Solar Energy Vol. 44, No. 6, pp. 333-341, 1990
Printed in the U.S.A.

0038-092X/90 $3.00 + .00

A METHOD TO GENERATE SYNTHETIC HOURLY SOLAR RADIATION GLOBALLY

V. A. GRAHAM* and K. G. T. HOLLANDS,†
*The University of the West Indies, St. Augustine, Trinidad, W.I.,
†University of Waterloo, Waterloo, Ontario Canada

Abstract—This paper develops a stochastic procedure for generating synthetic sets of hourly solar irradiation values, suitable for use in solar simulation design work. The daily atmospheric transmittance K_t for the day in question is broken down into hourly irradiation events by a stochastic disaggregation procedure. The necessary disaggregation models were constructed using the quasi-universal hourly atmospheric transmittance k_t, instead of the irradiation itself, as the random variable. Analyses of meteorological records revealed that both the marginal probability and stochastic features of sets of hourly events k_t within an individual day can be closely predicted using the K_t for the day. Results indicated that the parameters defining the disaggregation procedure were independent of geographical location. The values of k_t are found to be closely modeled by a Beta distribution, and a mean correlation coefficient between successive hourly values of k_t was found to be 0.54 ± 0.14. An algorithm is described from which unlimited hourly solar irradiation data may, therefore, be generated using only the 12 monthly means of daily events, $\bar{K}_t$.

1. INTRODUCTION

The analysis and design of solar energy converters by computer simulation generally requires a time sequence of hourly solar irradiation values for the location, to be used as input to the solar simulation program. For many locations around the world, however, historic records of sufficient detail are either not available or the limited data that are available are of questionable quality and have an appreciable number of missing values. (Obtaining quality records, without missing values, is indeed rare.) Thus the range of application of these popular and very useful computer simulation design tools is compromised.

But computer simulation design work need not necessarily be limited to using *measured* historic records as input. Recently, Graham *et al.*[1] proposed replacing the historic record with *synthetic* sets of irradiation values, generated using mathematical models of the irradiation process. They outlined a methodology for producing synthetic irradiation data, based on the theory of autoregressive time series[2]. Agiar *et al.*[3] recently described a methodology with similar capabilities, based on "transition probabilities."

The methology described by Graham *et al.* was incomplete. They described a procedure capable of generating synthetic sets of *daily* irradiation values. In order to meet the need of simulation programs requiring *hourly* irradiation values, this paper extends their daily irradiation model, by providing stochastic models for the hourly solar irradiation, given the daily values.

1.1 *Previous work*

Previous researchers have also attempted to develop models for the hourly irradiation. Goh and Tan[4] used Autoregressive-Moving Average (ARMA) techniques[2] for an annual sequence of hourly irradiation data from Singapore; Paassen[5] modelled irradiation data from Holland by detrending daily set of hourly data and representing the hourly residuals as an uncorrelated set; Hittle and Pedersen[6] studied an annual catenated series of hourly beam irradiation and used a Fourier series to describe deterministic trends in the annual set, and the ARMA stochastic process for the random component; Balouktsis and Tsaides[7] removed daily and annual periodicity from hourly irradiation in Greece, using Fourier analysis, and modeled a catenated series of hourly irradiation residuals.

The major shortcoming of these studies is that they studied the irradiation variable itself. The solar irradiation variable is specific to a given location, and models developed without a proper accounting for local latitude and changes in the local atmospheric states are bound to be applicable only to the locations studied. We note that solar irradiation outside the earth's atmosphere can be accurately predicted for any hour; it is the atmospheric transparency that induces the randomness. These arguments were recognised early by Liu and Jordan[8] who studied the statistical features of solar radiation, by treating the atmospheric transmittance as the random variable, instead of the irradiation variable itself. (The atmospheric transmittance, also called the clearness index, is denoted k_t for hourly events and K_t for daily events.) Liu and Jordan reported that the probability characteristics of this variable behaves in a quasi-universal manner.

Dagelman[9] attempted to model the hourly event using the atmospheric transmittance variable. He generated daily events using the probability distribution curves, and estimated hourly irradiation values from these. His research, however, was not based on established stochastic methodology: the correlation between daily events was determined by random drawings from a Gaussian distribution; and the statistical and stochastic features between hourly events were not evaluated.

Exell[10] modelled Thailand data using the ratio of the hourly irradiation to the clear sky irradiation, which is close to the atmospheric transmittance variable. But he did not account for autocorrelations in the random component of the variable they studied. Mustacchi *et al.*[11] used ARMA modelling for the hourly atmospheric transmittance variable. They did so, however, by catenating sets of hourly events from consecutive days to form an annual series of contiguous events. Catenating daily sets in this manner, by joining sunset and sunrise hourly atmospheric transmittance data from consecutive days, will alter the correlation structure, and hence, produce nonrepresentative models. Applying a single ARMA stochastic process to the catenated series also assumes (incorrectly) that the statistical and stochastic features of hourly events are constant over all parts of the contiguous sequence of events. Adequate representation of the hourly process requires that any variability in the sequence that are peculiar to different parts of the series be correctly represented.

1.2 *Varying probability features of the Hourly Atmospheric Transmittance*

The progression of hourly values of k_t could be fully described solely from models for their distribution function, if these hourly events, were independent. But for nonindependent events, such as occurrences of k_t, additional information on the factors affecting the probability of occurrence is needed. Stochastic model building is used to identify these factors and incorporate them into mathematical expressions.

Classical stochastic model building (e.g., Box–Jenkins techniques[2]) most often uses the premise that the (marginal and serial) probability structure of the random variable process is time-variant (i.e., the series is stationary); the only factor (apart from the frequency distribution) influencing the probability of the random variable is its relationship to events in the previous time intervals. The progression of k_t itself, however, cannot be described by such a typical process. For it is known that the probability of daily events change on a monthly basis, and there is also evidence that the probability of a specific k_t value will depend on the daily clearness index K_t for the day for which it occurred. For example, the probability of obtaining high values for k_t for a clear day is greater than for a cloudy day. This means that the individual hourly events within a given day cannot be dissociated from the aggregate daily event. Hence, the probability space for certain sections of the k_t series may not be the same as for other sections. (The series is not stationary.) Further, the k_t probability space may even change for periods shorter than one day; it may depend on the position of the sun in the sky when a particular k_t occurs. It follows, therefore, that models which are built by simply representing a contiguous series of hourly events by a stationary stochastic process will most likely fail to reproduce the probability feature inherent in the historic set.

This paper describes a study aimed at developing a time series model for the k_t sequence that is flexible in nature, and incorporates the varying nature of its probability features. To build such a model, we studied separately the effects of process history, daily clearness index, and solar zenith on the progression of hourly events. In this way the statistical and stochastic features peculiar to differing locations and differing levels of overall daily clearness index were evaluated. Of course the models thus developed will require values of K_t, as input, before they can be useful in generating hourly events. These daily clearness indices may themselves be generated using a separate stochastic model, such as was developed previously[1]. Alternatively, they could be obtained using an excellent historic daily record, where they exist. Full details of the model development are given in the thesis of Graham[12].

2. DISAGGREGATION OF K_t: TREND AND RANDOM COMPONENT

For each day there exists an ordered set $\{k_t\}$ of hourly values of the atmospheric transmittance values k_t. Given such a set and the corresponding set of ordered extraterrestrial irradiation values then the daily atmospheric transmittance K_t takes on a unique value. Thus, for a given location and date, for every $\{k_t\}$ there exist a unique value for K_t. Our purpose here is to obtain mechanisms for performing the inverse operation: that of obtaining the set $\{k_t\}$ from knowledge of K_t. This work involves disaggregation (i.e., distributing a variable of fixed aggregate over the hours of the day), using stochastic means. The concept of stochastic disaggregation is not new; it has been applied, for example in hydrology, to break down annual streamflow estimates into probable monthly estimates[13,14].

Models have been reported in the literature[15,16,17] which may be used in calculating the set $\{k_t\}$ on the basis that the day has a clear sky. These deterministic models implicitly assume that random fluctuations are small and can be ignored. For nonclear skies, however, the random fluctuations in k_t are more important, and their effect on simulated solar device performance can be significant[12].

The variation in k_t events consists of two components: a trend (or mean) component and a random component:

$$k_t = k_{tm} + \alpha \tag{1}$$

The trend component k_{tm} represents the atmospheric transmittance if the presence of radiation attenuators were uniformly distributed over the day; it is the general course observed in a trace of a $\{k_t\}$ set. The random component α incorporates the effect of unpredictable perturbations in the radiation attenuators, chiefly caused by varying cloud cover. For a clear day (little or no clouds) the random component will be small, but for nonclear days this component is significant.

Disaggregation will require identifying and characterizing the sets $\{k_{tm}\}$ and $\{\alpha\}$, for all possible values of K_t. Models for $\{k_{tm}\}$ and $\{\alpha\}$ will be constructed by studying the hourly historic records. We used ten years of historic hourly irradiation values measured at three Canadian cities having widely differing climate characteristics: Vancouver (coastal); Swift Current (continental); and Toronto (mixed). Data for these locations were obtained from the Canadian Atmospheric Environment Service. By isolating $\{k_t\}$ sets drawn from days with similar K_t, the effect of varying overall daily clearness were assessed.

3. MODEL CONSTRUCTION FOR THE TREND COMPONENT

The component k_{tm} may be itself split into two components: the beam trend component k_{bm} and the diffuse trend component k_{dm}. Thus, $k_{tm} = k_{bm} + k_{dm}$. Hottel[16] studied the beam trend component for a clear atmosphere. By treating the clear-day atmosphere as a spectral mixture of black, grey, and clear parts, he obtained a simple exponential equation for modelling the beam transmittance as a function of air mass. Following this lead, we apply a similar equation for the beam trend component, extending its use to all sky conditions: $k_{bm} = a_o + a_1 \exp(-\kappa m)$, where m is the air mass, taken at the center of the hour. We accommodate the general character of the sky by making the parameters a_o, a_1, and κ (unique) functions of the daily atmospheric transmittance, K_t.

The diffuse transmittance k_d is reported in the literature to bear a strong dependence on the beam transmittance k_b. Various researchers (e.g., Hollands[18], Orgill and Hollands[19], and Iqbal[20] have studied this dependence and have reported mathematical equations for its functional form: $k_d(k_b)$. This work suggests that k_{dm} may be a similar function of k_{bm}. However, by studying trends in the $\{k_t\}$ sets at different levels of K_t, we observed that variations in k_{tm} within a given day were relatively modest, and that k_{tm} maintained the decaying exponential character of k_{bm} with increasing air mass. Thus, within a given day, a simple linear approximation was applied to describe the relationship between k_{dm} and k_{bm}, in which the constants depend upon the K_t for the day.

It follows from all this that k_{tm} should be of the form:

$$k_{tm}(t) = \lambda + \epsilon \exp(-\kappa m) \tag{2}$$

in which the parameters λ, ϵ, and κ are unique functions of the clearness index K_t for the day in question.

3.1 *Parameter estimation*

The parameters λ, ϵ, and κ can be estimated from measured radiation data by a non-linear regression analysis of k_{tm} against m, using eqn (2) as the model equation. This regression analysis, however, can equally use k_t, rather than k_{tm} as the dependent variable, since α averages to zero. Indeed the regression analysis we used found the values of λ, ϵ, and κ which minimized, over a large set of observations, the rms value of $k_t - k_{tm}$, where k_{tm} is given by eqn (2).

In the regression procedure we sought to relate λ, ϵ, and κ to the overall K_t. One way of doing this would be to form sets of days in which K_t takes on specific values (or, more precisely, in which K_t takes on values inside a narrow range), then estimate λ, ϵ, and κ by a regression analysis on that set, repeat the process for various specific values of K_t, then develop empirical expressions which describe how the parameters depend on K_t. But this approach would not necessarily yield optimum estimates. Better estimates are possible if the expirical expressions relating λ, ϵ, and κ to K_t are obtained first, and then incorporated in the regression work. Thus, the regression analysis would not estimate λ, ϵ, and κ directly, but rather the unknown constants in their respective empirical expressions.

In order to develop these empirical expressions for $\lambda(K_t)$, $\epsilon(K_t)$, and $\kappa(K_t)$, we considered the physics of solar radiation transport through the atmosphere, using Hottles mixed-gas model for the atmosphere as a basis. The k_{tm} model chosen above requires (for any given day) k_{tm} to have a fixed (and minimum) component (λ) regardless of the solar zenith, even at infinite air mass. This component represents the transparent spectral window which models the transport of radiation through the atmosphere. The size of this window will depend on the clearness of the day. As K_t approaches unity, the atmosphere will become almost transparent, all the radiation is assumed to pass through the window; thus λ should approach K_t, while κ and ϵ will approach zero. As K_t approaches zero, on the other hand, the atmosphere become almost opaque (the black component of the gas model dominates), as radiation extinction becomes large and the window is assumed to close; thus λ must approach K_t, while κ must approach infinity (ϵ is finite, but $\epsilon \exp(-\kappa m)$ will also go to zero). Following a series of preliminary model discrimination studies, and guided by these arguments, we selected the following empirical expressions: $\lambda(K_t) = K_t + b_0 K_t^3(1 - K_t)$; $\epsilon(K_t) = b_1(1 - K_t)$; and $\kappa(K_t) = b_2(1 - K_t)/K_t$.

The constants b_0, b_1 and b_2 in k_{tm} were estimated using the irradiation data from the three cities. Results obtained showed the k_{tm} model to be highly significant; for example, the estimates b_0, b_1, and b_2 were also statistically very significant, being many times the standard error associated with the estimates. (These estimates were also observed to be highly correlated between themselves; thus modest changes in one, produced a compensating change in another, with no change in the total variations explained by the model.) When the parameter estimation process was repeated, using a merged data set from the three cities, we observed that the variations explained at each city by the resultant model was not statistically different from that explained using a model developed specifically for that city. The estimates could therefore be assigned fixed values that were independent of the cities studied. The regression results, with parameter estimates, obtained using the merged data set are shown in Table 1.

Results showed that the chosen model for k_{tm} is extremely significant, explaining approximately 96% of the sum of squares in the k_t data. In summary the trend transmittance k_{tm} is therefore represented by

Table 1. Summary results of nonlinear least-square parameter estimation

Constant	Estimate	Standard error
b_0	−1.167	0.066
b_1	0.979	0.049
b_2	1.141	0.038

eqn (2) with

$$\lambda(K_t) = K_t - 1.167K_t^3(1 - K_t);$$
$$\epsilon(K_t) = 0.979(1 - K_t);$$
$$\kappa(K_t) = 1.141\ (1 - K_t)/K_t. \qquad (3)$$

(For solar engineering calculations in which the effect of the random fluctuations in the hourly events is not important, e.g., passive heating of building[21], it is only necessary to estimate hourly irradiation trends during the individual days. Although workers have sometimes used correlations proposed [22] and [23] for estimating the long-term mean monthly hourly transmittance $\bar{k}_t$ from long-term daily transmittance $\bar{K}_t$, up to now, there has been no equation available for extracting the instantaneous hourly trend transmittances k_{tm} for an individual day of known K_t (a different problem). Equations (2) and (3) for k_{tm} should satisfy this need. These equations may also be useful in this problem of estimating $\bar{k}_t$ for a given $\bar{K}_t$. We note that the expectation of k_t for an hour is the same as the expectation of k_{tm} for that hour (since the expectation of α is zero) and that the expectation of k_{tm}, for a fixed time, is dependent only on the random variable K_t. Thus, using standard probability methods, and a probability model for the monthly distribution of K_t[23,24], the long-term expectation of k_t can be evaluated.)

4. MODEL CONSTRUCTION FOR THE RANDOM COMPONENT

In a vast number of solar engineering work the random fluctuations about the trends cannot be ignored. Research by Graham[12] has shown that these fluctuations effect the thermal performance of a number of solar energy devices (particularly those in which thermal output requires that the hourly irradiation exceed some threshold value). The performance in a given hour of solar conversion system with energy storage elements depends on how much solar energy was available in previous hours. It follows that the effect of past hourly fluctuations (i.e., α) on a current hourly event must be described to model these system. Stochastic modelling of α will do this.

4.1 *Effect of* K_t

To build a flexible stochastic model for α, we investigated the effect of process history, air mass, and daily clearness index. Having established k_{tm}, it is possible to find α for any hourly event from $\alpha = k_t - k_{tm}$. Hourly sets $\{\alpha\}$ of historic data were drawn from days with approximately equal daily clearness K_t, and grouped. Sixteen such groups, denoted $(\{\alpha\}:K_t)$, were formed from days with K_t centered at 0.1 and 0.175, 0.225, 0.257, . . . 0.875 (e.g., the 3rd group had $0.20 < K_t \le 0.25$). Models were developed by studying the statistics and stochastic features of each $(\{\alpha\}:K_t)$ group separately. We describe first the results obtained by relaxing the hypothesis of an air mass effect. Thus, for these results the probability (statistical and stochastic) features within $(\{\alpha\}:K_t)$ was assumed to depend only on the marginal distribution and serial relationship between events, and not on the time of day that the event occurs. The basic statistics of the fluctuations in α were investigated by studying its mean (presumably zero) and its standard deviation.

It was found that the standard deviation σ_α of α varied from group to group; σ_α is strongly dependent on K_t. The dependence appeared to be nearly sinusoidal (Fig. 1): σ_α was at its minimum, of approximately 0.05, when K_t was 0.1 (very cloudy days) and when K_t was greater than 0.85 (very clear days). However, as K_t increased above 0.1, σ_α increased reaching a maximum of approximately 0.16, at K_t approximately equal 0.45. It then decreased again as K_t increased from 0.45 to 0.85. Regression analysis of the $\{\sigma_\alpha, K_t\}$ data set indicated that the variations in σ_α could be described very well using

$$\sigma_\alpha = 0.16\ \sin(\pi K_t/0.90) \qquad (4)$$

This model successfully explained in excess of 97% of the total sum of squares of α, taken over all the data. A plot of this regression model together with the data from the different cities is shown in Fig. 1. The same statistical analysis also ensured that the mean of α was indeed statistically zero, over the complete range of data used. (The regression technique used in developing had k_{tm} required that an adequate model produce a zero mean for α.)

The effect of past values of α on the current value was studied by constructing stochastic models. These models were constructed for each group, thus allowing an investigation of the effect of daily clearness on the stochastic behaviour. As a first step, the serial autocorrelation coefficient r between values separated by one hour was estimated for each group. An estimate of zero for r suggests that past values do not influence current values, while a non-zero estimate suggests the opposite. For each group, pairs of $(\alpha(t), \alpha(t-1))$, within a given day, were formed, index t represents the hour. Care was taken not to take pairs from different days; thus a sunrise value was not paired with a sunset value from the previous day. By merging all the acceptable pairs of $\alpha(t)$ and $\alpha(t-1)$, within each group, the r value was estimated for each group.

The estimates of r were found to be non-zero, for all sixteen groups, and fairly independent of group, i.e. of value of K_t. The mean of all group estimates was 0.54, and assuming a constant value of r of 0.54 explained 97% of the total sum of squares of variations,

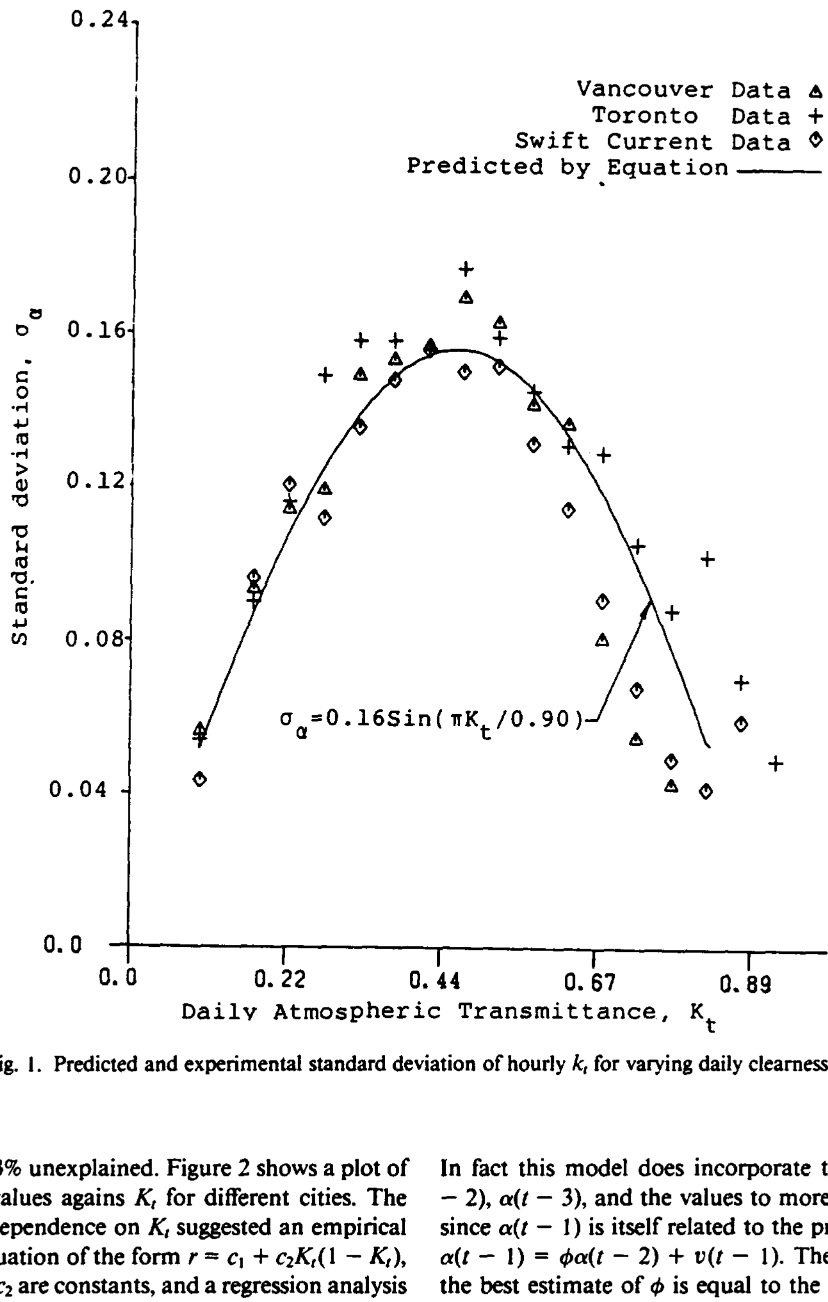

Fig. 1. Predicted and experimental standard deviation of hourly k_t for varying daily clearness K_t.

leaving only 3% unexplained. Figure 2 shows a plot of estimated r values agains K_t for different cities. The form of the dependence on K_t suggested an empirical modelling equation of the form $r = c_1 + c_2K_t(1 - K_t)$, where c_1 and c_2 are constants, and a regression analysis gave $c_1 = 0.35$ and $c_2 = 1.1$. This regression equation, plotted in Fig. 2, explained 62% of the remaining 3% sum of squares of the variations in r data.

Theoretical stochastic models for the effect of past values are most often represented by expressing the current value as a linear sum of weighted past values, plus a random value (v) which is drawn from an uncorrelated random set; that is, $\alpha(t) = \phi_1\alpha(t-1) + \phi_2\alpha(t-2) + \phi_3\alpha(t-3) + \cdots v(t)$. The coefficients ϕ_i will generally decay with increasing value of i. Based on past experience with forecasting in engineering application, we felt it was unnecessary to represent $\alpha(t)$ using a model which required more than $\alpha(t-1)$, the value in the immediately preceding hour. Thus we propose the simplest such model:

$$\alpha(t) = \phi\alpha(t-1) + v(t). \tag{5}$$

In fact this model does incorporate the effect of $\alpha(t-2)$, $\alpha(t-3)$, and the values to more distant events, since $\alpha(t-1)$ is itself related to the previous hour by $\alpha(t-1) = \phi\alpha(t-2) + v(t-1)$. Theory shows that the best estimate of ϕ is equal to the autocorrelation coefficient r. Since the standard deviation σ_α changes with K_t (eqn (4)) the standard deviation σ_v governing the distribution of $v(t)$ must also change with K_t. The theoretically derived relationship is given by

$$\sigma_v = \sigma_\alpha(1 - \phi^2)^{1/2}. \tag{6}$$

4.2 *Effect of air mass*

The previous analysis assumed that the probability of α does not depend on air mass, m (or time of day). In this section we assess this assumption by examining the σ_α and ϕ stochastic parameter, for different values of air mass. Following a similar procedure to that described in the previous sections the hourly values of α within groups of K_t were further stratified into subgroups, defined according to air mass, or more specifically zenith angle $\theta_z = \sec^{-1} m$. The zenith angle

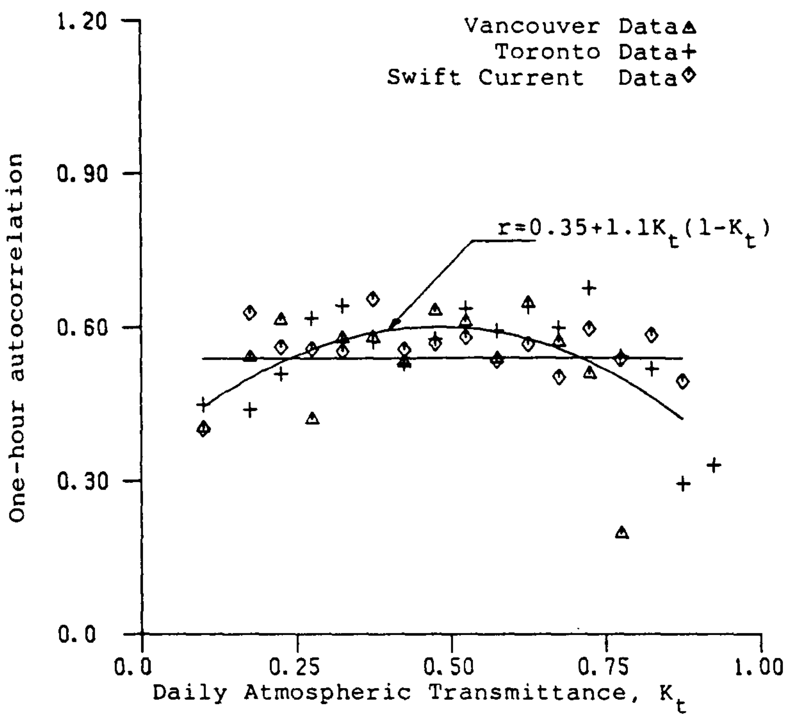

Fig. 2. Predicted and experimental serial correlation between hourly k_t separated by one hour vs. K_t.

form 0° to 90° was divided into 10 intervals, and pairs of α values corresponding to successive events which belongs to adjacent θ_z-intervals formed. (Special care was taken not to lose the order of occurrence of these hourly events; if two consecutive events could not be classified into adjacent zenith angle intervals, they were not used.) The autocorrelation coefficient r and the standard deviation of α for groups of adjacent zenith angle intervals were computed, using these pairs.

The results indicated a mild (but consistent) increase in σ_α with increasing zenith angle, but r was observed to be generally independent of zenith angle, with no consistent pattern being observed. The average variation in σ_α within a day was of the order of 25% of average σ_α with this pattern being observed in all K_t groups. This information suggests a small zenith angle effect on the probability behaviour of atmospheric transmittances, but the importance of this effect on engineering simulation predictions is doubtful.

4.3 *The uncorrelated random component*

To generate the daily sets α, the only further requirement is a definition of its (marginal) probability distribution. This is fixed by the distribution of v, the set of uncorrelated random values used to generate α from eqn (5). The standard deviation σ_v of v is obtained from eqn (6) with σ_α given by eqn (4).

It is popular in stochastic modelling to use a Gaussian distribution to represent the random variable v. This will (by mathematical arguments) produce a Gaussian distribution for α (and hence k_t). We investigated the frequency distribution of α by studying its frequency histogram for differing conditions, and found it to be non-Gaussian. Thus it was necessary to build a model for α's distribution function.

Except for a constant, the frequency of α is identical to the frequency of the corresponding k_t. For convenience, we address the probability features of k_t and apply our conclusions to α. A preliminary investigation in the marginal probability distribution of hourly values of k_t was performed by studying their frequency histograms within the groups of constant K_t. These histograms (Figs. 3, 4, 5) show that the distribution is unimodal and symmetrical about its mean, for K_t approximately equal to 0.45; for K_t less than 0.45, it was right-skewed, and left-skewed for greater values of K_t. The peakedness (the 4th central moment) increases as the distribution becomes more assymetrical (i.e., as K_t moves from approximately 0.45 and approaches its upper or lower limit). Thus, the distribution varies depending on the daily clearness.

The k_t random variable (unlike a Gaussean one) is bounded: that is, k_t values occur between an upper limit k_{tu} and a lower limit k_{tl}. This feature, coupled with similar patterns in skewness and peakedness, is also observed to be characteristic of the classical Beta probability distribution. Thus, we tested the applicability of this model to describing the observed frequency features in our data.

The k_t variable was first normalised to a random variable (u) within the (0, 1) range, for all values of k_t:

$$u = \frac{k_t - k_{tl}}{k_{tu} - k_{tl}}$$

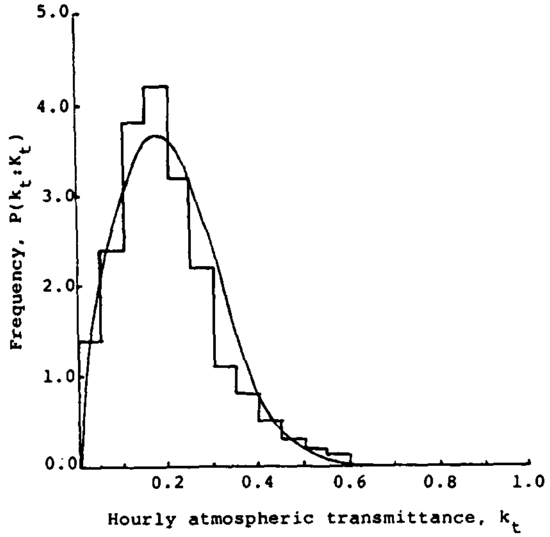

Fig. 3. Observed and predicted frequency distribution of k_t within individual days ($k_{tm} \approx 0.22$).

and the probability density function of u is represented by that of the Beta distribution: $P(u) = cu^{p-1}(1-u)^{q-1}$ where $c = \Gamma(p+q)/(\Gamma(p)\Gamma(q)$ and $\Gamma(\cdot)$ is the gamma function. The distribution of k_t is related to the distribution of u, through the $u(k_t)$ normalizing eqn (7): by equating elemental probabilities we have $P(k_t:K_t)/dk_t/ = P(u:K_t)|du|$, so $P(k_t:K_t) = P(u:K_t)/(k_{tu} - k_{tl})$. The mean u and standard deviation σ_u of the Beta distribution are as follows: $u = p/(p+q)$, $\sigma_u{}^2 = pq/[(p+q)^2(p+q+1)]$: parameters p and q are evaluated by equating the data estimates of the mean and variance of u with these calculated ones. Data estimates of the mean and variance of u are readily expressed in terms of k_{tm} and σ_α:

$$\bar{u} = \frac{k_{tm} - k_{tl}}{k_{tu} - k_{tl}} \quad \text{and} \quad \sigma_t = \frac{\sigma_\alpha}{k_{tu} - k_{tm}} \tag{8}$$

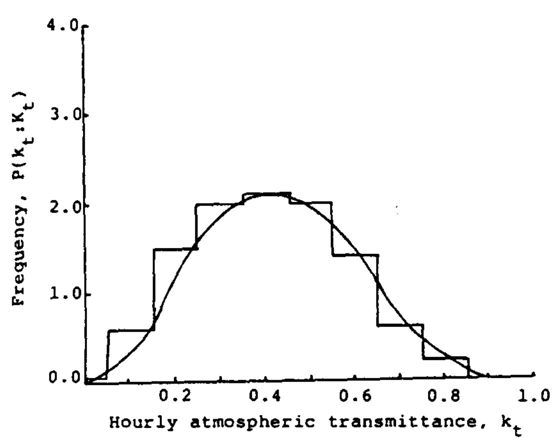

Fig. 4. Observed and predicted frequency distribution of k_t within individual days ($k_{tm} \approx 0.40$).

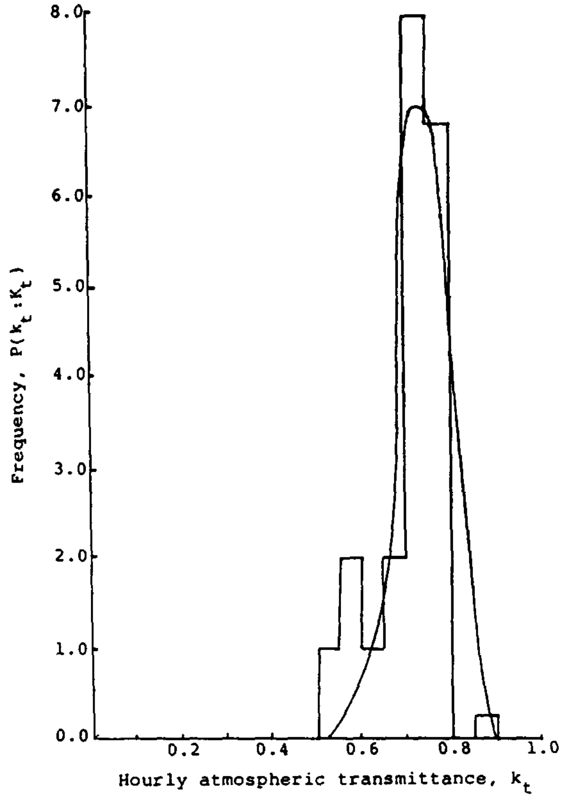

Fig. 5. Observed and predicted frequency distribution of k_t within individual days ($k_{tm} \approx 0.74$).

Also,

$$p = \frac{\bar{u}^2(1-\bar{u})}{\sigma_u^2} - \bar{u} \quad \text{and} \quad q = \frac{p(1-\bar{u})}{\bar{u}}. \tag{9}$$

The probability model for describing the frequency distribution of α is then given by

$$P(\alpha:K_t) = P(k_t:K_t) = \frac{\Gamma(p+q)u^{p-1}(1-u)^{q-1}}{\Gamma(p)\Gamma(q)(k_{tu} - k_{tl})} \tag{10}$$

Figures 3, 4, and 5 show plots of this theoretical probability model for k_t along with experimental density histograms, for specially chosen of k_{tu} and k_{tl}. The superimposed curved generally show good agreement with experimental data. The accuracy of these fits were influenced by the choice of k_{tl} and k_{tu}. Our experimental data show these extreme values to vary, depending on the clearness of the day. For example, for Fig. 3, k_{tl} and k_{tu} were approximately 0.0 and 0.6; for Fig. 4, these values were approximately 0.0 and 0.9; and for Fig. 5, 0.5 and 0.9, respectively. In general, the occurrence of a extreme value is a rare event, the probability of which is extremely small. For example, with a normal distribution, the probability of having an event deviating from the mean by more than three standard deviations is less than 0.12%, and by more than four standard deviations is less than 0.004%. Thus a safe rule would be to use the standard deviation σ_α, as a guide in defining these rare events; we used a deviation of $4\sigma_\alpha$ from the mean. There are, however, circumstances where a deviation of this magnitude is not physically feasible (i.e., it produces a value out of the allowable bounds), and the k_{tl} and/or k_{tu} must be adjusted to its boundary value. With this guideline, suitable values of k_{tl} and k_{tu} are obtained using:

$$k_{tl} = \max(0.0,\ k_m - 4\sigma_\alpha) \quad \text{and}$$
$$k_{tu} = \min(0.9,\ k_m + 4\sigma_\alpha). \quad (11)$$

Because the α values are distributed in a non-Gaussian way, special care must be exercised in using the stochastic model $\alpha(t) = \phi\alpha(t-1) + v(t)$ to generate synthetic data: merely drawing $v(t)$ from a Beta distribution is not sufficient. The mathematics required to manipulate non-Gaussian random variables is much more complex than with Gaussian ones. To solve this problem, the α random variable is mapped into another random variable β (zero mean and unit variance), which is Gaussian. This mapping can be done in the same way in which the non-Gaussian K_t was mapped into a Gaussian variable in the work of generating synthetic K_t[1]. Research by Graham (see Chapter 5 of thesis [12]) has shown that autocorrelation estimates for these Gaussian variables were not different from the same estimates for the non-Gaussian variables. The equation relating α to β is $\beta = \sqrt{2}\ \mathrm{erf}^{-1}[2F(\alpha{:}K_t) - 1]$, where $F(\alpha{:}K_t)$ is the cumulative probability of α, computed by integrating eqn (10). Since β is Gaussian, the serial relationship between hourly events is more easily embedded in its random set:

$$\beta(t) = \phi\beta(t-1) + b(t) \quad (12)$$

where $b(t)$ is drawn from an uncorrelated set of Gaussian random numbers, with mean zero and variance $(1 - \phi^2)$. These autocorrelated β events are then mapped from their domain into the domain of the α distribution, using the mapping equation. Thus, sets of $\alpha(t)$ values are produced with the correct frequency distribution and hourly serial relationship. The value of ϕ is the same as that for α. Thus, to a reasonable approximation $\phi = 0.54$.

5. GENERATING SYNTHETIC HOURLY SOLAR IRRADIATION: A SUMMARY

Synthetic hourly irradiation values (on a horizontal surface) can be produced by generating the hourly atmospheric transmittances and multiplying by the calculated corresponding extra-terrestrial solar irradiation. To generate the k_t values, the hourly trend k_{tm} in k_t is calculated using the K_t for that day and the air mass for the hour in question (eqn (2) and (3)). The standard deviation σ_α of the random component is calculated from eqn (4). Values of β are generated by use of eqn (12) with $\phi = 0.54$. These values are mapped into α values using the distribution of (eqn (10) along with values of p and q obtained from eqns (8), (9), and (11). Then k_t is obtained from eqn (1).

6. CONCLUSIONS

The progression of hourly solar radiation events is a complex, highly nonstationary process with probability mechanism that changes monthly, daily, and even hourly. To construct flexible mathematical models capable of mimicing such probability features of historic records we studied the marginal and serial (stochastic) probability behaviour of sets of hourly atmospheric transmittance events $\{k_t\}$ drawn from days with known daily clearness K_t. Results indicate the marginal probability and stochastic features of $\{k_t\}$ sets vary in a manner which depends predominantly on the overall clearness of the day from which they were drawn.

Models were developed for predicting the trends in a daily set $\{k_t\}$ and for estimating the likely magnitudes of random fluctuations in k_t about the trends. Results reveal that extremely good predictions were possible from knowledge of the associated K_t. The parameters in these models were also observed could be fixed at values which were independent for the climates studied, without any loss in their accuracy. An investigation to determine if the probability features of the random component k_t depend on the air mass, m was also performed, and only a modest change in the variance with air mass was observed. The probability density of the random fluctuations in k_t was found to be non-Gaussian, but fitted well by a Beta distribution. The characteristics of this distribution in terms of parameters dependent of K_t and air mass, have been identified. Finally, methods for generating synthetic values of k_t, given K_t and m, have been described.

Acknowledgments—The authors express their sincere appreciation to the WATSUN User Service for the use of the WATSUN solar simulation programs, and to Environment Canada for making their records of irradiation data available to us for use in the research. The financial support of the National Science and Engineering Research Council Canada through a strategic grant is also gratefully acknowledged.

REFERENCES

1. 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).
2. G. E. P. Box and G. M. Jenkins, *Time series analysis: Forecasting and control*, Holden-Day, San Francisco (1970).
3. R. J. Aguiar, M. Collares-Pereira, 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).
4. T. N. Goh and K. J. Tan, Stochastic modelling and forecasting of radiation data, *Solar Energy* **22**, 105–111 (1977).
5. A. H. C. van Paassen, *Indoor climate, outdoor climate*

and energy consumption, Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands (1981).
6. D. C. Hittle and C. O. Pedersen, Periodic and stochastic behaviour of weather data, *ASHRAE Transactions*, Vol. 87, part 2 (1981).
7. A Balouktsis and P. Tsaides, Stochastic simulation model of hourly total solar radiation, *Solar Energy* 37, 119-126 (1986).
8. B. Y. H. Liu and R. C. Jordan, The interrelationship and characteristic distributions of of direct, diffuse, and total solar radiation, *Solar Energy* 4, 1-19 (1960).
9. L. O. Dagleman, A weather simulation model for building energy analysis, ASHRAE Symposium on Weather Data, Seattle, WA, Annual Meeting, June 1976, ASHRAE Transactions, pp. 435-447 (1976).
10. R. H. B. Excell, A mathematical model for solar radiation in south-east Asia (Thailand), *Solar Energy* 26, 161-168 (1981).
11. C. Mustacchi, V. Cena, and M. Rocchi, Stochastic simulation of hourly global radiation sequences, *Solar Energy* 23, 47-51 (1979).
12. V. A. Graham, *Stochastic synthesis of the solar atmospheric transmittance*, Ph.D. Thesis, University of Waterloo, Waterloo, Ontario, Canada (1985).
13. M. B. Fiering, Multivariate techniques for synthetic hydrology, ASCE Journal of Hydraulic Div., HY5, 43-60 (1964).
14. V. D. Valencia and J. C. Schaake, Disaggregation processes in stochastic hydrology, *Water Resources Research* 9, 580-585 (1964).
15. R. King and R. O. Buckius, Direct solar transmittance for a clear sky, *Solar Energy* 22, 297-301 (1979).
16. H. C. Hottel, A simple model for estimating the transmittance of direct solar radiation through clear atmospheres, *Solar Energy* 18, 129-134 (1976).
17. R. O. Buckius and R. King, Diffuse solar radiation on a horizontal surface for a clear sky, *Solar Energy* 21, 503 (1978).
18. K. G. T. Hollands (1984), A derivation of the diffuse fraction's dependence on the clearness index, Proc. of Solar World Congress, 8th Biennial Congress of ISES, Perth, Australia, Pergamos Press, New York. pp. 2088-2092 (1983).
19. J. F. Orgill and K. G. T. Hollands, Correlation equation for hourly diffuse radiation on a horizontal surface, *Solar Energy* 19, 357-359 (1977).
20. M. Iqbal, Prediction of hourly diffuse solar radiation from measured hourly global radiation on a horizontal surface, *Solar Energy* 24, 491-503 (1980).
21. A. Athienitis, Personal communication, University of Waterloo. Waterloo, Ontario, Canada (1985).
22. 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, 155-164 (1979).
23. K. G. T. Hollands and R. G. Huget, A probability density function for the clearness index with applications, *Solar Energy* 30, 195-209 (1983).
24. P. Bendt, M. Collares-Periera, and A. Rabl, The frequency distribution of daily insolation values, *Solar Energy* 27, 1-5 (1981).