Solar Energy Vol. 37, No. 3, pp. 179–194, 1986
Printed in the U.S.A.

0038–092X/86 $3.00 + .00

MARKOV PROCESSES AND FOURIER ANALYSIS AS A TOOL TO DESCRIBE AND SIMULATE DAILY SOLAR IRRADIANCE

U. AMATO, A. ANDRETTA, B. BARTOLI, B. COLUZZI, V. CUOMO, F. FONTANA and C. SERIO
Dipartimento di Fisica Nucleare, Struttura della Materia e Fisica Applicata, Università di Napoli, Napoli, Italy

(*Received October* 1984; *revision received February* 1986; *accepted March* 1986)

Abstract—Time series of 20 years of daily solar irradiance data from four Italian stations are analyzed on a statistical basis. It is shown that the irradiation sequences are not stationary, both in the mean and in the variance. They can be determined by three components: (a) a mean, well described by a Fourier series with only one coefficient; (b) a variance about the mean, well fitted by a Fourier expansion with two coefficients; (c) a stochastic component. The stochastic component follows a first order Markov model. Since it has a non-normal distribution, a normalizing transform has been introduced which does not affect its statistical properties.

1. INTRODUCTION

The statistical modeling of global solar irradiation has many practical applications in calculations of the design and performance of solar energy systems. If the statistical distribution is known, a large quantity of data can be characterized with only a few parameters. A typical example is the Liu and Jordan method[1], which uses the frequency distribution of daily insolation values. This technique, referred to as the utilizability method, has been investigated by many authors (see for example [2–6]) and is presently one of the most important tools for predicting the long term average performance of solar energy systems.

The basic hypothesis in this method is that solar radiation is a random variable or process. However, in many cases, e.g., solar thermal or photovoltaic systems with limited storage capacity[7, 8], it is also necessary to know the sequential properties of solar irradiance data: That is, simulation of daily insolation sequences by Markov chain or other stochastic processes may significantly improve performance predictions of solar energy systems. The time constants involved in storage are much less than 1 month: In the case of buildings using passive solar elements the storage practically coincides with the building itself, and its time constant is in order of one to three days[9]; in the case of solar thermal collector systems[10] the storage unit is generally decoupled from the collector and its optimal size depends on the statistical characteristics of daily irradiation sequences (i.e., on the probability of persistence of clear sky conditions); this is also the case of photovoltaic plants[11].

For this reason the knowledge of sequential properties of solar irradiance data is often used in engineering analysis in order to determine proper system sizing and optimal cost/benefit ratios.

In a previous paper we have analyzed the sequences of daily global irradiance in 17 stations all over Italy[12]. We have found that the distribution functions of K_T ratios (daily solar irradiance on a horizontal plane/extra-atmospheric irradiance) depend only on their monthly average values and are independent of season and locality. Such results are in excellent agreement with those of Bendt, Collares-Pereira and Rabl[5] who analyzed data from 90 stations in the American north hemisphere.

In previous studies[13, 14] we have found that the sequences of K_T values are well described by a first order Markov process. Moreover, the first autocorrelation coefficient seems to be independent of season all over Italy. Similar results are quoted in [7, 15] for other European zones.

The results are very useful for two practical applications:

(a) For stations where solar radiation data are directly measured, they allow one to compile a Reference Year simply by starting from the monthly average value of the daily global solar irradiation. The Reference Year obtained has the same statistical properties, i.e., mean, variance and autocovariance, as the historical series from which it was derived.

(b) This method allows one to estimate sequences of daily global solar radiation for all stations where it is not directly measured, if the corresponding monthly average values can be estimated, e.g., using correlations with other atmospheric parameters[16].

The above work focused on monthly time series. Since these time series are quite short, the hypothesis of stationarity and homoscedasticity (see Section 3) are generally well satisfied[17, 18]. Consequently, analysis can be performed using quite simple tools. On the other hand this approach does not allow one to characterize the structure of longer term time series well, especially if one is interested in estimating the contribution of systematic periodic components as well as stochastic ones. Moreover, the estimation of autocorrelation coefficients for short period time series is affected by bias errors[17, 18].

In order to give better statistical significance to the results obtained in previous papers, in this study we analyzed measured daily horizontal global solar irradiance data from four Italian stations (listed in Table 1), for which 20 years of data (1958–1977) are available. We have found for all stations that the time series from daily solar irradiance data are nonstationary both in the mean and in the variance and can be well described by a first order Markov process when detrended correctly. We also show that daily means and standard deviations of the daily solar irradiance time series can be fitted by a Fourier series using only a few coefficients.

Table 1. List of the examined stations and corresponding latitude and longitude

Station	Latitude (N)	Longitude (E)
Alghero	40°38′	8°17′
Ancona	43°37′	13°31′
Napoli	40°51′	14°18′
Vigna di Valle	42°5′	12°13′

Figure 1 shows a map of Italy with the four meteorological stations whose data we analyzed. They cover three of the four climatic subzones in which Italy is traditionally divided[19], but it should be noted that Italy has a quite homogeneous climate. To confirm this, we here report, for stations located in the four zones, the annual trend of R (average monthly irradiation normalized to extraatmospheric irradiation (Fig. 2)) and the correlation of R

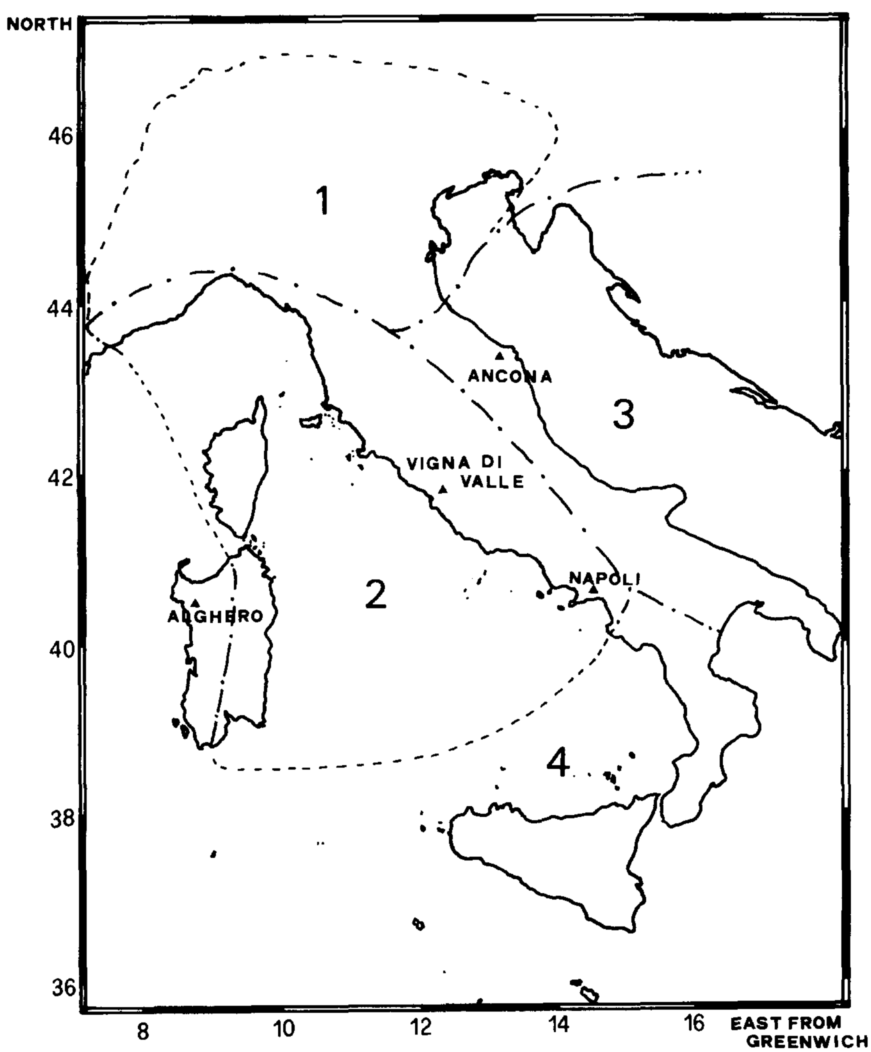

Fig. 1. Map of the meteorological stations whose data have been analyzed. The numbers 1, 2, 3 and 4 in the figure refer to the four climatic subzones in which Italy is divided.

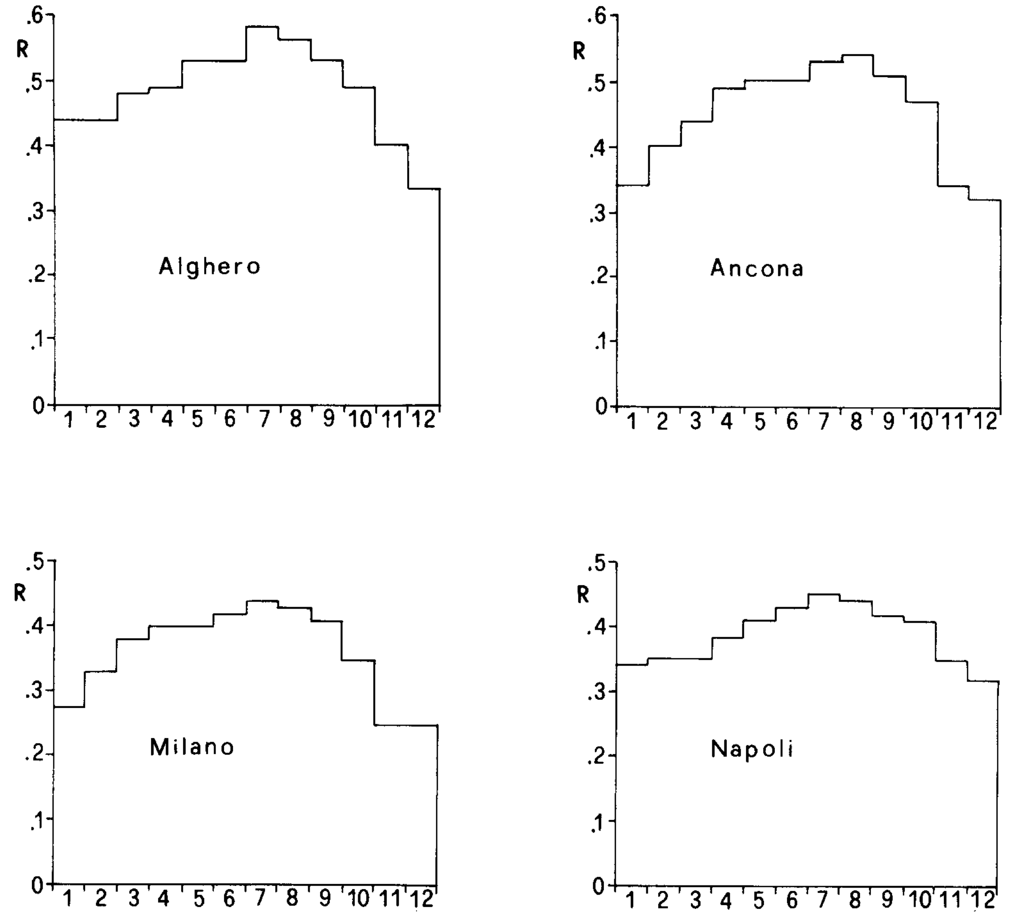

Fig. 2. Annual trend of the average monthly irradiance normalized to extra-atmospheric irradiance (R) for four stations located in the four Italian climatic sub-zones (Milano: Zone 1; Ancona: Zone 2; Napoli: Zone 3; Alqhero: Zone 4); (see also Fig. 1).

with I (monthly average of actual sunshine hours normalized to day length, Fig. 3).

2. DESCRIPTION OF THE MODEL: BASIC ASSUMPTIONS, EQUATIONS AND DEFINITIONS

In this section we introduce the model used to describe daily solar irradiance time series. We discuss the basic assumptions on which the model is based and describe the methods that allow us to compute its parameters.

Much research in geophysical time series analysis has been concerned with the use of simple stationary stochastic processes in order to describe experimental data. A stationary stochastic process is simply described by its autocovariance function or, equivalently, by its power spectrum, which is the Fourier transform of the autocovariance function.

However, geophysical time series generally are not stationary, e.g., rainfall sequences[20] and daily ambient temperature sequences[21]. Obviously, the stationarity assumption or hypothesis can be tested using suitable statistical procedures. If an empirical series is not stationary, various methods can be used to remove obvious trends, leaving behind a residual series for which it is reasonable to assume stationarity.

The mean solar irradiance exhibits an obvious

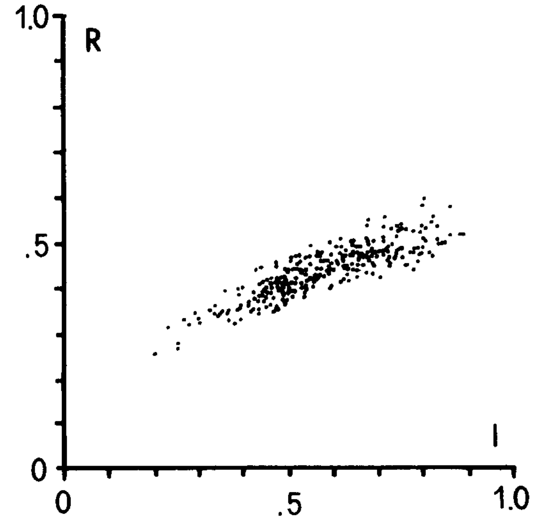

Fig. 3. Monthly values of R and I for 28 Italian meteorological stations. All data are averaged over the period 1958–1969.

seasonal trend; in Section 3 we show that its variance exhibits a seasonal trend as well.

If we assume that daily solar irradiance time series are nonstationary only in the lower order moments (i.e., mean and variance), then these 1-day discrete time series can be described according to the mathematical model[17, 18, 20]:

$$H(i, t) = \mu(t) + \sigma(t)\cdot X(i, t) \qquad (1)$$
$$t = 1, \ldots, \omega$$
$$i = 1, \ldots, n$$

where

- $H(i, t)$ is the experimental value of daily global solar irradiance recorded on the tth day of the ith year;
- ω is the basic period of the series expressed as a discrete number;
- n is the number of periods ω in the finite series;
- $\mu(t)$ is the trend in the mean;
- $\sigma(t)$ is the trend in the standard deviation;
- $X(i, t)$ is the stochastic component.

For daily solar irradiance the year is a basic period: For this reason we assume $\omega = 365$ so that n is the number of years (20 in our analysis) and $n\cdot\omega = 7300$ is the length of the sample series.

Equation (1) models the daily solar irradiance as a ciclostationary signal $\mu(t)$, i.e., the mean component of the global signal $H(i, t)$, with random shocks $X(i, t)$. The strength of the shock (i.e., its variance) is modulated by another ciclostationary signal $\sigma(t)$ in order to take into account the seasonal variability of the fluctuations in $H(i, t)$. By a physical point of view $\mu(t)$ and $\sigma(t)$ are low frequency signals which model the effects of climate. Both $\mu(t)$ and $\sigma(t)$ are assumed ciclostationary (i.e., they do not depend on the year: This dependence is explicited by index i in eqn (1)) because systematic variations of climate have a larger time scale than the time period considered in this paper.

The random variable $X(i, t)$ is a high frequency signal which models the short time effects of the weather.

Fundamental frequencies or harmonics induced in $H(i, t)$ by climate can be detected by Fourier analysing $\mu(t)$ and $\sigma(t)$ which have no weather fluctuations, all contained in the $X(i, t)$ term of eqn (1). Reflecting their physical significance, $\mu(t)$ and $\sigma(t)$ are low frequency signals, so it is expected that only the first two or three harmonics are needed to explain their time behaviour. Obviously $\mu(t)$ is determined mostly by the theoretical sunshine hour number in a day. This number is a periodic function of time with a period equal to one year, therefore it is expected that $\mu(t)$ would have the same time behaviour.

Experimental data show that, independently of the climatic zone, $\mu(t)$ has a time behaviour of the type:

$$\mu(t) = A_0 + A_1 \cos \frac{2\pi t}{\omega} + A_2 \sin \frac{2\pi t}{\omega},$$

as shown in Section 5 for our case and in [7] for other climatic zones.

On the other hand we have that the first two harmonics are both significant for $\sigma(t)$ (see Section 5), but we have no information for other climatic zones.

The value of the variance around $\mu(t)$, i.e., $\sigma(t)$, is mostly determined by clouds and their distribution over the year, so it has a time behaviour which is quite dependent on the type of climate. However, it is experimentally well established[20] that geophysical time series (on a daily basis) never exhibit more than five or six significant harmonics if short term fluctuations are filtered out.

Now we briefly quote the relations we use to estimate $\mu(t)$, $\sigma(t)$ and $X(i, t)$.

For each fixed t (day of the year) the estimates $\hat{\mu}(t)$ of $\mu(t)$, and $\hat{\sigma}(t)$ of $\sigma(t)$ are given by:

$$\hat{\mu}(t) = \frac{\sum_{i=1}^{n} H(i, t)}{n} \qquad (2)$$

$$\hat{\sigma}(t) = \left[\frac{1}{n-1}\sum_{i=1}^{n} ((H(i, t) - \hat{\mu}(t))^2\right]^{1/2}. \qquad (3)$$

With $\hat{\mu}(t)$ and $\hat{\sigma}(t)$ computed from data according to eqns (2) and (3), it is then possible to compute the values of the residual time series $X(i, t)$ from eqn (1):

$$X(i, t) = \frac{H(i, t) - \hat{\mu}(t)}{\hat{\sigma}(t)}. \qquad (4)$$

This transformation in a new series with zero mean and unit standard deviation completely removes trends both in the mean and in the standard deviation from the original series $H(i, t)$. Then, if the basic model assumptions, i.e., stationary higher moments, are well satisfied, it is expected that $X(i, t)$ will not display more complex form of nonstationarity, so that a simple mathematical model for $X(i, t)$ may be developed.

In Section 4 we use $\hat{\mu}(t)$ and $\hat{\sigma}(t)$ to build the residual time series $X(i, t)$. We find that the variable X is not independent and that its empirical distribution function is not gaussian. We also show that a first order autoregressive model of the form

$$X(i, t) = \rho\cdot X(i, t - 1) + Y(i, t) \qquad (5)$$

adequately decribes the persistence properties of the series $X(i, t)$, with the same autocorrelation coefficient ρ for the four stations we studied. If we use the linear model of eqn (5) to simulate the series

X, we obtain data with the same mean, variance and autocorrelation of the experimental data but we cannot reproduce the distribution function since it is not gaussian. To avoid this difficulty we introduce a nonlinear transform.

In Section 5 we show that the series $\mu(t)$ and $\sigma(t)$ (see eqn (1)) are well fitted by a Fourier expansion with only one or two harmonic terms.

Finally, in Section 6, we compare the statistical properties of data simulated by our model with the experimental data.

3. VARIANCE ANALYSIS AND THE HOMOSCEDASTICITY HYPOTHESIS

If a time series does not exhibit a trend in the standard deviation, i.e., if the standard deviation is constant along the series within the sample variations, then the series is said to be homoscedastic.

We have tested the homoscedasticity hypothesis for the daily solar irradiance by using the F-test for variance comparison which gives significant results when applied to small samples (20 data samples in our case) from non-normal populations[22].

For each examined station the corresponding daily solar irradiance time series, $H(i, t)$, gives ω ($\omega = 365$) samples:

$$\{H(i, 1), i = 1, \ldots, n\}, \ldots, \{H(i, \omega), i = 1, \ldots, n\},$$

with $n = 20$ (n is the number of years of available data), then the F-test is applied to these ω samples. The results indicate that the daily standard deviations exhibit a seasonal trend.

In Fig. 4 for each examined station we show the plot of $\hat{\sigma}(t)$ (estimated by eqn (3)) vs. $\hat{\mu}(t)$ (estimated by eqn (2)). These plots not only confirm that $\sigma(t)$ varies, but also show that $\sigma(t)$ is correlated with $\mu(t)$ albeit not linearly. Consequently, it is expected that $\hat{\sigma}(t)$ and $\hat{\mu}(t)$ will show different harmonics (different frequencies) and phases when fitted with a Fourier series.

4. STATISTICAL ANALYSIS OF THE RESIDUAL TIME SERIES $X(i, t)$

Using the experimental data of $H(i, t)$ and the computed values of $\hat{\mu}(t)$ and $\hat{\sigma}(t)$ (eqns (2) and (3)) we can obtain the residual time series $X(i, t)$ according to eqn (5). In Fig. 5 we show the empirical frequency distribution of $X(i, t)$ for the four examined stations; it is interesting to note that these distributions are not gaussian.

The time series $X(i, t)$ can be described as an $n \cdot \omega$ matrix: the ith row of this matrix ($i = 1, \ldots, n$) contains the ω residuals corresponding to the ith year, while the tth column of this matrix contains the n residuals corresponding to the same day across the years:

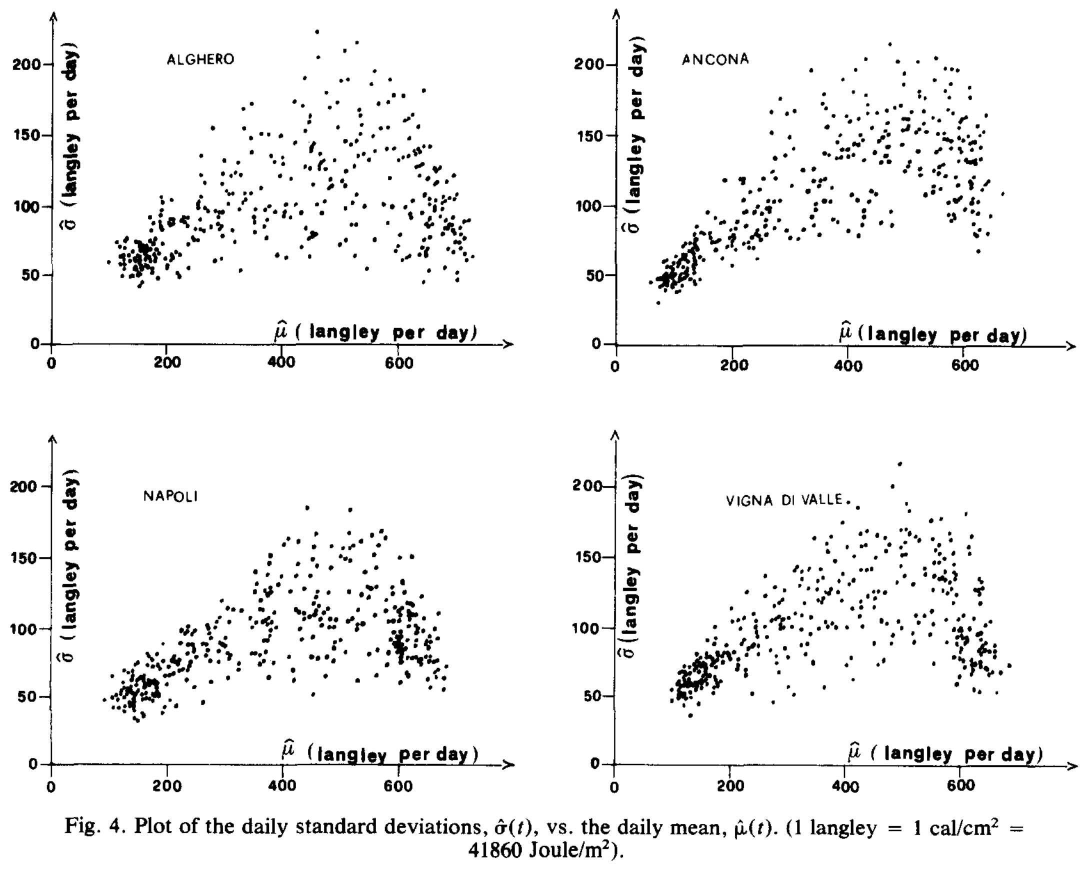

Fig. 4. Plot of the daily standard deviations, $\hat{\sigma}(t)$, vs. the daily mean, $\hat{\mu}(t)$. (1 langley = 1 cal/cm^2 = 41860 Joule/m^2).

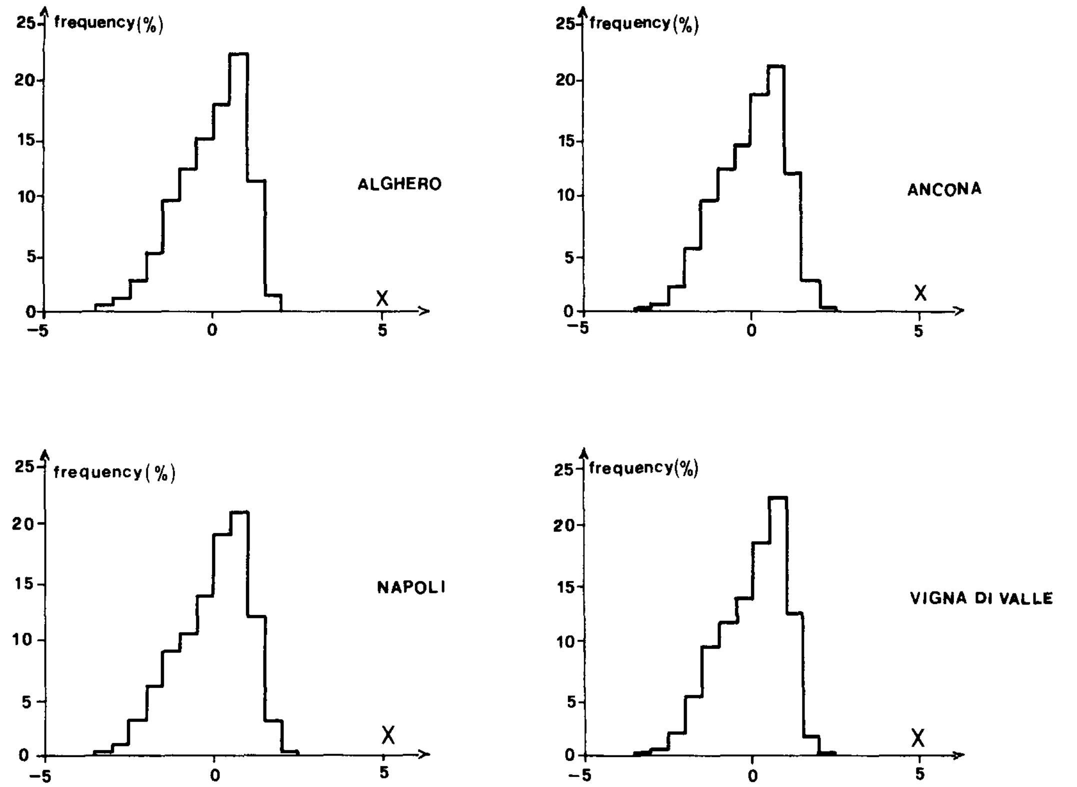

Fig. 5. Empirical frequency distributions of the residual time series X.

$$\begin{array}{cccc|c} X(1, 1) & \cdots\cdots & X(1, \omega) & & \hat{T}_1^0 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & & \cdots\cdots \\ X(n, 1) & \cdots\cdots & X(n, \omega) & & \hat{T}_n^0 \\ \hline \hat{T}_1 & \cdots & \hat{T}_\omega & & \end{array}$$

Then we check the hypothesis that the residuals $X(i, t)$ are independent by rows and columns, using the rank-test[18]. The rank-test statistic, $\hat{T}$, is computed across the ensemble of the series ($\hat{T}_t$, $t = 1, \ldots, \omega$) and along the discrete series ($\hat{T}_i$, $i = 1, \ldots, n$).

The results indicate that the data are truly random by columns but autocorrelated by rows.

In order to eliminate this persistence, we tried a first order autoregressive model of the form:

$$X(i, t) = \rho \cdot X(i, t - 1) + Y(i, t). \tag{6}$$

An estimate $\hat{\rho}$ of ρ was obtained by[17, 18, 20]:

$$\hat{\rho} = \frac{\dfrac{1}{n \cdot \omega - 1} \sum_k X(k) \cdot X(k + 1)}{\left[\dfrac{1}{n \cdot \omega - 1} \sum_k X^2(k) \dfrac{1}{n \cdot \omega - 1} \sum_k X^2(k + 1) \right]^{1/2}}, \tag{7}$$

where $X(k)$, $k = 1, \ldots, n \cdot \omega - 1$ represents the kth term of the series $X(i, t)$ ordered year by year. For the four examined stations we obtain the same value for the autocorrelation coefficient, $\hat{\rho} = 0.31$.

We have then computed the series:

$$Y(i, t) = X(i, t) - \hat{\rho} \cdot X(i, t - 1) \tag{8}$$

and we have checked the hypothesis that the residuals $Y(i, t)$ are independent by rows and columns, using the rank-test. This test confirms at a significance level of 5% that $Y(i, t)$ is an independent variable.

This result suggests that the time series $X(i, t)$ can be represented by a first order autoregressive model, but the adequacy of the model (6) can be shown by comparing the statistics of simulated and experimentally observed data. A simulated series $X(i, t)$ can be computed with the following procedure:

(a) generate a series of $n \cdot \omega$ random independent numbers with the frequency distribution of Y.

(b) compute values of X using recursively eqn (6) with $\rho = 0.31$, i.e.,

$$\begin{cases} X(0) = Y(0) \\ X(1) = \rho \cdot X(0) + Y(1) \\ \cdots\cdots\cdots\cdots \\ X(j) = \rho \cdot X(j - 1) + Y(j) \end{cases}$$

$$j = 1, \ldots, n \cdot \omega - 1. \tag{9}$$

Using this procedure we have computed 7300 $X(i, t)$ values for each examined station; in Fig. 6

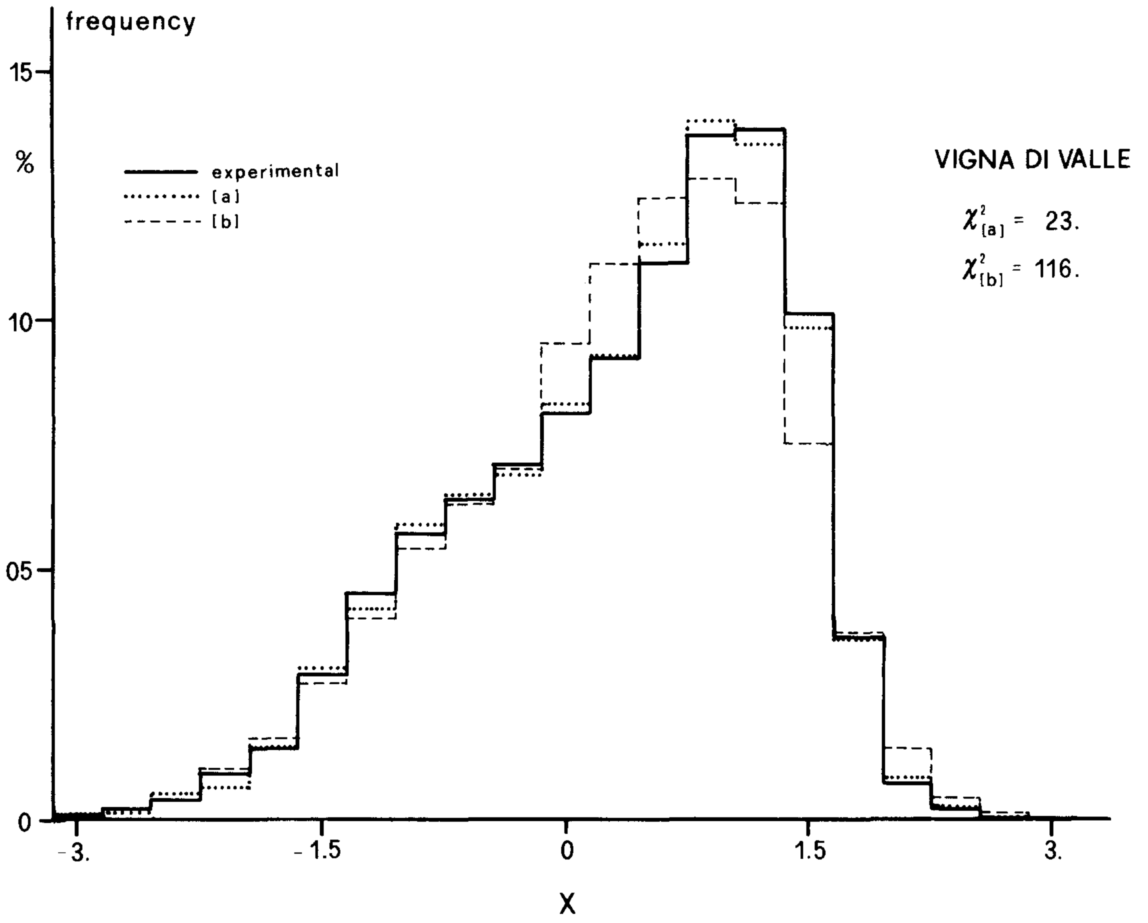

Fig. 6. Comparison between the empirical frequency distribution of the residual X for the Vigna di Valle station and the sample frequency distribution of 7300 simulated residuals X. The dashed curve refers to the simulated values obtained according to eqn 9; for this case the usual χ^2-test shows that the experimental and simulated frequency distributions are statistically different. The dotted curve refers to the simulated values obtained using eqn 13 and eqn 11. In this case no statistical difference is evidenced between experimental and simulated frequency distributions. The observed values of the χ^2-test statistic are shown in the figure for both the curves; the number of freedom degrees is 20.

we show the frequency distributions of the simulated and experimental data for the Vigna di Valle station. These two distributions are very different. Similar results were obtained for the other three stations.

The failure of the linear model (6) can be explained as follows. The difference equations (9) integrate to:

$$X(j) = \rho^j \cdot Y(0) + \sum_{i=1}^{j} \rho^{j-i} \cdot Y(i), \tag{10}$$

so that $X(j)$ is the summation of $j + 1$ independent stochastic variables. Then as j increases the frequency distribution of $X(j)$ becomes similar to a gaussian distribution, as predicted by the Central Limit Theorem of Statistics. So, there is no way to generate sequences with a specified asymmetric distribution using linear autoregressive models.

Since classical methods in time series analysis[17, 18, 20] allow us to analyze the structure of series and give good estimates of related statistical parameters only when dealing with normal stochastic processes, we introduce a transform in order to build a new normal series from the original non-normal series $X(i, t)$.

The empirical distribution function of $X(i, t)$ is shown in Fig. 5; the corresponding integrated distribution curve, $F(X)$ (dropping indexes i and t for simplicity) can be used to transform the sequence of $X(i, t)$ data in the following manner. We set

$$\operatorname{erf}(Z) = F(X), \tag{11}$$

where erf(Z) is the error function defined as:

$$\operatorname{erf}(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} \left[\exp\left(-\frac{x^2}{2}\right)\right] dx. \tag{12}$$

Equation (11) when solved for $Z(i, t)$ gives a new time series with zero mean, unit standard deviation and normal distribution. The transform (11) also preserves the structure of Markov processes, i.e., if $X(i, t)$ is a Markov process then $Z(i, t)$ is also a Markov process, and conversely. These properties of transform (11) can be demonstrated mathematically in a way quite general. However they are intuitive, since the transform (11) maps Z into X (and conversely) in a one-to-one correspondence.

We found for all stations that $Z(i, t)$ can be described as a first order autoregressive process:

$$Z(i, t) = \rho \cdot Z(i, t-1) + U(i, t) \qquad (13)$$
$$t = 1, \ldots, \omega$$
$$i = 1, \ldots, n$$

where

ρ is the first autocorrelation coefficient;
$U(i, t)$ is a normal, independent variable with zero mean.

An estimation $\hat{\rho}$ of ρ is given by [17, 18, 20]:

$$\hat{\rho} = \frac{\frac{1}{n\cdot\omega - 1}\sum_k Z(k)\cdot Z(k+1)}{\left[\frac{1}{n\cdot\omega - 1}\sum_k Z^2(k)\cdot\frac{1}{n\cdot\omega - 1}\sum_k Z^2(k+1)\right]^{1/2}}, \qquad (14)$$

where $Z(k)$, $k = 1, \ldots, n\cdot\omega - 1$ represents the kth value of the series $Z(i, t)$ obtained by setting the data in progressive year by year order.

For the four examined stations we obtain the same value $\hat{\rho} = 0.33$, computed according to eqn (14). Moreover, the series $Z(i, t)$ appears to be stationary; in other words, the autocorrelation coefficient is independent of time. Such a result was inferred in this way. We recall that for normal autoregressive processes the variances $\sigma^2(U)$, $\sigma^2(Z)$ and the first autocorrelation coefficient ρ are related by:

$$\sigma^2(U) = (1 - \rho^2)\cdot\sigma^2(Z). \qquad (15)$$

In our case $\sigma^2(Z) = 1$; so that the problem of systematic variations of ρ along the time series $Z(i, t)$ can be investigated by testing the homoscedasticity of the time series $U(i, t)$. Again, we use the F-test for variance comparison[22], applied to the twenty yearly samples of the time series $U(i, t)$:

$$\{U(1, t), t = 1, \ldots, \omega\}, \ldots, \{U(20, t), t = 1, \ldots, \omega\}.$$

Experimental values of $U(i, t)$ were obtained inverting eqn (13), i.e.,

$$U(i, t) = Z(i, t) - \hat{\rho}\cdot Z(i, t-1), \qquad (16)$$

with $\hat{\rho} = 0.33$.

The F-test results confirm at a 5% significance level that variations of $\sigma^2(U)$, and therefore of ρ, are not statistically significant.

So far our analysis has been concerned with the statistical properties of $Z(i, t)$. Recall that the transform (11) leaves the Markovian structure invariant, so that all results obtained for $Z(i, t)$ hold for $X(i, t)$ as well. In other words $X(i, t)$ follows a stationary Markov process. Furthermore, this process admits a linear representation in the "Z domain" but not in the "X domain". This is the main advantage of using the transform (11), because in the X domain the process $X(i, t)$ can be only described by its joint distribution function, while in the Z domain it is described by the simple linear relation (13); we also note that the transform uses only the distribution function, $F(X)$.

Using eqns (13) and (111) we can now compute a simulated series $X(i, t)$ which has the same mean, variance, autocorrelation and distribution function as the experimental data. The procedure is as follows:

(a) Generate a series of $n\cdot\omega$ random numbers $U(i, t)$, $i = 1, \ldots, n$; $t = 1, \ldots, \omega$, with zero mean, variance equal to $(1 - \rho^2)$, $\rho = 0.33$, and gaussian distribution;

(b) Use relation (13) to build the series

$$Z(i, t) = \rho\cdot Z(i, t-1) + U(t).$$

(c) Use relation (11), i.e.,

$$F(X) = \mathrm{erf}(Z),$$

to find the values of $X(i, t)$.

Using this procedure we have computed 7300 simulated values, $X(i, t)$, for each one of the examined stations. Figure 6 shows the distribution functions of the simulated and experimental data for the Vigna di Valle station; similar results were obtained for the other stations.

Finally in Fig. 7 we show the sample integrated distribution function of the variable X obtained by grouping the $X(i, t)$ values for all the examined stations.

A χ^2 test confirms that there is no significative difference among the distribution function so obtained and the four distribution functions of X corresponding to each one of the stations examined.

5. ANALYSIS OF THE SERIES $\mu(t)$ AND $\sigma(t)$

The model obtained in the previous section for $X(i, t)$ can be used in eqn (1) to simulate values of daily irradiance. Of course the series $\mu(t)$ and $\sigma(t)$ are also needed. We could use the values of $\mu(t)$ and $\sigma(t)$ computed on the basis of the historical time series of daily solar irradiance. For practical applications, however, such approach is ill-advised, since it requires a knowledge of 365 × 2 parameters.

In this section we show that a short Fourier se-

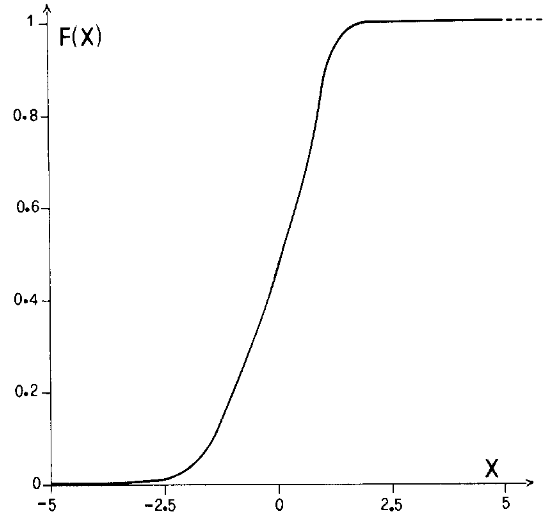

Fig. 7. Cumulative frequency distribution of the variable X obtained grouping the X values for all the examined stations.

ries can well describe the computed values of $\mu(t)$ and $\sigma(t)$.

5.1 *The periodic component in the mean,* $\mu(t)$*: estimation of the significant harmonics*

From each one of the empirical daily solar irradiance time series, $H(i, t)$, we compute the $\omega(\omega = 365)$ daily means $\hat{\mu}(t)$, according to eqn (2). The Fourier coefficients, A_j and B_j, for these series are estimated according to [17, 18, 20]:

$$A_j = \frac{2}{\omega}\sum_{t=1}^{\omega}(\hat{\mu}(t) - \langle\mu\rangle)\cdot\cos(2\pi f_j t), \quad (17.1)$$

$$B_j = \frac{2}{\omega}\sum_{t=1}^{\omega}(\hat{\mu}(t) - \langle\mu\rangle)\cdot\sin(2\pi f_j t), \quad (17.2)$$

with f_j being the ordinary frequency (i.e., $f_j = j/\omega$, $j = 1, 2, \ldots$) and with

$$\langle\mu\rangle = \frac{1}{\omega}\sum_{t=1}^{\omega}\hat{\mu}(t). \quad (18)$$

We recall that

$$\frac{1}{2}\sum_{j=1}^{\omega}(A_j^2 + B_j^2) = \sigma^2(\hat{\mu}(t)), \quad (19)$$

where $\sigma^2(\hat{\mu}(t))$ is the variance of the 365 computed daily means, $\hat{\mu}(t)$, i.e.,

$$\sigma^2(\hat{\mu}(t)) = \frac{1}{\omega}\sum_{t=1}^{\omega}(\hat{\mu}(t) - \langle\mu\rangle)^2. \quad (20)$$

Therefore the cumulative discrete spectrum, $S(k)$, shows how the variance of the process is distributed with frequency:

$$S(k) = \sum_{j=1}^{k}\frac{A_j^2 + B_j^2}{2\cdot\sigma^2(\hat{\mu}(t))}. \quad (21)$$

$$j = 1, 2, \ldots.$$

In all cases we found that $S(k)$ has a very sharp rise in correspondance with the first harmonic; for $k > 1$ the rise is very slow and becomes practically linear in k. Such results are reported in Fig. 8. In all cases the first harmonic, corresponding to the period ω, explains more than 95% of the total variance, $\sigma^2(\hat{\mu}(t))$. Consequently, we conclude that only the first harmonic is significant.

Therefore, the empirical series $\hat{\mu}(t)$ can be described mathematically as:

$$\hat{\mu}(t) = \langle\mu\rangle + A_1\cdot\cos\left(\frac{2\pi t}{\omega}\right) + B_1\cdot\sin\left(\frac{2\pi t}{\omega}\right).$$

$$t = 1, \ldots, \omega \quad (22)$$

For each locality the values of $\langle\mu\rangle$, A_1, B_1, $\sigma^2(\hat{\mu}(t))$ and the ratio $(A_1^2 + B_1^2)/(2\cdot\sigma^2(\hat{\mu}(t)))$ are given in Table 2. Figure 9 shows the fitted function, $\mu(t)$, and the 365 computed daily means, $\hat{\mu}(t)$.

5.2 *The periodic component in the standard deviation* $\sigma(t)$*: estimation of the significant harmonics*

From each one of the empirical daily solar irradiance time series, $H(i, t)$, we compute the ω daily standard deviations, $\hat{\sigma}(t)$, $t = 1, \ldots, \omega$, according to eqn (3). The Fourier coefficients D_j and E_j for these series are estimated according to

$$D_j = \frac{2}{\omega}\sum_{t=1}^{\omega}(\hat{\sigma}(t) - \langle\sigma\rangle)\cdot\cos(2\pi f_j t), \quad (23.1)$$

$$E_j = \frac{2}{\omega}\sum_{t=1}^{\omega}(\hat{\sigma}(t) - \langle\sigma\rangle)\cdot\sin(2\pi f_j t), \quad (23.2)$$

where f_j is always the ordinary frequency ($f_j = j/\omega$) and $\langle\sigma\rangle$ the mean of the 365 computed daily standard deviations, $\hat{\sigma}(t)$, i.e.,

$$\langle\sigma\rangle = \frac{1}{\omega}\sum_{t=1}^{\omega}\hat{\sigma}(t). \quad (24)$$

In this case the cumulative discrete spectrum, $SS(k)$, is given by

$$SS(k) = \sum_{j=1}^{k}\frac{D_j^2 + E_j^2}{2\cdot\sigma^2(\hat{\sigma}(t))}, \quad (25)$$

where $\sigma^2(\hat{\sigma}(t))$ is the variance of the 365 computed daily standard deviations, $\hat{\sigma}(t)$; i.e.,

$$\sigma^2(\hat{\sigma}(t)) = \frac{1}{\omega}\sum_{t=1}^{\omega}(\hat{\sigma}(t) - \langle\sigma\rangle)^2. \quad (26)$$

In all cases the plot of $SS(k)$ vs. k (see Fig. 10)

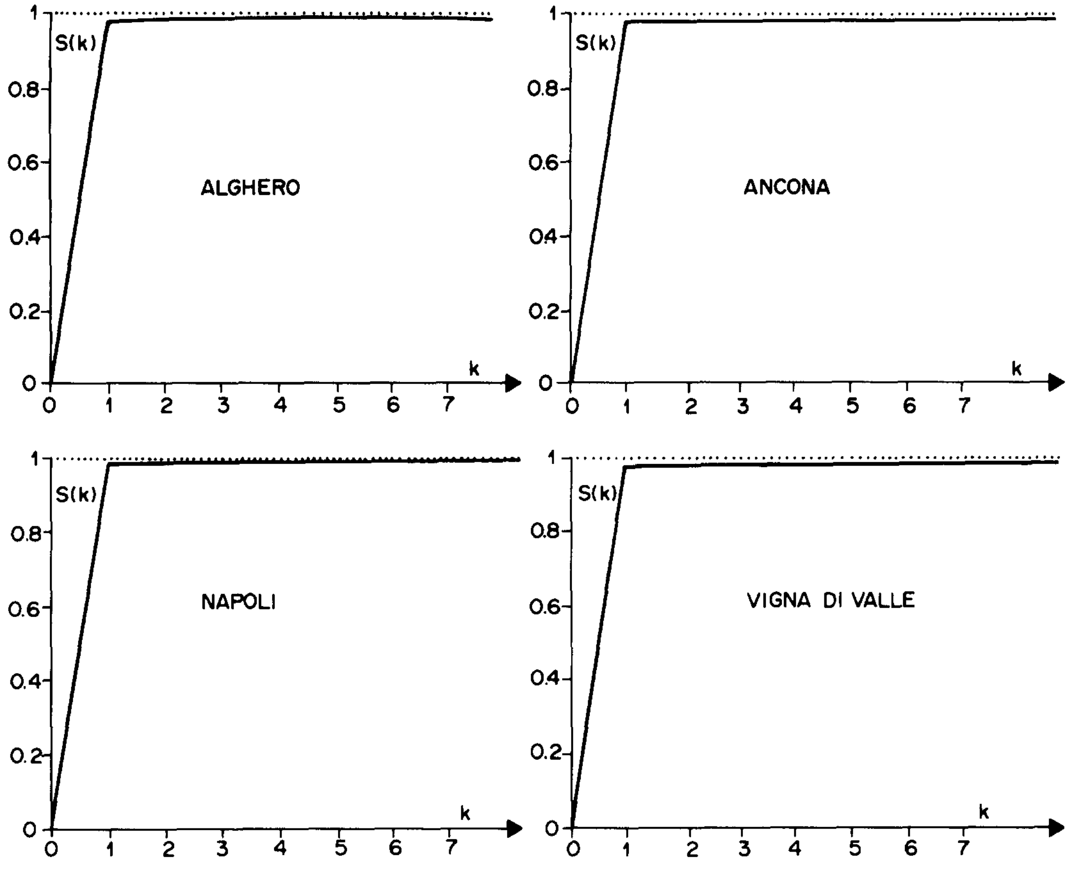

Fig. 8. Variance, S, explained by K harmonics in the daily mean series $\hat{\mu}(t)$.

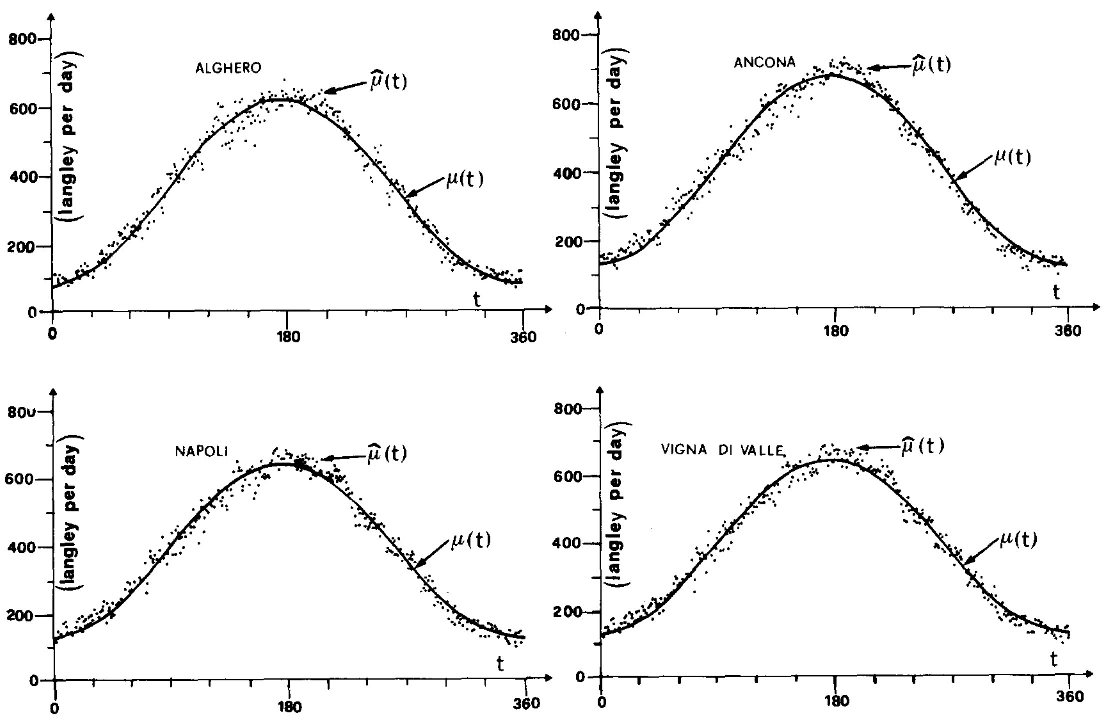

Fig. 9. Fit of the periodic function of the 365 values of the daily means; $\mu(t)$ indicates the fitted function and $\hat{\mu}(t)$ the 365 computed daily means. (1 langley = 1 cal/cm^2 = 41860 J/m^2).

Table 2. Computed values of the parameters related to the Fourier series expansion about $\hat{\mu}(t)$. A_1 and B_1 indicate the Fourier coefficients of the first harmonic; $\langle\mu\rangle$ the mean of the series $\hat{\mu}(t)$; $\sigma(\hat{\mu}(t))$ the standard deviation of the series $\hat{\mu}(t)$; the last column of the table gives the variance explainable by the first harmonic. A_1, B_1, $\langle\mu\rangle$ and $\sigma(\hat{\mu}(t))$ are in KJ/m^2 per day

Station	A_1	B_1	$\langle\mu\rangle$	$\sigma(\hat{\mu}(t))$	$\frac{A_1^2 + B_1^2}{2\sigma^2(\hat{\mu}(t))}$
Alghero	−11440	622	16487	8194	98%
Ancona	−11176	892	14667	8018	98%
Napoli	−10638	748	16054	7622	98%
Vigna di Valle	−10602	692	15438	7610	97%

shows a sharp rise of the curve $SS(k)$ up to $k = 2$ and then a slow rise for $k > 2$. Consequently, we conclude that only the first two harmonics are significant. Therefore, the empirical series $\hat{\sigma}(t)$ can be adequately described mathematically as:

$$\hat{\sigma}(t) = \langle\sigma\rangle + D_1\cdot\cos\left(\frac{2\pi t}{\omega}\right) + D_2\cdot\cos\left(\frac{4\pi t}{\omega}\right) + E_1\cdot\sin\left(\frac{2\pi t}{\omega}\right) + E_2\cdot\sin\left(\frac{4\pi t}{\omega}\right). \tag{27}$$

For each locality the values of $\langle\sigma\rangle$, E_1, E_2, D_1, D_2, $\sigma^2(\hat{\sigma}(t))$ and the ratio $\sum_{j=1}^{2} D_j^2 + E_j^2/2\cdot\sigma^2(\hat{\sigma}(t))$ are given in Table 3. Figure 11 shows the fitted function, $\sigma(t)$, and the 365 computed standard deviations $\hat{\sigma}(t)$.

In this case the first two harmonics (see Table 3) explain more than 65% of the variance $\sigma^2(\hat{\sigma}(t))$.

The overall variance, σ_H^2, of the time series $H(i, t)$:

$$\sigma_H^2 = \frac{1}{n\cdot\omega}\sum_{i=1}^{n}\sum_{t=1}^{\omega}(H(i, t) - \langle\mu\rangle)^2 \tag{28}$$

can be expressed in terms of $\sigma^2(\hat{\mu}(t))$ (eqn (20)), $\langle\sigma\rangle^2$ (eqn (24)) and $\sigma^2(\hat{\sigma}(t))$ (eqn (26)), i.e.,

$$\sigma_H^2 = \sigma^2(\hat{\mu}(t)) + \langle\sigma\rangle^2 + \sigma^2(\hat{\sigma}(t)), \tag{29}$$

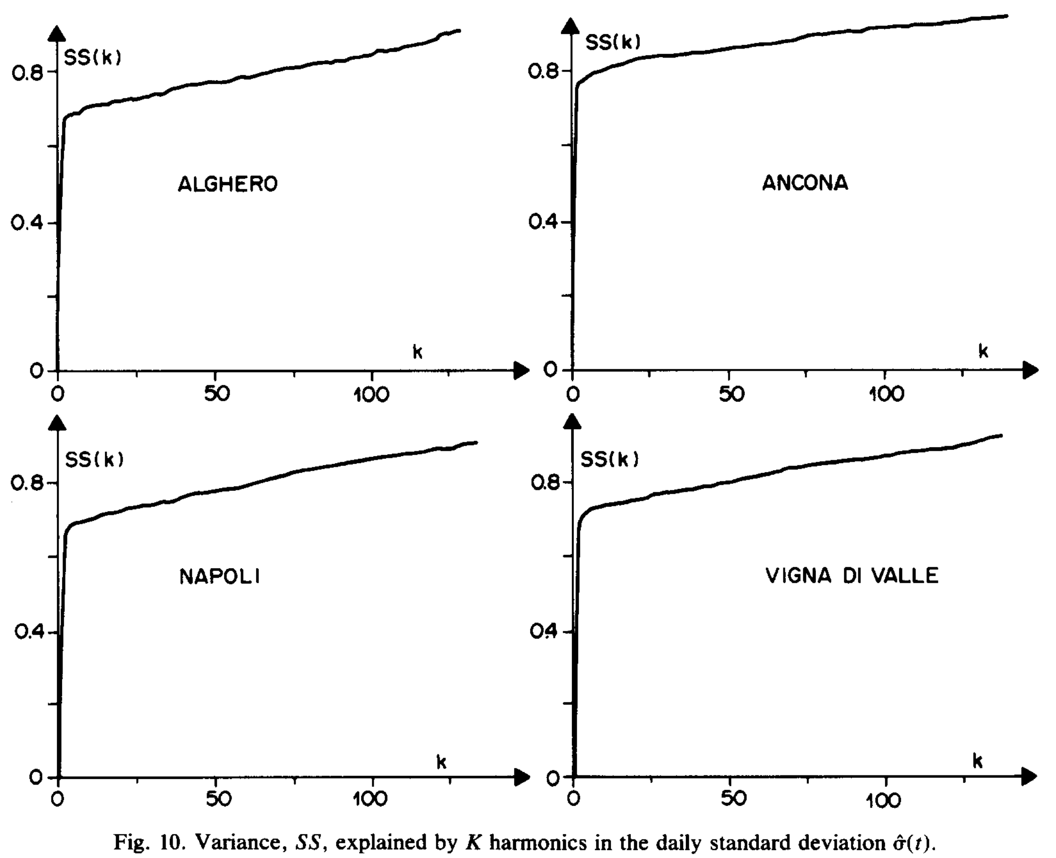

Fig. 10. Variance, SS, explained by K harmonics in the daily standard deviation $\hat{\sigma}(t)$.

Table 3. Computed values of the parameters related to the Fourier series expansion $\hat{\sigma}(t)$. D_1, E_1, D_2 and E_2 indicate respectively the Fourier coefficients of the first and second harmonic; $\langle\sigma\rangle$ the mean of the series $\hat{\sigma}(t)$; $\sigma(\hat{\sigma}(t))$ the standard deviation of the series $\hat{\sigma}(t)$; the last three rows of the table give the variance explainable by the first harmonic only, by the second harmonic only and by the first and second harmonics respectively. D_1, E_1, D_2, E_1, $\langle\sigma\rangle$ and $\sigma(\hat{\sigma}(t))$ are in KJ/m^2 per day

	Alghero	Ancona	Napoli	Vigna di Valle
D_1	−1010	−1929	−1053	−1165
E_1	1107	998	783	1072
D_2	−854	−662	−786	−841
E_2	−452	−172	−239	−274
$\langle\sigma\rangle$	4160	4500	3836	4238
$\sigma(\hat{\sigma}(t))$	1548	1836	1336	1540
$\dfrac{D_1^2 + E_1^2}{2\sigma^2(\hat{\sigma}(t))}$	47%	70%	48%	53%
$\dfrac{D_2^2 + E_2^2}{2\sigma^2(\hat{\sigma}(t))}$	19%	7%	19%	16%
$\dfrac{\sum_{j=1}^{2} D_j^2 + E_j^2}{2\sigma^2(\hat{\sigma}(t))}$	66%	77%	67%	69%

so that the Fourier expansion for $\hat{\sigma}(t)$ and $\hat{\mu}(t)$ explain a fraction, GF, of σ_H^2 given by:

$$GF = \frac{A_1^2 + B_1^2 + 2\cdot\langle\sigma\rangle^2 + D_1^2 + D_2^2 + E_1^2 + E_2^2}{2\cdot\sigma_H^2}. \tag{30}$$

In all cases GF is more than 97% (see Table 4).

It is a characteristic of the climate of the Mediterranean Europe to have the four seasons with weather conditions quite different, as it can be inferred by the bimodal character of the $\sigma(t)$ curves (Fig. 11). Boileau[7] finds a similar result analyzing daily solar irradiance data from Carpentras, France. The bimodal character of the standard deviation distributed over a year's time at the four examined

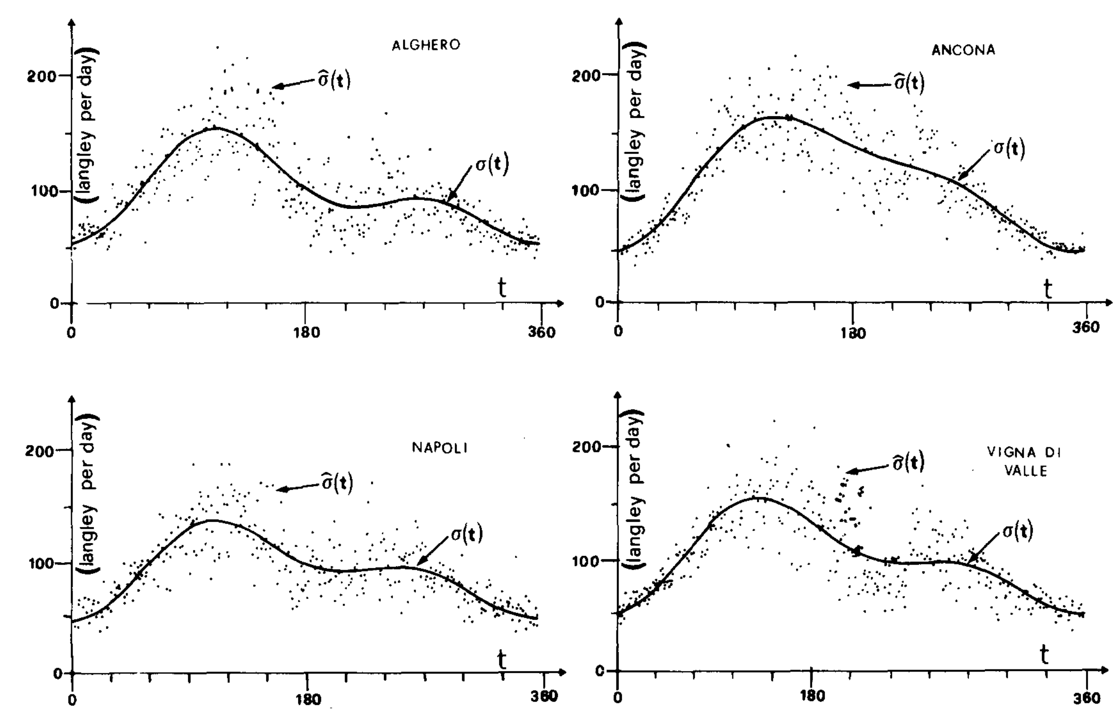

Fig. 11. Fit of the periodic function of the 365 values of the daily standard deviations; $\sigma(t)$ indicates the fitted function and $\hat{\sigma}(t)$ the 365 computed daily standard deviations. (1 langley = 1 cal/cm^2 = 41860 Joule/m^2).

Table 4. Global standard deviation, σ_H (KJ/m^2 per day), of the series $H(i, t)$ and the amount (GF) of the global variance, σ_H^2, explainable by both the fits for $\hat{\mu}(t)$ and $\hat{\sigma}(t)$

Station	σ_H	GF
Alghero	9250	98, 8%
Ancona	9290	99, 3%
Napoli	8625	97, 8%
Vigna di Valle	8750	99, 3%

stations (Fig. 11) is mainly due to the variations in cloud cover through the seasons. In fact in Italy winter and summer months are characterised by weather conditions much more uniform than spring and autumn months: In other words, in winter the sky is prevalently cloudy, in summer it is clear, while in spring and autumn clear and cloudy weather conditions alternate.

As an example, in Fig. 12 we show the function distributions of daily sunshine hours for winter and summer.

6. SIMULATED DAILY SOLAR IRRADIANCE DATA

In the previous sections we showed that a good model for the empirical sequences of daily horizontal solar irradiance, $H(i, t)$, is given by:

$$H(i, t) = \mu(t) + \sigma(t) \cdot X(i, t) \tag{31}$$

where

$\mu(t)$ is the periodic component in the mean given by eqn (22);

$\sigma(t)$ is the periodic component in the standard deviation given by eqn (27);

$X(i, t)$ is a dependent stochastic variable whose distribution is shown in Fig. 7.

We now have all the elements to simulate sequences of daily radiation data using only a few parameters as input. The procedure, easily implemented with a computer program, is the following:

(a) Generate a series of $n \cdot \omega$ random numbers $U(i, t)$, $i = 1, \ldots, n$; $t = 1, \ldots, \omega$ with zero mean, gaussian distribution and standard deviation equal to $(1 - \rho^2)$, $\rho = 0.33$;

(b) Use relation (15), i.e.,

$$Z(i, t) = \rho \cdot Z(i, t - 1) + U(i, t),$$

always with $\rho = 0.33$, to construct a sequence of $n \cdot \omega$ numbers with zero mean, unit standard deviation and normal distribution;

(c) Use relation (13), i.e.,

$$\mathrm{erf}(Z) = F(X)$$

to find the values of $X(i, t)$ corresponding to the $Z(i, t)$ values from Fig. 7;

(d) Obtain the daily global solar radiation sequence from eqn (31).

The daily solar irradiance time series so obtained has the same statistical properties as real data, including the distribution, and can be used for solar systems performance predictions.

As an example, using this procedure, we have computed 7300 simulated values of $H(i, t)$ for each one of the examined stations. In Fig. 13 we show the distribution functions of the simulated data and

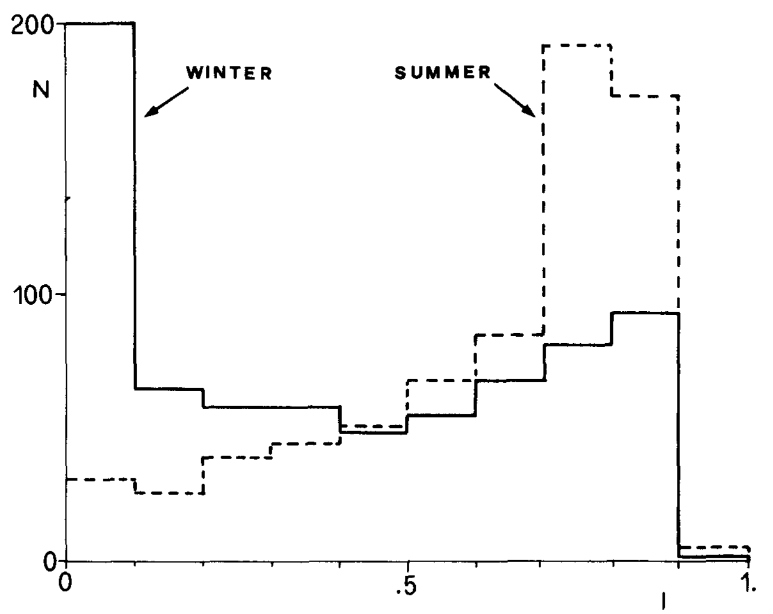

Fig. 12. Typical frequency distribution of daily I values (actual sunshine hours normalized to day length for December–March (continuous curve) and for June–September (dashed curve).

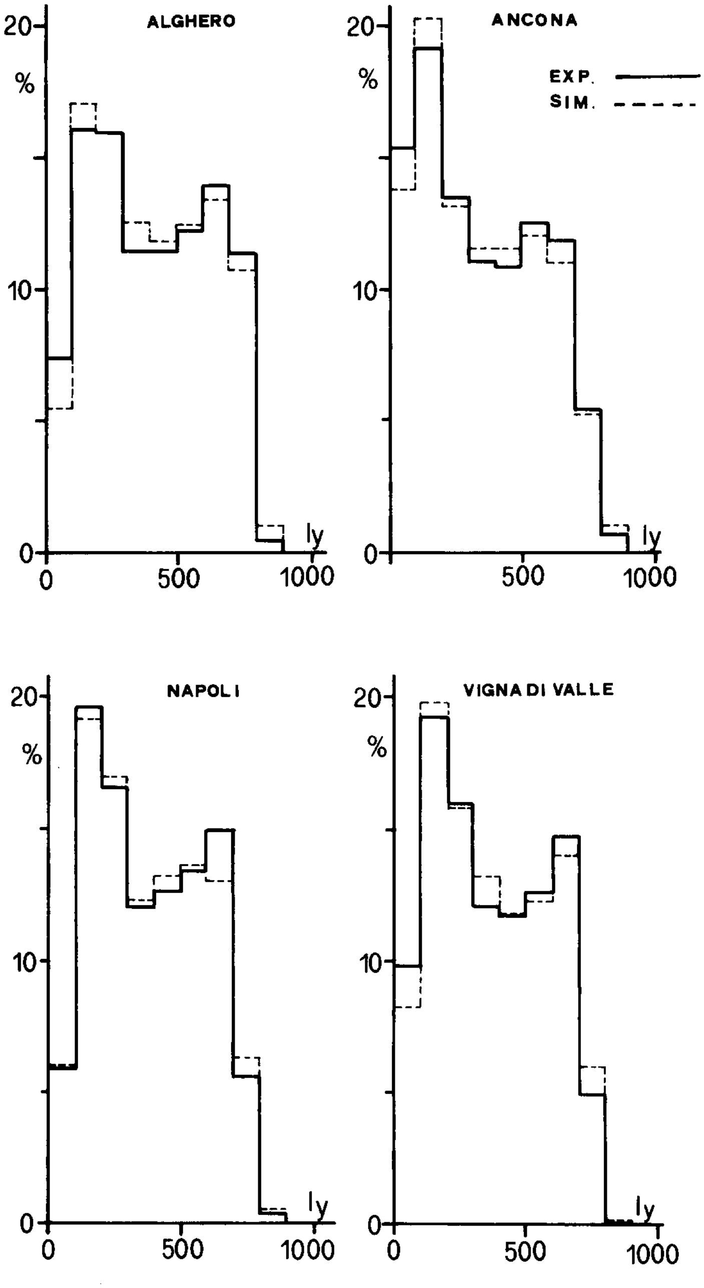

Fig. 13. Frequency distributions of the daily solar irradiance values: comparions between experimental data (continuous line) and simulated data (dashed line) 1 langley = 1 cal/cm^2 = 41860 J/m^2).

Table 5. Mean and standard deviation of the experimental and simulated data (values in KJ/m^2 per day)

Station	Experimental Data		Simulated Data	
	mean	st. dev.	mean	st. dev.
Alghero	16487	9250	16612	9125
Ancona	14667	9290	14753	9170
Napoli	16054	8625	16054	8440
Vigna di Valle	15438	8750	15510	8690

the ones from the experimental data; while in Table 5 we show mean and standard deviation of the experimental and simulated data.

7. CONCLUSIONS

We have analyzed the time series $H(i, t)$ of 20 years of daily global solar irradiance data from four stations in Italy.

We have shown that the time series can be parametrized as:

$$H(i, t) = \mu(t) + \sigma(t) \cdot X(i, t), \tag{32}$$

where both $\mu(t)$ and $\sigma(t)$ are well fitted by expansions in Fourier series (see Section 5).

The series, $X(i, t)$, is a dependent variable with non-normal distribution which follows a first order Markov process.

We have introduced the transform

$$\mathrm{erf}(Z) = F(X), \tag{33}$$

where $\mathrm{erf}(Z)$ is the error function, in order to treat residual series with non-gaussian distributions and consequently deal with relations which leave the frequency distributions invariant.

We have shown that transform (33) does not affect the statistical characteristics of the process, that the series Z is stationary, and that it can be expressed as:

$$Z(i, t) = \rho \cdot Z(i, t - 1) + U(i, t), \tag{34}$$

where U is a normal random variable.

We have described a method which uses relations (32), (34) and the transform (33) to simulate daily solar irradiance values which have the same statistical features (i.e. average, variance and autocovariance) as historical time series. The method requires the following parameters:

(a) the average yearly value of $\mu(t)$ and $\sigma(t)$;

(b) the coefficients of the first term of the expansion in a Fourier series of $\mu(t)$;

(c) the coefficients of the first and the second term of the expansion in a Fourier series of $\sigma(t)$;

(d) the value of the first autocorrelation coefficient, ρ;

(e) the distribution of the variable X in order to perform transform (33).

Variable Z is constructed to be gaussian with zero mean and unitary variance and consequently variable U is gaussian with zero mean and variance $\sigma^2(U) = 1 - \rho^2$.

The autocorrelation coefficient has the same value for the four stations examined. Moreover, the variable X can be described by the same parent distribution independent of season.

It then seems plausible to suggest that parameters (d) and (e) may be constant all over Italy.

NOMENCLATURE

A_j, B_j: jth Fourier coefficients (Fourier series expansion about $\mu(t)$).
D_j, E_j: jth Fourier coefficients (Fourier series expansion about $\sigma(t)$).
f_j: frequency of the jth harmonic of Fourier series expansion: $f_j = j/\omega, j = 1, 2, \ldots$
$F(X)$: frequency distribution of the residual time series $X(i, t)$.
$H(i, t)$: experimental value of the daily horizontal solar irradiance recorded on the tth day of the ith year.
i: index of the year ($i = 1, \ldots, 20$)
t: index of the day of the year ($t = 1, \ldots, 365$).
n: number of years ($n = 20$).
$S(k)$: cumulative discrete spectrum of the series $\hat{\mu}(t)$; $k = 1, 2, \ldots$; see eqn 21.
$SS(k)$: cumulative discrete spectrum of the series $\hat{\sigma}(t)$; $k = 1, 2, \ldots$; see eqn 25.
$X(i, t)$: standardized values of $H(i, t)$:

$$X(i, t) = \frac{H(i, t) - \hat{\mu}(t)}{\hat{\sigma}(t)}$$

$Y(i, t)$: filtered values of $X(i, t)$; see eqn (8).
$Z(i, t)$: normalized values of $X(i, t)$; see eqn (11).
$U(i, t)$: filtered values of $Z(i, t)$; see eqn (16).
$\mu(t)$: trend in the mean of the time series $H(i, t)$.
$\hat{\mu}(t)$: computed values of $\mu(t)$; see eqn (2).
$\langle\mu\rangle$: overall mean value of $H(i, t)$.
ρ: first autocorrelation coefficient.
$\hat{\rho}$: computed value of ρ.
$\sigma(t)$: trend in the standard deviation of the time series $H(i, t)$.
$\hat{\sigma}(t)$: computed values of $\sigma(t)$; see eqn 3.
σ_H^2: overall variance of the time series $H(i, t)$.
$\langle\sigma\rangle$: arithmetic mean of the series $\hat{\sigma}(t)$; $t = 1, \ldots, 365$.
$\sigma^2(\hat{\mu}(t))$: computed variance of the series $\hat{\mu}(t)$; $t = 1, \ldots, 365$.
$\sigma^2(\hat{\sigma}(t))$: computed variance of the series $\hat{\sigma}(t)$; $t = 1, \ldots, 365$.
ω: basic period of the series $H(i, t)$ expressed as a discrete number; $\omega = 365$ (one year).

REFERENCES

1. B. Y. H. Liu and R. C. Jordan, The long term performances of flat plate solar collectors. *Solar Energy* **7**, 53 (1963).
2. S. A. Klein, Calculation of a flat plate collector utilisability. *Solar Energy* **21**, 393 (1978).
3. M. Collares-Pereira and A. Rabl, Simple procedure for predicting long term average performance of nonconcentrating and of concentrating solar collectors. *Solar Energy* **23**, 235 (1979).
4. S. A. Klein and W. Beckman, A general design method for closed-loop solar energy systems, *Solar Energy* **22**, 269 (1979).
5. P. Bendt, M. Collares-Pereira and A. Rabl, The frequency distribution of daily insolation values. *Solar Energy*, **27**, 1 (1981).
6. J. M. Gordon and M. Hochman, On the random nature of solar radiation. *Solar Energy* **32**, 337 (1984).
7. E. Boileau, Use of some simple statistical models in solar meteorology. *Solar Energy* **30**, 333 (1983).
8. M. D'Ambrosio, Identificazione dell'Anno di Riferimento dei valori di radiazione. Thesis (1979).
9. V. Cuomo, F. Fontana and C. Serio, Trombe walls and green-houses: an analytical approach to long term

performance analysis. *Rev. Phys. App.* **20,** 589 (1985).
10. F. Bloisi, S. Catalanotti, V. Cuomo, S. De Stefano and L. Vicari, Heat storage and solar system performances. *Appl. En.* **7,** 19 (1980).
11. B. Bartoli, U. Coscia, V. Cuomo, F. Fontana and V. Silvestrini, Statistical approach to long term performances of photovoltaic systems. *Rev. Phys. Appl.* **18,** 281 (1983).
12. B. Bartoli, S. Catalanotti, V. Cuomo, M. Francesca, C. Serio, V. Silvestrini and G. Troise, Statistical correlation between daily and monthly averages of solar radiation data. *Il Nuovo Cimento* **2C,** 222 (1979).
13. B. Bartoli, B. Coluzzi, V. Cuomo, M. Francesca and C. Serio, Autocorrelation of daily global solar radiation, *Il Nuovo Cimento* **4C,** 113 (1981).
14. U. Amato, A. Andretta, B. Bartoli, B. Coluzzi, V. Cuomo and C. Serio, Stochastic modeling of solar radiation data. Submitted to *Il Nuovo Cimento* (1984).
15. B. J. Brinkworth, Autocorrelation and stochastic modeling of sequences. *Solar Energy* **19,** 343 (1977).
16. A. Andretta, B. Bartoli, B. Coluzzi, V. Cuomo, M. Francesca and C. Serio, Global solar radiation estimation from relative sunshine hours in Italy. *J. Appl. Met.* **21,** 1377 (1982).
17. G. M. Jenkins and D. G. Watts, *Spectral Analysis and its applications*. Holden Day, San Francisco, U.S.A. (1968).
18. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*. Charles Griffin & Co. Ltd., London (1976).
19. C. Mennella, *Il Clima d'Italia*. E.D.A.R.T., Napoli (1967).
20. V. Yevjevich, *Stochastic Processes in Hydrology*. Water Resources Publications, Fort Collins, Colorado, U.S.A. (1972).
21. V. Cuomo, F. Fontana and C. Serio, Behaviour of ambient temperature on daily basis in Italian climate. *Rev. Phys. App.* **21,** 211 (1986).
22. H. Scheffè, *The Analysis of Variance*. J. Wiley & Sons, New York (1959).