.0	8.0	6.8	6.8
Apr	0.48	0.48	0.47	13.1	13.1	14.6	12.6
May	0.51	0.51	0.49	18.4	18.4	18.7	18.6
Jun	0.54	0.57	0.52	23.8	23.8	24.4	22.9
Jul	0.54	0.56	0.54	25.2	25.2	24.7	25.5
Aug	0.55	0.56	0.56	23.9	23.9	24.1	24.0
Sep	0.52	0.53	0.54	20.4	20.4	21.7	19.5
Oct	0.49	0.51	0.48	14.0	14.0	15.8	14.1
Nov	0.40	0.38	0.40	6.6	6.6	6.7	6.5
Dec	0.38	0.38	0.37	1.6	1.6	5.5	2.1

LT=Long-Term
G(c)=Corrected Generated
TMY=Typical Meteorological Year
G(uc)=Uncorrected Generated

correction is that the entire day's hourly clearness index values must be generated at the beginning of each day and stored. This correction was not applied to the generated radiation data reported in this paper.

6. STATISTICAL COMPARISON OF RADIATION DATA

The statistics computed from the generated data are compared to both the long-term and the TMY data. There are statistically significant discrepancies between the generated and long-term data, mainly due to the limitations of the correlations; the comparison of the generated data statistics to the TMY statistics is made to get a feeling for the magnitude of these discrepancies. A comparison of simulation results using generated and long-term data could be made; however, the information gained from such a comparison is not necessarily useful since the particular characteristics of the system would influence and confound the results. A system could be chosen for which a particular discrepancy would be significant in affecting the simulation results; similarly a different system could be chosen which is highly insensitive to the same discrepancy. For example, a solar energy system having no storage capacity is independent of the radiation autocorrelation; a system which has some storage capacity and meets a high percentage of the load is highly sensitive to the autocorrelation.

A sampling of the statistics calculated from the long-term, TMY, and generated data for Albuquerque, NM, Madison, WI, and New York, NY is shown in Tables 2 and 3. The maximum deviation of the generated

Table 3. Comparison of annual lag one autocorrelations of daily clearness index and daily average temperature

	K_t			T (°C)		
	ALB	MAD	NYC	ALB	MAD	NYC
Long Term	0.25	0.21	0.13	0.71	0.66	0.60
Generated (uc)	0.20	0.20	0.21	0.50	0.50	0.50
Generated (c)	-	-	-	0.50	0.50	0.50
TMY	0.21	0.21	0.11	0.73	0.62	0.50

ALB=Albuquerque MAD=Madison NYC=New York

monthly-average K_t values is only 0.03; that of the TMY values is only 0.04. Monthly-average K_t values for Madison are listed in Table 2.

The annual lag one autocorrelation values of the generated daily K_t series are all approximately 0.2, which is lower than the value of 0.29 that the sequence (Table 1) was chosen to create (although for these locations it actually provides better agreement with the long-term than the corrected values). The reason that the lag one autocorrelation differs from 0.29 is that the 'target' K_t value for the day is not identical to the K_t value actually generated. This difference is not expected to have a significant effect on simulation results since the autocorrelation is weak; a simple fix for this problem would be to change the sequence so as to generate a higher lag one autocorrelation value. This change was not made for the results presented in this paper so as to indicate the effect of the simulation process.

The average diurnal variations of the generated, TMY, and long-term radiation data are comparable, although the generated data is slightly better than the TMY data at replicating the long-term diurnal variation. As an example, the average diurnal variations of the generated, TMY, and long-term data for Madison, WI are shown in Fig. 4.

Examination of the RMS error associated with the K_t distributions revealed that in Madison the generated distributions provide a slightly better approximation to the long-term curves than the TMY data; in New York, the opposite is true, while for Albuquerque the TMY and generated data are equally able to reproduce the long-term distribution. The discrepancies between the generated and the long-term distributions are due to the limitations of the Bendt *et al.* model [24]. As shown in Fig. 5, a worst case example is the August distribution for New York; the long-term distribution is noticeably different from the distribution obtained from the generated data, while the TMY data in this instance tends to better replicate the shape of the long-term curve. As shown in Fig. 6, February in Madison

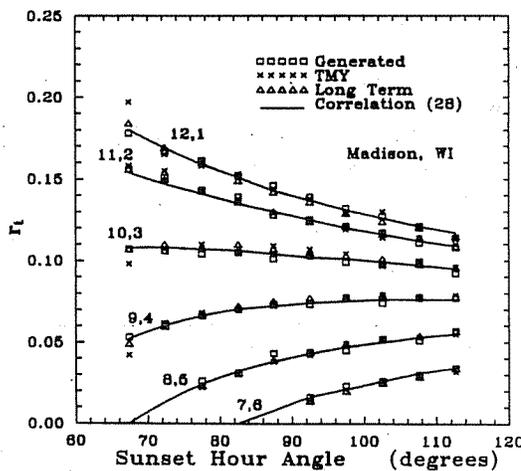


Fig. 4. Comparison of long-term, generated and TMY radiation diurnal variations for Madison, WI.

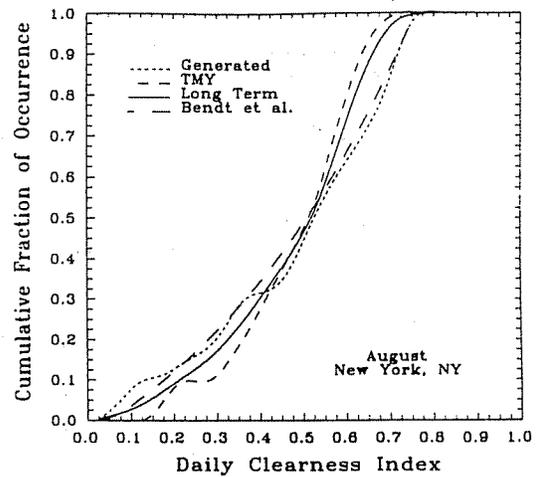


Fig. 5. Comparison of long-term, generated, and TMY daily clearness index distributions for a month in which the Bendt *et al.* correlation [24] differs from the long-term distribution.

is a month in which the opposite is true; the generated distribution more closely approximates the long-term distribution than does the TMY data.

7. GENERATION OF AMBIENT TEMPERATURE DATA

To generate ambient temperature values, Degelman [16,17] first selects the daily average and maximum temperatures from a normal distribution and orders them according to predetermined 31-day sequences similar to the manner used for daily radiation data generation. The particular sequences used by Degelman are shown in Table 1. The hourly temperatures are then obtained by fitting a cosine curve between the maximums and minimums. This deterministic approach does not attempt to preserve the autocorrelation structure of the hourly ambient temperatures or the

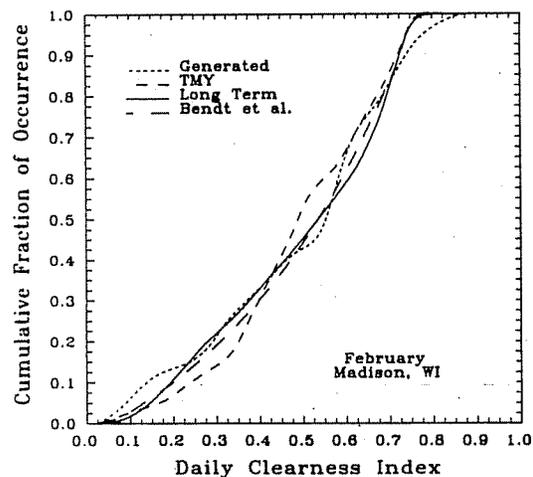


Fig. 6. Comparison of long-term, generated, and TMY daily clearness index distributions for a month in which the Bendt *et al.* correlation [24] accurately reproduces the long-term distribution.

distribution and diurnal variation documented by Erbs *et al.* [36]. Hollands *et al.* [38] studied the effect of neglecting the random component in hourly temperature data for various solar process heat systems. They found that the error in solar fraction introduced by not including the random component for these systems is on the order of 1%; taking the ambient temperature constant at its monthly-average value resulted in errors of only as much as 5%. These results indicate that for some systems, the extra complexities of including the random component of the hourly ambient temperature is unwarranted. However, there may be systems for which its inclusion is important. The following model generates an hourly ambient temperature series with the random component included.

A logical extension of the method for generating the hourly k_i values would be to generate the hourly ambient temperatures from the daily averages in a similar manner. However, unlike the hourly clearness index series, the ambient temperature series is continuous. Once the daily average is fixed, it is difficult to generate hourly values stochastically without introducing discontinuities between the last hour of one day and the first hour of the next day.

An alternative to predetermining the daily values and ordering them with a 31-day sequence is to predetermine the 24 hourly monthly-average temperatures from the diurnal temperature variation relation (eqn (10)). The hourly temperatures can then be generated by a stochastic model, again by generating a Gaussian random variable and transforming it to a temperature value through the cumulative distribution function. The diurnal variation is maintained through the different mean value of the distribution used for each hour of the day.

First, 24 hourly monthly-average temperatures are estimated from eqn (10). \bar{T} , the monthly-average daily temperature, is known, and the amplitude, A , is estimated from eqn (11).

Each hour, a new X value is generated according to a second order autoregressive model (AR2), i.e., eqn (9) with N equal to 2. The AR2 model was found to be adequate based on analysis of the autocorrelation structure and model residuals [18] of TMY data for Albuquerque NM, Madison WI, and Miami FL. Hittle and Pedersen [11] also found an AR2 model to be appropriate for modeling hourly ambient temperature. ϕ_1 and ϕ_2 were estimated using a linear least-squares regression routine, and were found to vary from month to month and location to location with no apparent pattern; it was believed that they were to an extent compensating for each other. The increase in the residual sum of squares from using a single set of coefficients as opposed to the location and month specific coefficients is small; the residual standard deviation increased less than 1°C when values of $\phi_1 = 1.178$ and $\phi_2 = -0.202$ were used for all months and locations [18].

The generated x value is transformed to an hourly temperature by equating the cumulative distribution functions of x (eqn (16)) and hourly ambient tem-

perature. The cumulative distribution function of the hourly ambient temperature is given by eqn (12), where for this application, \bar{T}_h is substituted for \bar{T} in eqn (13). The shape of the distribution is the same throughout the month, but the mean is dependent upon the hour of the day. Solving for the hourly temperature, T , gives

$$T = \bar{T} - \frac{\sigma_m \sqrt{N/24}}{3.396} \times \ln \left[\frac{1}{0.5 \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right]} - 1 \right] \quad (22)$$

The stochastic method just described does not guarantee that the generated monthly-average daily temperature is as originally specified. This discrepancy can be eliminated by comparing the generated monthly-average to the long-term average and adding the difference to each hourly temperature. Again, a disadvantage associated with this correction is that the temperature values must be generated and stored; in this case, the entire month's hourly values must be stored. This correction was applied to the generated temperature data reported in this paper.

8. STATISTICAL COMPARISON OF AMBIENT TEMPERATURE DATA

The monthly-average ambient temperatures for the generated data and the long-term data are identical when the generated values are adjusted; the TMY monthly-average values differ by as much as 1.7°C, and the uncorrected generated values by as much as 4°C. The monthly-average temperatures for Albuquerque are listed in Table 2. The generated daily autocorrelations are low; instead of yielding values of about 0.6 or 0.7, the generated values have a mean of 0.5 (Table 3). This indicates that there is a weakness in the AR2 hourly temperature model. The following statistics are for the corrected series of ambient temperature values.

Examination of the RMS error associated with the monthly-average hourly diurnal variation of ambient temperature indicates that the TMY data is often better than the generated data at reproducing the long-term diurnal variation, the reason being the accuracy of the diurnal variation correlation (eqn (10)). As shown in Fig. 7, a worst case example is the diurnal variation for November in Albuquerque. For some months, however, the opposite is true, and the generated data more accurately represent the long-term data, for example March in Madison, as shown in Fig. 8.

The distributions of the daily-average ambient temperatures about their monthly-average value again indicate that the limitations of the generated data lie in the correlation accuracy. As shown in Fig. 9, slight discrepancies are apparent for some months, such as August in Albuquerque. For many of the months,

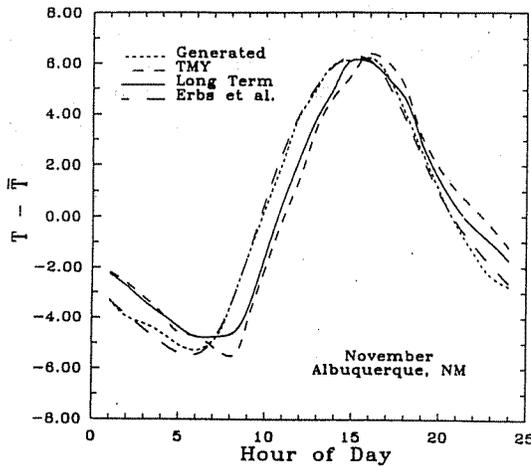


Fig. 7. Comparison of long-term, generated, and TMY monthly-average ambient temperature diurnal variations for a month in which the Erbs *et al.* correlation [36] differs from the long-term diurnal variation.

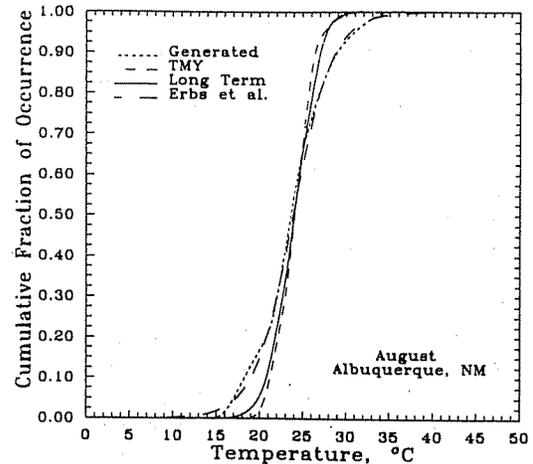


Fig. 9. Comparison of long-term, generated, and TMY daily-average ambient temperature distributions for a month in which the Erbs *et al.* correlation [36] differs from the long-term distribution.

however, the distribution replication is quite good, for example November in Madison, as shown in Fig. 10. Again, substitution of a more accurate expression for the temperature distribution will improve results, and should be done whenever available.

9. COMMENTS ON THE GENERATION OF HUMIDITY AND WINDSPEED DATA

Models for generating hourly relative humidity and windspeed were also developed by Degelman [16,17]. Analysis [18] has indicated a need for improvement of both models. In particular, the monthly-average diurnal variations and the monthly distributions of daily average values are significantly different from the long-term diurnal variations and distributions for both relative humidity and windspeed. The methods presented

in this paper for the synthetic generation of radiation and/or temperature could easily be extended. Correlations for the diurnal variation and distribution of relative humidity are presented by Erbs *et al.* [39]. Stochastic models for windspeed are available, for example [13-15].

10. CONCLUSIONS

Models for generating hourly series of radiation and ambient temperature values have been presented. The aim of these particular models is to replicate the statistics of long-term data with just one year of generated data, similar to the concept of a "Typical Meteorological Year" [19], and ultimately to reproduce long-term simulation results with one year's simulation.

The statistics presented (means, distributions,

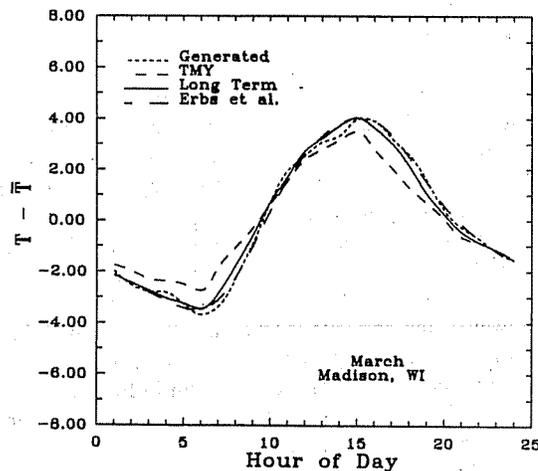


Fig. 8. Comparison of long-term, generated, and TMY monthly-average ambient temperature diurnal variations for a month in which the Erbs *et al.* correlation [36] accurately reproduces the long-term diurnal variation.

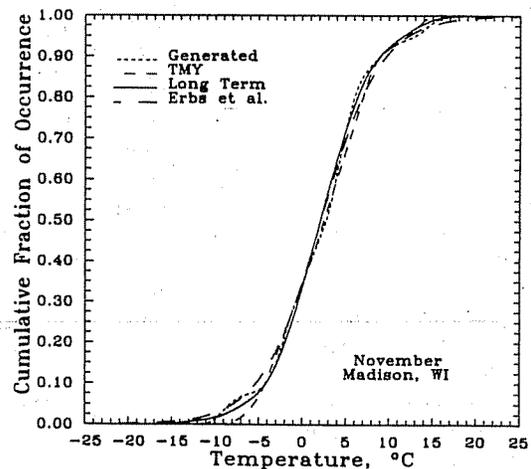


Fig. 10. Comparison of long-term, generated, and TMY daily-average ambient temperature distributions for a month in which the Erbs *et al.* correlation [36] accurately reproduces the long-term distribution.

diurnal variations and autocorrelations) indicate that the accuracy of the generated data in reproducing the statistics is on a par with that of the TMY data. Most of the current discrepancies are due to the limitations of the existing correlations; by improving these correlations, the accuracy of the generated data can be increased.

The main emphasis of this paper is on the methodology used to produce a Type II year of generated data. The weakness in this type of weather data generator is its inability to reproduce weather characteristics that are not classified in a location-independent form. For example, it is not capable of reproducing location-specific peculiarities such as morning fog. Cross-correlations, which vary from location to location and month to month, are also not directly modeled.

This approach to weather data generation offers many advantages: simulations run with generated data for one year periods will yield results quite similar to those which would be obtained from simulations driven by many years of recorded (or Type I synthetic) data but will require significantly less computation. Only limited meteorological information will be needed as inputs to synthesize data, so simulations will be able to be run in many more locations than presently possible. Finally, the inconvenience of handling large amounts of weather data will be avoided. However, further verification to a wider range of climate types is necessary before general universality can be claimed.

NOMENCLATURE

A	long-term monthly average amplitude of ambient temperature, °C
F_{k_t}	cumulative distribution function of the hourly clearness index
F_{normal}	cumulative distribution function of a normally distributed variable
F_{temp}	cumulative distribution function of hourly ambient temperature
H	daily global solar radiation on a horizontal surface
H_o	daily extraterrestrial global solar radiation on a horizontal surface (see [22])
I	hourly global solar radiation on a horizontal surface
I_o	hourly global extraterrestrial solar radiation on a horizontal surface (see [22])
\bar{I}	monthly-average hourly total radiation at a particular hour
I_{LT}	long-term average value of hourly total radiation for a given location, day, hour, and daily total radiation value
k_t	hourly clearness index; the ratio of hourly global radiation on a horizontal surface to hourly extraterrestrial radiation, I/I_o
k_{tm}	long-term average value of the hourly clearness index for a particular daily clearness index, sunset hour angle and hour angle
K_t	daily clearness index; ratio of daily global solar radiation on a horizontal surface to daily extraterrestrial radiation, H/H_o
\bar{K}_t	monthly average clearness index; ratio of monthly-average global solar radiation on a horizontal surface to monthly-average daily extraterrestrial radiation, \bar{H}/\bar{H}_o
N_m	number of hours in the month
p	indicates the lag in an autocorrelation computation

r_t	ratio of hourly global solar radiation on a horizontal surface to daily global radiation on a horizontal surface, I/H
t	time (hours)
T	hourly ambient temperature
\bar{T}	monthly-average daily ambient temperature
T_h	monthly-average ambient temperature at a particular hour

Greek

χ	normally distributed stochastic variable with a mean of 0 and a variance of 1
ϵ	normally distributed random disturbance with mean of 0 and variance of σ^2
ϕ	coefficient in autoregressive stochastic model
ρ	autocorrelation coefficient, subscript indicates lag
σ^2	variance of the normally distributed random disturbance ϵ_t
σ_{k_t}	standard deviation of hourly clearness indices (k_t) about their long-term average value (k_{tm})
σ_m	standard deviation of a month's daily average ambient temperature about the long-term average value for that month
$\sigma_{m,h}$	standard deviation of a month's average ambient temperature at a particular hour about the long-term average value for that month at that particular hour
σ_{yr}	standard deviation of the 12 monthly-average daily ambient temperatures about the yearly average daily temperature
ω	hour angle, in degrees
ω_s	sunset hour angle, in degrees (see [22])

REFERENCES

1. P. W. Suckling, Estimating daily solar radiation values in selected mid-latitude regions by extrapolating measurements from nearby stations, *Solar Energy* 35, 491-495 (1985).
2. V. A. Graham, K. G. T. Hollands, and T. E. Unny, Stochastic variation of hourly solar radiation over the day, *Advances in Solar Energy Technology: Proceedings of the Biennial Congress of the International Solar Energy Society*, Vol. 4, 3796-3800 (1987). Also see ref. [4].
3. V. A. Graham, K. G. T. Hollands, and T. E. Unny, Stochastic modeling of the daily solar atmospheric transmittance, K_t , *Intersol 85: Proceedings of the Ninth Biennial Congress of the International Solar Energy Society*, Vol. 4, 2473-2477 (1985).
4. V. A. Graham, Stochastic synthesis of the solar atmospheric transmittance, *Ph.D. Thesis*, Mechanical Engineering Dept., University of Waterloo, Waterloo, Ontario, Canada (1985).
5. B. J. Brinkworth, Autocorrelation and stochastic modelling of insolation sequences, *Solar Energy* 19, 343-347 (1977).
6. U. Amato, A. Andretta, B. Bartoli, B. Culuzzi, V. Cuomo, F. Fontana, and C. Serio, Markov processes and Fourier analysis as a tool to describe and simulate daily solar irradiance, *Solar Energy* 37(3), 179-194 (1986).
7. V. A. Graham, K. G. T. Hollands, and T. E. Unny, A time series model for K_t with application to global synthetic weather generation, *Solar Energy* 40(2), 83-92 (1988).
8. R. J. Aguiar, R. Collares-Perreira, and J. P. Conde, Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices, *Solar Energy* 40(3), 269-279 (1988).
9. J. M. Gordon and T. A. Reddy, Time series analysis of daily horizontal solar radiation, *Solar Energy* 41(3), 215-226 (1988).
10. J. M. Gordon and T. A. Reddy, Time series analysis of hourly global horizontal solar radiation, *Solar Energy* 41(5), 423-429 (1988).

11. D. C. Hittle and C. O. Pedersen, Periodic and stochastic behavior of weather data, *ASHRAE Trans.* **87**(2), (1981).
12. M. J. Brook and B. A. Finney, Generation of bivariate solar radiation and temperature time series, *Solar Energy* **39**(6), 533-540 (1987).
13. A. Baklouktis, D. Tsanakas, and G. Vachtsevanos, Stochastic simulation of hourly and daily average wind speed sequences, *Wind Engineering* **10**(1), 1-11 (1986).
14. K. C. Chou and R. B. Corotis, Simulations of hourly wind speed and array wind power, *Solar Energy* **26**(3), 199-212 (1981).
15. M. Blanchard and G. Desrochers, Generation of auto-correlated wind speeds for wind energy conversion system studies, *Solar Energy* **33**(6), 571-579 (1984).
16. L. O. Degelman, A weather simulation model for building energy analysis, *ASHRAE Trans.*, Symposium on Weather Data, Seattle, WA, 435-447 (1976).
17. L. O. Degelman, Monte Carlo simulation of solar radiation and dry bulb temperatures for air conditioning purposes. Report No. 70-9, Sponsored by National Science Foundation under Grant No. GK-2204, Department of Architectural Engineering, The Pennsylvania State University, University Park, PA (September 1970).
18. K. M. Knight, Development and validation of a weather data generation model, *M.S. Thesis*, Mechanical Engineering Dept., University of Wisconsin-Madison, Madison, WI (1988).
19. I. J. Hall, R. R. Prairie, H. E. Anderson, and E. C. Boes, Generation of a typical meteorological year, *Proceedings of the 1978 Annual Meeting*, American Section of the International Solar Energy Society, Vol. 2.2, 669-671 (1988).
20. W. Rudloff, *World-climates: With tables of climatic data and practical suggestions*, Wissenschaftliche Verlagsgesellschaft, Stuttgart (1981).
21. SOLMET, Volume 2—Final Report, Tape Deck 9724, U.S. Department of Commerce, National Climatic Center, Asheville, NC (1979).
22. J. A. Duffie and W. A. Beckman, *Solar engineering of thermal processes*, Wiley, New York (1980).
23. B. Y. H. Liu and R. C. Jordan, The interrelationship and characteristic distribution of direct, diffuse and total solar radiation, *Solar Energy* **4**(1), 1-19 (1960).
24. P. Bendt, M. Collares-Pereira, and A. Rabl, The frequency distribution of daily insolation values, *Solar Energy* **27**(1), 1-5 (1981).
25. K. G. T. Hollands and R. G. Huget, A probability density function for the clearness index, with applications, *Solar Energy* **30**(3), 195-209 (1983).
26. M. E. Herzog, Estimation of hourly and monthly average daily insolation on tilted surfaces, *M.S. Thesis*, Trinity University, San Antonio, TX (1985).
27. G. Y. Saunier, T. A. Reddy, and S. Kumar, A monthly probability distribution function of daily global irradiation values appropriate for both tropical and temperate locations, *Solar Energy* **38**, 169-177 (1987).
28. R. W. Stuart and K. G. T. Hollands, A probability density function for the beam transmittance, *Solar Energy* **40**(5), 463-467 (1988).
29. M. Collares-Pereira and A. Rabl, The average distribution of solar radiation—Correlations between diffuse and hemispherical and between daily and hourly insolation values, *Solar Energy* **22**(2), 155-164 (1979).
30. H. C. Hottel and A. Whillier, Evaluation of flat-plate collector performance, *Transactions of the Conference on the Use of Solar Energy*, Vol. 2, 74-104 (1955).
31. G. E. P. Box and G. M. Jenkins, *Time series analysis: Forecasting and control*, Holden-Day, Oakland (1976).
32. G. M. Jenkins and D. G. Watts, *Spectral analysis and its applications*, Holden-Day, San Francisco (1968).
33. S. A. Klein, A design procedure for solar heating systems, *Ph.D. Thesis*, Chemical Engineering Dept., University of Wisconsin-Madison, Madison, WI (1976).
34. P. R. Armstrong, How weather persistence affects performance of solar heating systems, *M.S. Thesis*, Mechanical Engineering Dept., Colorado State University, Fort Collins, CO (1980).
35. S. A. Klein and W. A. Beckman, Loss-of-load probabilities for stand-alone photovoltaic systems, *Solar '87: Proceedings of the 1987 Annual Meeting*, American Solar Energy Society, Solar Energy Society of Canada, 18-25 (1987).
36. D. G. Erbs, S. A. Klein, and W. A. Beckman, Estimation of degree-days and ambient temperature bin data from monthly-average temperatures, *ASHRAE Journal* **25**(6), 60-65 (1983). Also see ref. [40].
37. S. A. Klein and W. A. Beckman, Review of solar radiation utilizability, *Journal of Solar Energy Engineering* **106**, 393-402 (1984).
38. K. G. T. Hollands, L. T. D'Andrea, and I. D. Morrison, Effect of random fluctuations in ambient air temperature on solar system performance, *Solar Energy* **42**(4), 335-338 (1989).
39. D. G. Erbs, S. A. Klein, and W. A. Beckman, A simple distribution method for two-dimensional temperature/humidity bin data, *ASHRAE Transactions* **91** (1985). Also see ref. [40].
40. D. G. Erbs, Models and applications for weather statistics related to building heating and cooling loads, *Ph.D. Thesis*, Mechanical Engineering Dept., University of Wisconsin-Madison, Madison, WI (1984).