CHARACTERIZATION AND SIMULATION OF HOURLY EXPOSURE SERIES OF GLOBAL RADIATION

LL. MORA-LÓPEZ * and M. SIDRACH-DE-CARDONA **†

*Dpto. Lenguajes y C. Computación, E.T.S.I. Informática,
Complejo Tecnológico de la Información Campus Teatinos, Universidad de Málaga,
29071 Málaga, Spain and ** Dpto. Física Aplicada II, E.T.S.I. Informática,
Complejo Tecnológico de la Información Campus Teatinos, Universidad de Málaga,
29071 Málaga, Spain

(Received 22 December 1995; revised version accepted 16 January 1997)

(Communicated by AMOS ZEMEL)

Abstract—A statistical model which captures the main features of hourly exposure series of global radiation is proposed. This model is used to obtain a procedure to generate radiation series without imposing, a priori, any restriction on the form of the probability distribution function of the series. The statistical model was taken from the stationary stochastic processes theory. Data were obtained from ten different Spanish locations. As monthly hourly exposure series of global radiation are not stationary, they are modified in order to remove the observed trends. A multiplicative autoregressive moving average model with regular and seasonal components was used. It is statistically accepted that this is the true model which generates most of the analyzed sequences. However, the underlying parameters of the model vary from one location to another and from one month to another. Therefore, it is necessary to examine further the relationship between the parameters of the model and the available data from most locations.

© 1997 Elsevier Science Ltd.

1. INTRODUCTION

When solar energy systems (sizing, design, evaluation, optimization) with memory (storage capacity) from one to a few hours are studied, it is necessary to take into account the random nature of solar radiation. In general, measured series of radiation are only available for a small number of locations. Moreover, the recorded values usually correspond to daily intervals, whereas for this kind of systems it would be necessary to have hourly values. Practical examples of such systems are some specific applications of photovoltaic solar energy (grid connected systems, hybrid systems, pumping systems, solar electricity systems with buffer storage), certain solar absorption air conditioning systems, certain control problems in solar building, and, in general, systems with high threshold. In this sense, values of the utilizability function computed with hourly data, are always greater than those computed with just daily radiation.

A possible way to study the stationary and sequential characteristics in series is by means of time series statistical models. The statistical analysis of time series is common practice in areas such as economics, hydrology or meteorology. Basically, its purpose is to analyze past data in order to infer statistical properties which capture the observed empirical regularities of the series and, then, use these properties to forecast its future behavior. The theory of stationary stochastic processes provides a general methodology to study infinite series of random variables when the correlation they exhibit depends on its time proximity.

In recent years this methodology has been used in a solar radiation series context by many authors, e.g. Brinkworth (1977), Bartoli et al. (1981), Bendt et al. (1981), Aguiar et al. (1988), Athienitis (1990) and Aguiar and Collares-Pereira (1992b).

As it is well-known, hourly exposure series of global radiation are not stationary. They exhibit two classes of trends, a daily one and a seasonal one. Therefore, all articles which study these series first propose a method to capture and remove the observed trends. Balouktsis and Tsalides (1986) propose to use average values (or moving averages) obtained from these series with Fourier functions. The new series are not normal and the authors propose procedures to transform them into normal series. Boch et al. (1981), Amato et al. (1988) and Aguiar and Collares-Pereira (1992a) use values of extraterrestrial hourly radiation. Boch et al. (1981) modify the new series of hourly clearness index
using average hourly values and the variances of these series. Guinea et al. (1988) and Palomo (1989) use the enveloping or maximum value among recorded values.

The characteristics of the series obtained after removing the different trends depend on the method which is used. It is possible to specify for these series a theoretical statistical model studying its statistical properties, such as mean value, variance, probability distribution function (pdf), autocorrelation function (acf) and partial autocorrelation function (pacf).

Palomo (1989) proposes to use first-order Markov models for hourly series, though, as she says in her article, these models are not universal. These models were previously proposed by Aguiar et al. (1988) to analyze and generate daily series of the clearness index. In this case the resultant model is universal.

A different approach to study hourly exposure series of global radiation is adopted by Graham and Hollands (1990), who use stochastic models for generating synthetic sets of hourly irradiation values, using daily values as input (these daily values are also generated using mathematical models of the irradiation process) and by Aguiar and Collares-Pereira (1992b), who propose a time-dependent autoregressive Gaussian model (TAG). The main disadvantage of these models is that it is necessary to generate daily series first, and then generate hourly values.

For the time being, this methodology of a "cascade of models" is the most widely accepted and the one which yields best results. However, in this article the possibility of using a different methodology is introduced: each monthly series of values of hourly exposure of global radiation is directly analyzed and characterized and, then, new series with the same characteristic parameters are generated. This analysis was performed in several steps. First, an expression which makes it possible to remove the observed trends in the original series is used; this expression is only a function of the solar height at each hour. Then, the difference operator is used to modify the original series. Afterwards, an underlying statistical model for the modified series is proposed and its parameters are estimated for each month. Then, the validity of assumptions and results is analyzed from a statistical point of view. A procedure to generate synthetic series is proposed. Finally, the applicability of the model is studied and the unsolved problems are discussed—basically, how to apply this model in locations with no past data, that is to say, how to find the relationship between the parameters of the underlying statistical model and available data, such as the monthly average of the daily clearness index.

2. DATA SET

The data of hourly exposure series of global radiation, \( \{G_h(t)\} \), used in this work were recorded over several years in various Spanish meteorological stations. These data are contained in the Data Base of Institute of Renewable Energy (CIEMAT). In Table 1 the number of available observations is reported, specifying year and latitude of the place where they were recorded (North latitude). As some months in each year are not available, the total number of months for which observations are available in each location appears in the last column. This table also includes annual average values of daily global solar radiation, \( G_{day} \).

3. STATISTICAL ANALYSIS OF HOURLY EXPOSURE SERIES OF GLOBAL RADIATION

The hourly exposure series of global radiation have been constructed in an "artificial" way because data from different days were linked together: the last observation of each day is followed by the first observation of the following day. The initial assumption of this work is that these series are homogeneous. The statistical results will allow us to accept or reject this hypothesis.

In hourly exposure series of global radiation one observes the influence of two different kinds of factors: deterministic ones and random ones. In these series two main characteristics are observed: on the one hand, a trend which is caused by the influence of variation in the relative sun-earth position for different days of year (this is a seasonal trend); on the other hand, a daily trend which is caused by the day-night sequence, that is to say, in a fixed day the values of radiation display a trend caused by the variation in the extraterrestrial radiation. These characteristics render hourly exposure series of global radiation non-stationary (neither in mean nor in variance), a condition which must be met if the theory of stationary stochastic processes is to be used.

The mathematical function proposed to characterize the seasonal component of these series
Characterization and simulation of hourly exposure series of global radiation

Table 1. Data set

<table>
<thead>
<tr>
<th>Location</th>
<th>Years</th>
<th>Months</th>
<th>Latitude (°N)</th>
<th>( G_{\text{h}} ) (MJ m(^{-2}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Badajoz</td>
<td>1976-1983</td>
<td>84</td>
<td>38.89</td>
<td>16.6</td>
</tr>
<tr>
<td>Castellón</td>
<td>1979-1984</td>
<td>55</td>
<td>39.95</td>
<td>15.8</td>
</tr>
<tr>
<td>Logroño</td>
<td>1981-1984</td>
<td>48</td>
<td>42.46</td>
<td>15.1</td>
</tr>
<tr>
<td>Madrid</td>
<td>1979-1986</td>
<td>96</td>
<td>40.45</td>
<td>16.6</td>
</tr>
<tr>
<td>Málaga</td>
<td>1977-1984</td>
<td>94</td>
<td>36.66</td>
<td>16.9</td>
</tr>
<tr>
<td>Murcia</td>
<td>1977-1984</td>
<td>96</td>
<td>38.00</td>
<td>16.9</td>
</tr>
<tr>
<td>Oviedo</td>
<td>1977-1984</td>
<td>96</td>
<td>43.35</td>
<td>11.1</td>
</tr>
<tr>
<td>Mallorca</td>
<td>1977-1984</td>
<td>94</td>
<td>39.33</td>
<td>15.5</td>
</tr>
<tr>
<td>Sevilla</td>
<td>1977-1984</td>
<td>78</td>
<td>37.42</td>
<td>17.6</td>
</tr>
<tr>
<td>Tortosa</td>
<td>1980-1984</td>
<td>52</td>
<td>40.81</td>
<td>15.1</td>
</tr>
</tbody>
</table>

is an expression for the maximum hourly exposure of global radiation, \( G_{\text{h, max}}(t) \), similar to that proposed by Brichambaut et al. (WMO, 1981). This is a clear-sky model estimated in Mora (1995), where the coefficients in the expression are computed using hourly exposure of global radiation data recorded in various Spanish locations. In this expression, the angle of solar height, \( \alpha \), depends on time. Thus, the function, \( f_h(t) \), which is used in order to remove trends is:

\[
f_h(t) = G_{\text{h, max}}(t) = 3.96 \sin(\alpha)^{1.05} \text{ (MJ m}^{-2}\text{)}
\]

(1)

This maximum radiation shows the influence of both the variation of sun-earth relative position and the variation of the air mass which radiation has passed through. The new series, \( \{X_h(t)\} \), exhibit variations caused by the diversity of atmospheric components (dust, aerosols, presence of clouds, ...). These series are:

\[
X_h(t) = G_h(t)/G_{\text{h, max}}(t)
\]

(2)

In order to use series which are stationary in mean, variance and covariance, monthly series must be considered. If periods of time which cover more than one month (seasons within a year, years, ...) are used, then stationarity properties no longer hold. For example, sample means and variances of the series \( \{X_h(t)\} \) recorded at Madrid in each month of the year 1982 are reported in Table 2. Here it is observed that it seems unreasonable to consider that the mean or the variance is the same for all these series. Moreover, it was checked that it is not possible either to use "global months" (series obtained simply by joining observations of the same month for different years), as Gordon and Reddy (1988) suggest with daily data; in most cases there are some specific months whose characteristics (mean, variance) differ consider-

ably from those of all other series corresponding to the same month, but different year. For example, sample means and variances of the series \( \{X_h(t)\} \) recorded at Madrid in January of eight different years are reported in Table 3. It is also observed here that it seems unreasonable to consider that the mean or the variance is the same for all these series. Graham and Hollands (1990), among others, have already come to the same conclusion (it is not possible to use global months).

When using monthly series, it is observed that the assumption of stationarity is still not acceptable for these series; see, for example, Fig. 1, where the series \( \{X_h(t)\} \), for July, 1978 in P. Mallorca is depicted. In this figure (which is similar in all series which were analysed), two main features are observed: the first one is that there is still a strong daily trend; the second one is that the first recorded observation in each day and the last one are clearly lower than all others. The problems which this first feature might cause were circumvented using the difference operator which makes it possible to remove the daily trends. The second feature appears because the first observation in each day does not correspond to an hourly value but to the value from sunrise to the next hour (sometimes just 5 min); similarly, the last observation in each day corresponds to the value from the previous hour to sunset (it is not an hourly value, either). Moreover, the recorded values at these intervals are quite low. In almost all locations and months, each day the ratio between energy received in first and last hour of the day, and total energy received in the day is less than 0.006. Moreover, the energy which is received at those hours (in all cases less than 50 kJ m\(^{-2}\)) is not exploited in most solar systems because of losses, functioning threshold, etc. Therefore, these intervals are not going to be considered.
Table 2. Sample mean and variance for the series \( \{X_h(t)\} = \{G_{h(t)}/G_{h_{\text{max}}(t)}\} \). Data: Year 1982, Madrid

<table>
<thead>
<tr>
<th>Year</th>
<th>January</th>
<th>February</th>
<th>March</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.559</td>
<td>0.609</td>
<td>0.757</td>
<td>0.709</td>
<td>0.694</td>
<td>0.690</td>
<td>0.737</td>
<td>0.710</td>
<td>0.682</td>
<td>0.579</td>
<td>0.563</td>
<td>0.518</td>
</tr>
<tr>
<td>Variance</td>
<td>0.292</td>
<td>0.282</td>
<td>0.272</td>
<td>0.254</td>
<td>0.232</td>
<td>0.217</td>
<td>0.219</td>
<td>0.210</td>
<td>0.233</td>
<td>0.293</td>
<td>0.239</td>
<td>0.302</td>
</tr>
</tbody>
</table>

Table 3. Sample mean and variance for the series \( \{X_h(t)\} = \{G_{h(t)}/G_{h_{\text{max}}(t)}\} \). Data: Month: January, from year 1979 to 1986, Madrid

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.482</td>
<td>0.559</td>
<td>0.757</td>
<td>0.559</td>
<td>0.689</td>
<td>0.578</td>
<td>0.561</td>
<td>0.579</td>
</tr>
<tr>
<td>Variance</td>
<td>0.305</td>
<td>0.273</td>
<td>0.242</td>
<td>0.293</td>
<td>0.185</td>
<td>0.276</td>
<td>0.297</td>
<td>0.265</td>
</tr>
</tbody>
</table>

As a consequence of not using those intervals, the number of hours considered for each month is constant and equal for all locations considered. The number of hours considered for each month are: 10 for January, February, November and December; 12 for March, April, September and October; 14 for May, June, July and August.

As mentioned above, the difference operator was used to remove the daily trends. In Box and Jenkins (1976) some examples of the use of this operator can be found. The order of this operator will coincide, for each month, with the number of recorded hours in each day (this number is denoted as \( s \)). This operator has already been used for daily series by Boileau (1983), (although his proposed daily model never yielded good results).

The differenced series are thus obtained according to:

\[
Y(t) = D_s(X_h(t)) = X_h(t) - X_h(t-s)
\]  

Using this operator in each monthly series the non-stationarity of the series is removed. This fact can be observed in Fig. 2, which corresponds to a differenced monthly series (same month and location as in Fig. 1).

Observe that the model which is proposed is not in contradiction with the hypothesis that \( X_h(t) \) can be decomposed as:

\[
X_h(t) = C_h(t) + W_h(t)
\]  

where \( C_h(t) \) is a deterministic seasonal trend—satisfies \( C_h(t) = C_h(t-s) \)—and \( W_h(t) \) a stationary time series.

The distribution functions of these series may be identified with Gaussian processes (or processes which asymptotically behave as Gaussian ones). This fact is common to all differenced series, as can be seen, for instance, in Fig. 3, where the probability distribution function, pdf, of a differenced series is shown (Month: July, 1978; location: P. Mallorca).
4. IDENTIFICATION OF THE MODEL

In order to identify the underlying model the sample autocorrelation function, acf, and partial autocorrelation function, pacf, are used—the expression for these functions are described in Box and Jenkins (1976), pp. 27–28 and pp. 64–65, respectively. These sample acf and pacf have the same shape in all the studied months. Figures 4 and 5 show the sample acf and pacf of the differenced series with data from Malaga, July 1980.

In the sample acf, it is systematically observed that there are several values significantly different from zero (sometimes even the fifth order coefficient is different from zero). The size of these correlations decreases rapidly and is virtually zero for almost all lags of order greater than five. Moreover, there is always a negative value, significantly different from zero, for the lag which coincides with the number of hours of the corresponding day; and the values for lags \( s - 1 \) and \( s + 1 \) are also different from zero though its absolute value is usually lower than that obtained for lag \( s \).

As regards the sample pacf of the analysed series, it is observed that for small lags there is only one value which is significantly different from zero (lag 1). Moreover, in the neighborhood of \( s \) (number of recorded hours in the day) there appear various values which are also significantly different from zero. The value of the coefficient of lag \( s \) in the sample pacf is negative. The preceding value is also negative and the subsequent one is positive and both are also significantly different from zero, though their absolute values are, in both cases, lower than that obtained for lag \( s \). And these characteristics also appear in lags \( ks - 1 \), \( ks \), \( ks + 1 \), when \( k \) is a positive integer close to 1; however, as \( k \) grows the values obtained for these lags decrease (in absolute value).

Following the usual methodology in the analysis of stationary time series, see e.g. Box and Jenkins (1976), a seasonal multiplicative ARMA model is proposed, that is to say, the regular and seasonal parts of the series are analyzed separately. The regular part captures the relation between consecutive observations. The seasonal part captures the relation between observations from the same season (in this case,
The properties of the regular part in the series are typical of a first-order autoregressive model $\text{ARMA}(1,0)$ and the properties of the seasonal part in the series are typical of a first-order moving-average model $\text{ARMA}(0,1)$ (see Box and Jenkins (1976), for a detailed description of $\text{ARMA}(1,0)$ and $\text{ARMA}(0,1)$ time series).

These characteristics make us propose the use of a multiplicative $\text{ARMA}$ model with a first-order autoregressive process in the regular part and a first-order moving average process in the seasonal part, that is:

$$Y(t) = \phi_1 Y(t-1) + \epsilon(t) - \theta_1 \epsilon(t-s) \quad (5)$$

where $\{Y(t)\}$ is the analyzed series and $\{\epsilon(t)\}$ is a white noise error term.

This model is usually written as $\text{ARMA}(1,0) \times (0,1)_s$. A multiplicative model is used because it explains the fact that the values for lags $s-1$ and $s+1$ are significantly different from zero (see Mora (1995), pp. 200–202 and Box and Jenkins (1976), pp. 329).

The introduction of a seasonal component makes this model different from first-order Markov models as those proposed by Palomo (1989) or Guinea et al. (1988). In these articles no seasonal part is mentioned. The reason why this happens may be twofold: on the one hand, because sample acf and pacf are only computed for the first lags; on the other hand, because they do not use, for each month, days with the same number of recorded hours.

According to this model, the value of the series in an hour, $Y(t)$, is a linear combination between the value in the previous hour, $Y(t-1)$, and an error term, $\epsilon(t)$. This error term is not a white noise, but can be expressed as:

$$\epsilon(t) = \epsilon(t) - \theta_1 \epsilon(t-s) \quad (6)$$

where $\{\epsilon(t)\}$ is a white noise. Thus, the error term, $\epsilon(t)$, in an hour is highly correlated with the error term which took place the previous day at the same hour, $\epsilon(t-s)$. These properties could be related with the fact that changes with respect to the maximum levels of radiation which take place at different hours in a day are not random.

In contrast, the results allow us to confirm that the initial assumption (the "artificial" constructed series are homogeneous) can be accepted.

5. ESTIMATION OF PARAMETERS

For the series $\{Y(t)\}$, an $\text{AR}(1)$ model for the regular part and a $\text{MA}(1)$ model for the seasonal part are proposed. This seasonal multiplicative $\text{ARMA}(1,0) \times (0,1)_s$ model may be written as in Fig. 5.

Therefore, parameters $\phi_1$, $\theta_1$, and the variance of the white noise process, $\sigma^2$, must be estimated for each series. These parameters are estimated using maximum likelihood and assuming a Gaussian distribution with zero mean for the white noise process. This method of estimation requires optimization and selects, among all possible values of the parameters, those which make more likely the data set used for the estimation.

The estimated parameters for each month and location have been compared in order to analyze whether it is possible to use the same parameters for all series. The results show that the proposed model is universal but the parameters of this model vary from one series to another.

6. DIAGNOSIS OF THE MODEL

A number of tests were performed in order to validate the model. Specifically, various tests were carried out in order to:

1. Analyze whether parameters $\phi_1$, $\theta_1$, are significantly different from zero;
2. Analyze whether it is admissible to assume that $\{\epsilon(t)\}$ is a Gaussian process;
3. Analyze whether the hypothesis that $\{\epsilon(t)\}$ in (5) is a white noise is admissible.

Specifically, $t$ statistics associated with $\phi_1$ and $\theta_1$ were computed using standard maximum likelihood techniques (part a); the Bera–Jarque statistic based on asymmetry and skewness coefficients (part b) has also been computed (see Spanos (1986), pp. 453–454 for a description of this statistic). For (c), series $\{\epsilon(t)\}$ cannot be used because it is not observable. From the estimation described in previous section, it is possible to obtain series $\{\epsilon^*(t)\}$ (residual series) which behaves similarly to $\{\epsilon(t)\}$ (see Box and Jenkins, 1976, pp. 289 for a description of how this series is obtained).

The hypothesis that $\{\epsilon(t)\}$ is a white noise was tested using the portmanteau test, computed with series $\{\epsilon^*(t)\}$ and 20 lags (see Box and Jenkins, 1976, pp. 290–291 for a description of this test).

In almost all series both coefficients $\phi_1$ and
Characterization and simulation of hourly exposure series of global radiation

6. GENERATION OF HOURLY EXPOSURE SERIES

In this section it is described how the proposed model can be used to generate hourly exposure series of global radiation. The procedure consists of the following steps.

7. GENERATION OF HOURLY EXPOSURE SERIES

In this section it is described how the proposed model can be used to generate hourly exposure series of global radiation. The procedure consists of the following steps.
Fig. 10. Frequency Histogram for the original series \( \{ Y(t) \} \) (□) and generated series \( \{ Y'(t) \} \) (■). Data: February 1982, Murcia. Parameters of the ARMA model used to obtain \( \{ Y'(t) \} \): \( \phi_1 = 0.790, \theta_1 = 0.904, \sigma^2 = 0.0338 \).

Fig. 11. Frequency Histogram for the original series \( \{ Y(t) \} \) (□) and generated series \( \{ Y'(t) \} \) (■). Data: November 1984, Oviedo. Parameters of the ARMA model used to obtain \( \{ Y'(t) \} \): \( \phi_1 = 0.534, \theta_1 = 0.877, \sigma^2 = 0.0576 \).

1) Generate a stationary ARMA(1,0)×(0,1), time series, \( \{ Y(t) \} \), using a Gaussian white noise \( \{ \varepsilon(t) \} \) (from \( t = -s \) to \( t = T \)) and the relation:

\[
Y(t) = \phi_1 Y(t-1) - \theta_1 \varepsilon(t-s) + \varepsilon(t) \quad t = 1 \ldots T
\]

where \( T \) is the number of hours for each month.

In order to obtain \( \{ Y(t) \} \) it is, thus, necessary to use a value of \( \sigma^2 \) for the variance of the white noise and values \( \phi_1, \theta_1 \) for the autoregressive and moving-average parameters. For each month these unknown values were replaced by
Figure 12. Frequency Histogram for the original series \( \{Y(t)\} \) (●) and generated series \( \{Y'(t)\} \) (■). Data: November 1983, Murcia. Parameters of the ARMA model used to obtain \( \{Y'(t)\} \): \( \phi_1 = 0.832, \theta_4 = 0.918, \sigma^2 = 0.0269 \).

their estimates. Thus, they are different in each generated series.

Observe that in order to use eqn (7) to obtain \( \{Y'(t)\} \) (from \( t = 1 \) to \( t = T \)) it is also necessary to use an initial value \( Y(0) \). It was checked that the resultant series does not virtually depend on this initial value. Specifically, series obtained with different initial points only differ in the first few initial values.

All generated series \( \{Y'(t)\} \) obtained in this way are similar to the original series, \( \{Y(t)\} \). For each month and location, the mean and the variance of \( \{Y(t)\} \) and \( \{Y'(t)\} \) were compared using statistical tests. In most cases (93%) it is statistically accepted (level of significance 0.05) that they are the same. Moreover, the pdf, acf and pacf of both series show similar behaviour. In Figs 6–12 the acf, pacf and pdf of both series are depicted for various month and locations.

2. Now, for each month and location, a series \( \{X'_h(t)\} \), which is similar to \( \{X_h(t)\} \), is obtained using \( \{Y'(t)\} \). This series is obtained using the expression:

\[
X'_h(t) = Y'(t) + X'_h(t-s) \quad t = 1, \ldots, T \quad (8)
\]

Obviously, in order to obtain \( \{X'_h(t)\} \) (from \( t = 1 \) to \( t = T \)) using eqn (8), it is necessary to use \( s \) initial values \( \{s_j, j = -(s-1), \ldots, 0\} \). If no restriction on the \( s \) initial values is imposed, then the series \( \{X'_h(t)\} \) may take on values which are less than zero or greater than one, in some cases. The original series, \( \{X_h(t)\} \), are normalized values of radiation (to the maximum radiation) and, therefore, \( X_h(t) \) is never less than zero. A possible way to circumvent the problem of the appearance of negative values in \( \{X'_h(t)\} \)—a feature which was already observed by Boileau (1983)—is to use an iterative process in order to modify the \( s \) initial values. Moreover, the new series must have the same mean value as the original series. Taking into account these conditions, in Appendix A an algorithm to calculate the initial \( s \) values and the series \( \{X'_h(t)\} \) is explained.

The obtained series \( \{X'_h(t)\} \) is not stationary (if the difference operator is applied to it then the resultant series will be stationary). Moreover, it has the same mean value as the original series and the estimate of its variance is similar to the variance of the ARMA model whose parameters are those which were used to generate the series.

The distribution function of this series is not normal (the distribution function of the original series was not normal either); in fact this distribution corresponds to a mixture of normal distributions.

In order to compare original and generated series—\( \{X_h(t)\} \) and \( \{X'_h(t)\} \)—it was tested whether it is possible to accept that the underlying pdf of original and generated series is the same. Using the Kolmogorov–Smirnov two-
sample test, the null hypothesis of equal pdf is accepted in 90% of cases (see Shorack and Wellner (1986), pp. 401–403 for a description of this test). Thus, these results show that original and generated series have similar characteristics. For example, in Figs 13–15 the cumulative probability distribution function, cpdf, for real and simulated series \( \{X_h(t)\} \) and \( \{X'_h(t)\} \), for different months and locations, are depicted. In Figs 16–19 real and simulated series, \( \{X_h(t)\} \) and \( \{X'_h(t)\} \), are depicted for various month and locations.

(3) Finally, using the series \( \{X'_h(t)\} \), the hourly exposure series of global radiation can be obtained, taking into account that:

\[
G_h(t) = G_{h,\text{max}}(t) X'_h(t) \tag{9}
\]

To analyze whether the recorded series, \( \{G_h(t)\} \), and the generated series, \( \{G'_h(t)\} \), are similar, statistical tests were used to compare their means, variances and probability distribution functions. The results show that, in most cases (approximately 90%), it is possible to
Characterization and simulation of hourly exposure series of global radiation

1.2
1.0
0.8
0.6
0.4
0.2
0

Fig. 17. Series $\{X'_h(t)\}$, obtained from the series $\{Y'(t)\}$. Parameters of the ARMA model used to obtain $\{Y'(t)\}$: $\phi_1 = 0.723$, $\theta_1 = 0.791$, $\sigma^2 = 0.036$, derived from data: May 1980, Madrid.

Fig. 18. Series $\{X_h(t) = G_h(t)/G_{h,max}(t)\}$, obtained from recorded data. Data: April 1983, Oviedo.

Fig. 19. Series $\{X'_h(t)\}$, obtained from the series $\{Y'(t)\}$. Parameters of the ARMA model used to obtain $\{Y'(t)\}$: $\phi_1 = 0.746$, $\theta_1 = 0.852$, $\sigma^2 = 0.0297$, derived from data: April 1983, Oviedo.

Fig. 20. Cumulative probability distribution function for the recorded series $\{G_h(t)\}$ (---) and generated series $\{G'_h(t)\}$ (---). Data: May 1980, Madrid. (MJ m$^{-2}$).

accept that recorded and generated series are equal. For example in Figs 20–22 some pdf for these two series are presented. Also, in Figs 23–26 the recorded and generated series for various months and locations are shown.

8. CONCLUSIONS

The objective of this work was to find a statistical model which enables the generation of hourly exposure series of global radiation, with properties similar to those of recorded series. Data of hourly exposure series of global radiation recorded in 10 different Spanish locations, during 735 months, were used for that purpose.

The theory of stationary stochastic processes was used. In order to obtain stationary series the observed daily and seasonal trends were removed from the original series. An expression for the maximum hourly global radiation was proposed in order to characterize the trends of observed series. Then, the difference operator was used to remove the persistent daily trend; the use of this difference operator makes it possible to obtain series whose distribution function is normal. The properties of the series

$\phi_1 = 0.723$, $\theta_1 = 0.791$, $\sigma^2 = 0.036$
obtained in this way were analyzed using the simple and partial autocorrelation functions. These functions suggest that these series can be identified with a seasonal multiplicative ARMA model whose regular part follows a first-order autoregressive model and whose seasonal part follows a first-order moving average model (i.e. an ARMA\((1,0) \times (0,1)_s\) model, where \(s\) is the number of sunshine hours per day corresponding to each month). According to this model the value for each hour is a linear combination of the value in the previous hour and an error term. This error term is highly correlated with the error term corresponding to the same hour at the previous day. This approach solves the problem which arises from the use of ARMA processes to generate and model hourly exposure series of global radiation: the non-normality of the analyzed series, as pointed out by different authors; for example, Aguiar and Collares-Pereira (1992b), or Amato et al. (1988).

The unknown parameters of the proposed ARMA model (regular AR coefficient, seasonal MA coefficient and error variance) were estimated. Then, these estimated parameters were compared in order to analyze whether it is possible to use the same parameters for all series. The results show that the proposed model
Characterization and simulation of hourly exposure series of global radiation

4000
3000
2000
1000
0
0 100 200 300 400 500
hour

Fig. 25. Recorded series of hourly exposure of global radiation \( (G_h(t)) \). Data: July 1983, Oviedo (MJ m\(^{-2}\)).

4000
3000
2000
1000
0
0 100 200 300 400 500
hour

Fig. 26. Generated Series of hourly exposure of global radiation \( (G_h(t)) \). From data: July 1983, Oviedo (MJ m\(^{-2}\)).

is universal but the parameters of this model vary from one series to another.

Finally a method to generate monthly series which, \textit{a priori}, imposes no restriction on the form of the probability distribution function of radiation values was proposed. With this method, it is now possible to solve a problem which has already been pointed out by Boileau (1983): the appearance of negative values in the generated series.

The most important conclusion of this work is that it is possible to use ARMA models to simulate hourly global radiation series. It remains to solve the problem of the non universality of the parameters. The authors of this work are currently studying the relationship between these parameters and other data which are commonly available, such as the monthly average clearness index of daily radiation.

REFERENCES


APPENDIX A

Algorithm to calculate the series \{X_i(t)\}

First, observe that, as the variable \(t\) is a function of the hour and day, the series \{Y'(t)\} can be written as:

\[
Y'(t) = Y'(h,d), \quad h=1,...,s, \quad d=1,...,N \tag{A1}
\]

where \(s\) is the number of hours for the day and \(N\) is the number of days for each month. Accordingly, the series \{X_i(t)\} can be written as:

\[
X'(h,d) = Y'(h,d) + X'(h,d-1), \quad h=1,...,s, \quad d=1,...,N \tag{A2}
\]

But \(s\) initial values \(X'(h,0)\) \((h=1,...,s)\) are required if (A.3) is to be used. These \(s\) initial values must be chosen in such a way that all values in \{X_i(t)\} are greater than zero and less than one. In order to be sure that this condition is met, these \(s\) initial values are obtained with an iterative process which is now described.

First, a series \{Z_0(h,d)\} is obtained as follows:

\[
Z_0(h,0) = 0, \quad h=1,...,s \tag{A4}
\]

\[
Z_0(h,d) = Y'(h,d) + Z_0(h,d-1), \quad h=1,...,s, \quad d=1,...,N \tag{A5}
\]

Then, the minimum and maximum values which appear in this series for each hour are searched:

\[
m(h) = \min\{Z_0(h,j),\quad j=0,...,N\} \tag{A6}
\]

\[
M(h) = \max\{Z_0(h,j),\quad j=0,...,N\} \tag{A7}
\]

Then it is checked whether:

\[
ABSM(m(h)) < 1 - M(h) \tag{A8}
\]

If (A.8) is not satisfied at an hour, then the second minimum is set as \(m(h)\) and (A.8) is checked again. The process continues until values \(m(h)\) and \(M(h)\) which satisfy (A.8) are found for each hour. Usually, the minimum and maximum values or second minimum, second maximum values satisfy (A.8).

In order to reproduce the trend observed in the series \{X_i(t)\}, a new series \{Z_1(h,d)\} is then obtained as follows:

\[
Z_1(h,0) = 1 - M(h) \tag{A9a}
\]

\[
Z_1(h,0) = ABS(m(h)) + \frac{(1 - M(h) - ABS(m(h)))}{2} \tag{A9b}
\]

\[
Z_1(h,0) = ABS(m(h)) \tag{A9c}
\]

\[
Z_1(h,d) = Y'(h,d) + Z_1(h,d-1), \quad h=1,...,s, \quad d=1,...,N \tag{A10}
\]

where the expression:

(1) (A.9.a) is used for the 4 h around noon: from \((s/2-2)\) to \((s/2+2)\)

(2) (A.9.b) is used for the hours: from \((s/2-4)\) to \((s/2-2)\)

(3) (A.9.c) is used for the remaining hours.

For this series, its mean value is computed and compared with the observed mean value. Specifically, it is checked if:

\[
\bar{Z}_1 \in (0.98\bar{X}, 1.02\bar{X}) \tag{A11}
\]

where:

\[
\bar{Z}_1 = \{ \sum_{d=1}^{N} Z_1(h,d) \}/(Ns)
\]

and \(\bar{X}\) is similarly obtained.

If (A.11) is satisfied, then the process stops and it is set:

\[
X'(h,0) = Z_1(h,0), \quad h=1,...,s
\]

and, therefore:

\[
X'(h,d) = Z_1(h,d), \quad h=1,...,s
\]

Otherwise, values \(Z_2(h,0)\) \((h=1,...,s)\) are obtained as follows. First, one of the following conditions is checked:

\[
Z_1(h,0) + 0.02 [ABS(m(h),1 - M(h))] \tag{A12a}
\]

\[
Z_1(h,0) - 0.02 [ABS(m(h),1 - M(h))] \tag{A12b}
\]

where (A.12.a) is used if \(Z_1 < \bar{X}\) and (A.12.b) is used if \(Z_1 > \bar{X}\). Then it is defined:

\[
Z_2(h,0) = Z_1(h,0) + 0.02 \tag{if \(Z_1 < \bar{X}\) and (A12.a) is satisfied}
\]

\[
Z_2(h,0) = Z_1(h,0) - 0.02 \tag{if \(Z_1 > \bar{X}\) and (A12.b) is satisfied}
\]

\[
Z_2(h,0) \tag{otherwise}
\]

(13)

With these initial values then a series \{Z_2(h,d)\} is obtained using:

\[
Z_2(h,d) = Y'(h,d) + Z_2(h,d-1), \quad h=1,...,s, \quad d=1,...,N \tag{A14}
\]

Then the mean of this series is compared with \(\bar{X}\); that is, it is checked whether \(Z_2(h,d)\) satisfies (A.11). If this is the case the process stops and it is set \(X'(h,0) = Z_2(h,0)\). Otherwise, (A.13) is used again to obtain new values \(Z_2(h,d)\). The process continues until a series \{Z_2(h,d)\} which satisfies (A.11) is found and then it is set \(X'(h,0) = Z_2(h,0)\) and, therefore \(X'(h,d) = Z_2(h,d)\). This iterative process could fail because it may happen that a certain series \(Z_2(h,d)\) does not satisfy (A.11) and \(Z_2(h,0)\) does not satisfy (A.12.a) or (A.12.b) for any \(h\). That is to say, after some steps it might happen that, with the limits established in (A.12.a) or (A.12.b), it is not possible to obtain new initial values \(Z_2(h,d)\) which do not coincide with \(Z_2(h,d)\). In this case, \(Z_{next}(h,d) = Z_2(h,d)\), and the iterative process would never end. When this iterative process fails then \(m(h)\) and \(M(h)\) are redefined using the next minimum and maximum values; then the process starts anew.