PERGAMON

Available online at www.sciencedirect.com
SCIENCE DIRECT

RENEWABLE ENERGY

Renewable Energy 28 (2003) 1523–1544

www.elsevier.com/locate/renene

Methodology for generating daily clearness index index values K_t starting from the monthly average daily value $\bar{K}_t$. Determining the daily sequence using stochastic models

J.M. Santos *, J.M. Pinazo, J. Cañada

Department of Applied Thermodynamics, E.T.S.I. Industriales. Polytechnic University of Valencia, Camino de Vera s/n, 46022 Valencia, Spain

Received 23 September 2002; accepted 3 October 2002

Abstract

The facility to generate weather data from limited inputs and independently of specific locations would allow simulations of energetic systems to be run at locations for which detailed weather records do not exist. This article presents a methodology to calculate synthetic daily solar radiation values and describes how sequences of daily global radiation can be generated using as input the monthly average radiation. A stochastic model, ARIMA(1,1,1) is presented, as well.

Keywords: Daily clearness index; Monthly average daily clearness index; Stochastic models; Autocorrelation correlograms; Partial correlation correlograms; ARIMA; Time series

1. Introduction

Programs such as the TRNSYS [1] have been developed for the specific purpose of simulating thermal facilities (solar energy use, air-conditioning/heating in buildings, greenhouse design, etc.). Generally, these programs base their estimations on atmospheric data such as temperature, radiation, humidity, etc., taken at regular time intervals. It is difficult to get hold of very detailed records for the vast majority of

* Corresponding author.
E-mail address: josanna@uvigo.es (J.M. Santos).

0960-1481/03/$ - see front matter © 2003 Published by Elsevier Science Ltd.
doi:10.1016/S0960-1481(02)00217-3

Nomenclature

$\bar{K}_t$	Monthly average daily clearness index;
K_t	Daily clearness index;
k_t	Hourly clearness index;
$K_{t,min}$	Minimum daily clearness index;
$K_{t,max}$	Maximum daily clearness index;
F	Cumulative distribution function or fraction of days in which the daily clearness index is less than a certain specific value.;
nd_k	Number of the day of the month (1, 2, … ndm);
ndm	Number of days in the month (31, 30 or 28);
t	Number of the day;
K_{t-1}	Daily clearness index for the day before day t;
K_{t-2}	Daily clearness index for two days before day t;
a_t	Residue value on day t;
a_{t-1}	Residue value on the day before day t;Greek
δ	declination of the place (°);
γ	Parameter that defines the exponential distribution proved by Bouguer law of absorption of radiation through the atmosphere.;
λ	latitude of the place (°);
Φ_1	Parameter of the autoregressive part of the stochastic model;
θ_1	Parameter of the moving averages part of the stochastic model;
σ	Standard deviation of the stochastic model residues.

locations around the world, either because of a limitation in the number of years during which information has been available, or because of the questionable reliability of the information also, there are often periods for which this data has been lost (measuring systems are temporarily unavailable, data is damaged, etc.).

On the other hand, even assuming that this data exists, the records for at least an 11 year period (standard estimated cycle of atmospheric conditions) would be needed in order to be able to analyse the simulated behaviour of a facility.

When atmospheric data are not available in a certain area, two methods for estimating the aforementioned meterological variables are used: extrapolation and synthetic generation. Extrapolation involves the use of similar data from areas or climates with similar behaviour, to deduct the meterological variables of locations from which data cannot be obtained. The errors which arise when using this method are much greater than those associated with synthetic generation. Synthetic generation tries to develop a methodology to obtain, from a limited set of data, the different meterological variables of a certain area. On the whole, the data obtained will cover an annual cycle, and is known as a Typical Meterological Year (e.g. Petraski et al. [2]; Pissimanis et al. [3]).

Numerous researchers have generated models for obtaining series of synthetic data (either on a daily and/or hourly scale) (e.g., Aguiar et al. [4]; Aguiar et al. [5]; Amato et al. [6]; Graham et al. [7]; Graham and Hollands [8]; Gordon and Reddy [9]; Knight et al. [10]; Mora-López and Sidrach de Cardona [11]; Mora-López and Sidrach de Cardona [12]; Petraski et al. [2]; Pissimanis et al. [3]; Zeroual and Ankrim [13]). The general aim has been to express daily distributions and variations without depending on the location (Bendt et al. [5]; Hollands and Huget [12]; Liu and Jordan [15]).

The aim of this work is to generate a sequence of daily solar radiation data starting from the monthly average daily solar radiation value. Such a sequence of data should represent the behaviour of solar radiation in the area, with respect to the values observed, the monthly average value and its distribution (sequence of "good and bad" days.

This study is based on the dimensionless *clearness index* variable, defined as the quotient between the horizontal global solar radiation and the horizontal global extraterrestrial solar radiation, and which will be defined as a monthly character ($\bar{K}_t$), a daily character (K_t), an hourly character (k_t) or an instantaneous character.

2. Antecedents

As a general rule, we can say that the meterological variable solar radiation is neither completely random nor completely deterministic. This variable is highly random for short periods of time (days and hours), highly deterministic for long periods of time (months and years). Extraterrestrial solar radiation can be predicted for any place and time interval, since the area's atmospheric conditions determine the random status of solar irradiation at ground level.

Liu and Jordan [16] studied the statistical characteristics of solar radiation using atmospheric transmittance (also called clearness index) as a random variable. These researchers showed how the daily clearness index was related to the monthly average daily clearness index. Bendt et al. [14], Hollands and Huget [15] subsequently developed an expression for Liu and Jordan's distributions.

Bendt et al. [14] proposed a frequency distribution of daily clearness index values (K_t) starting from the monthly average daily clearness index ($\bar{K}_t$); in order to do this, they studied solar radiation in 90 locations in the USA over a period of approximately 20 years. The basic starting point was that the daily clearness indexes (K_t) had an exponential distribution throughout the month ranging between the minimum and maximum values recorded ($K_{t\ min}$ $K_{t\ max}$). The occurrence frequency curves (cumulative distribution function) F, or the fraction of days in which the daily clearness index is less than a certain given specific value K_t, that is, $F{=}nd_k/ndm$) were obtained using the expression:

$$F[K_t,\ \bar{K}_t] \quad = \quad \frac{\exp(\gamma\ K_{t,\ min})-\exp(\gamma K_t)}{\exp(\gamma\ K_{t,\ min})-\exp(\gamma K_t, max)} \tag{1}$$

where γ is the value that defines the particular exponential distribution.

The basic idea leading to Eq. (1) can be argued from Bouger-Lambert's exponential law of radiation absorption when passing through a homogeneous medium, which would produce an exponential distribution of (K_t) if the extinction coefficient of the medium is considered (the sum of factors that absorb solar radiation, clouds, aerosols, the presence of gases …) to be a variable with a linear distribution between a maximum value (different from $+\infty$) and a minimum one (different from 0). However, Bouger-Lambert's law can only be applied to direct radiation through an uniform medium with radiant properties independent from the wavelength, although variable from day to day.

The value of γ, which defines the particular exponential distribution is easily determined if the minimum value $K_{t\ min}$, the maximum value $K_{t\ max}$ and the average monthly value $\bar{K}_t$ are known (in a continuous exponential distribution, the minimum and maximum value would be 0 and 1 respectively, although in our case, there will be a minimum value greater than zero, "worst day", and a maximum value less than one, "best day". With these limits, the average value of a continuous exponential distribution is obtained using the expression:

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

from which the value of γ can be obtained.

Alternatively Herzog [17] gave an explicit relation for obtaining the value of γ in this way:

$$\gamma = -1.498 + \frac{1.184\frac{K_{t,\ min}-K_{t,\ max}}{K_{t,\ min}-\bar{K}_t}-27.182\cdot\exp\left(-1.5\frac{K_{t,\ min}-K_{t,\ max}}{K_{t,\ min}-\bar{K}_t}\right)}{K_{t,\ min}-K_{t,\ max}} \tag{3}$$

Bendt et al. [5] recommended the value 0.05 for $K_{t,min}$, which was assumed to be constant and independent of locality, but they did not provide an expression for the estimation of $K_{t,max}$ (a maximum value of 0.864 was assumed).

Continuing Bendt's work, Hollands and Huget [15] correlated, for the areas studied, the value of the maximum monthly clearness index $K_{t,max}$ with the monthly average clearness index $\bar{K}_t$ using:

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

They obtained the exponential distribution of the daily clearness index from the value of his monthly average $\bar{K}_t$ was consequently obtained.

Recently Abdulla et al. [18] have proposed a new correlation to obtain the maximum clearness index:

$$K_{t,\ max} = 0.51585 + 0.34847\bar{K}_t + 2.302810^{-4}\ \delta + 3.410810^{-4}\lambda - 9.570910^{-6}z \tag{5}$$

here, the solar declination (δ), the latitude of the place (γ) and the altitude (z) of the area, allow for a more accurate index value for each month considered.

Knight et al. [10] implemented a specific procedure to obtain the different daily clearness index values (K_t) over the period of a month. In this case they assumed that the most probable value of each clearness index occurs in the middle of the interval between each two cumulative fractions, therefore, from Eq. (1) they obtain:

$$K_t = \frac{\ln\left[\left(1-\frac{nd_k-{}^1/_2}{ndm}\right)\exp(\gamma K_{t,\ min}) + \frac{nd_k-{}^1/_2}{ndm}\exp(\gamma K_{t,\ max})\right]}{\gamma} \tag{6}$$

in which:$nd_k = 1, 2, \ldots\ ndm$

Applying this procedure, Bendt's cumulative distribution function for different monthly average clearness index values $\bar{K}_t$ (assuming a month of 31 days) is shown in Fig. 1. From the aforementioned graph and reading it in an alternative way, we can obtain the 31 daily clearness index values corresponding to the 31 days of the month. However, the sequence of days in which these succeed each other is not known, and obviously these do not follow an ascending or descending order, but rather they present a random occurrence sequence.

Knight et al. [10] and Graham et al. [7,8] apply this methodology in order to obtain the 31 daily clearness indexes which succeed each other in a month (with 31 days) and they propose a particular sequence to organise clearness indexes, (a function simultaneously expresses the monthly average daily clearness index value), shown in Table 1. This technique is currently used to generate typical years in simulation programs such as the TRNSYS [1]. This form of determining the values will be called model 1. A priori establishing a particular pattern for a sequence of days

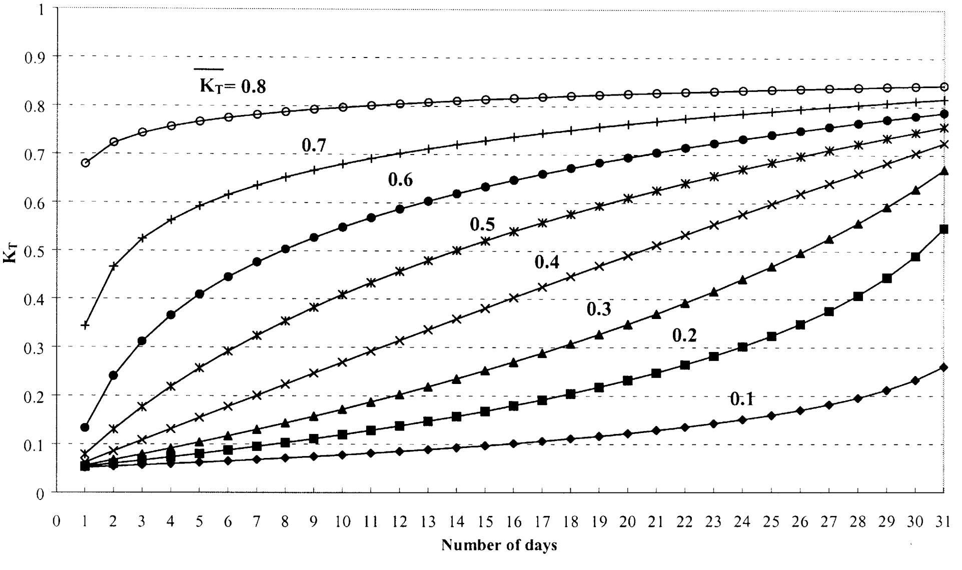

Fig. 1. Cumulative distribution of the clearness index as a function of the $\bar{K}_t$ for the Correlation of Bendt, et al.

Table 1
Sequence of values of the daily clearness indexes K_t

≤ 0.45	24, 28, 11, 19, 18, 3, 2, 4, 9, 20, 14, 23, 8, 16, 21, 26, 15, 10, 22, 17, 5, 1, 6, 29, 12, 7, 31, 30, 27, 13, 25
$0.45<\bar{K}_t<0.55$	24, 27, 11, 19, 18, 3, 2, 4, 9, 20, 14, 23, 8, 16, 21, 7, 22, 10, 28, 6, 5, 1, 26, 29, 12, 17, 31, 30, 15, 13, 25
$055\leq\bar{K}_t$	24, 27, 11, 4, 18, 3, 2, 19, 9, 25, 14, 23, 8, 16, 21, 26, 22, 10, 15, 17, 5, 1, 6, 29, 12, 7, 31, 20, 28, 13, 30

with no statistical procedure, does not seem to be coherent, or susceptible to be applied to any location.

Graham et al. [7] presented an alternative modelling based on the time series analysis of the daily clearness index values, ARMA (1,0)

$$\chi(t) = 0.29(t - 1) + \tilde{w}_t \quad (7)$$

with $\tilde{w}(t)$ white noise and incorporating an acumulative probability distribution function for K_t and $\chi(t)$ the Gauss Series is

$$F[K_t, \bar{K}_t] = G(\chi(t)) = \left[\frac{1}{2} + \frac{1}{2}erf\left(\frac{0.29\chi(t - 1) + \tilde{w}(t)}{\sqrt{2}}\right)\right] \quad (8)$$

This method for obtaining the values of the daily clearness index will be called model 2. Gordon et al. [9] set out to solve the problem using time series (ARMA), and directly generating the daily clearness index variable. Some researchers subsequently suggested directly generating the hourly clearness index variable; (Gordon and Reddy [19], Mora-López and Sidrach de Cardona [12]). In these cases, the clearness index is obtained from a reminder of the aforementioned variable at a previous moment or moments (days), plus a random value (normal variable) and even a reminder of the random value used when calculating the clearness indexes at previous moments. Generally, this procedure produces sequences (succession of “good and bad” days) which are reasonably acceptable (at least for the location studied), although the particular daily clearness index value can become absurd (values lower than $K_{t,min}$ or higher than $K_{t,max}$) due to the random component used. Equally, this procedure produces mean values for the monthly values clearness index that do not correspond to the average observed $\bar{K}_t$. Therefore the calculation process weights each daily clearness index obtained in order to match the mean values of the average monthly clearness index (observed and calculated).

Aguiar et al. [5] and Amato et al. [6] propose generating the daily sequence with Markov’s Processes, using transition matrixes. The procedure consists of assigning a probability to that which happened the day before, in order to carry out the transition from one value to the other. In this study the yearly sequence of days is used.

The methodology proposed in this article, model 3, tries to combine the best features of all these works, determining the daily clearness index values using Bendt’s procedure and obtaining the sequence of days by means of time series. Next, the application is implemented on a specific location, Valencia (Spain), in order to demonstrate its accuracy.

3. Data base

It should be pointed out that having an adequate series of data (several years) of hourly horizontal global solar radiation, (with the purpose of evaluating the daily solar radiation, and therefore the daily clearness index value, with a solar constant of 1367 W·m^{-2}), with systematically calibrated, standard measuring equipment, proves to be very problematic.

The solar radiation database used was obtained from radiation data from Valencia and Madrid. The solar radiation database for Valencia was obtained from the Radiometric and Meteorological Laboratory of the Department of Applied Thermodynamics, the Polytechnic University of Valencia. The station is located near the Mediterranean Sea (approximately 500 m) at 39.48° latitude north. The horizontal global solar irradiance was measured with an Eppley 8.48 pyranometer. The values of hourly irradiation (kJ·m^{-2}) and daily irradiation (MJ·m^{-2}) are calculated from the irradiance data measured every minute and stored on disk every 5 min.

We have been working with an uninterrupted (38,325 inputs) hourly database of horizontal solar radiation recorded from April 1991 to March 1998, both inclusive. For Madrid we have been working with an uninterrupted (43,800 inputs) hourly database recorded from May 1977 to April 1985, both inclusive (Fig. 2).

The hourly clearness index values (k_t) are estimated from the values stored, and the daily (K_t) and monthly $\bar{K}_t$ average daily clearness indexes are obtained from (K_t).

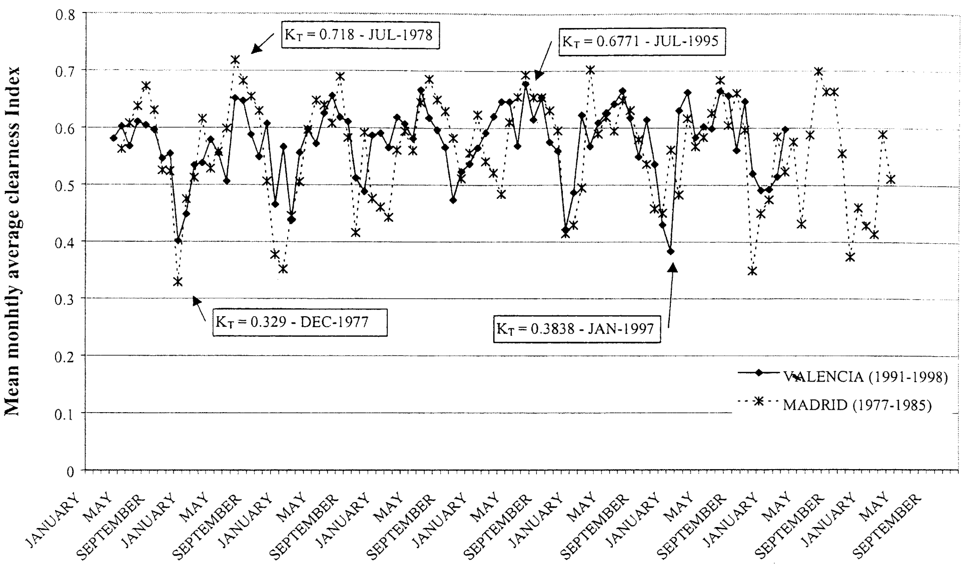

Fig. 2. Value of monthly-average daily clearness index for Valencia and Madrid.

4. Methodology used

The proposed validation is subdivided in two sections: 1. validating Bendt et al. [14] model, improved by Hollands et al. [15], for the data measured in Valencia, 2. generating a specific proposal based on time series to define the sequencing of the daily clearness indexes obtained.

4.1. Study of Bendt's correlation

In Fig. 3, and as an example, the measured and calculated clearness indexes corresponding to the month of March 1998 for Valencia are illustrated. The behaviour for the remaining months of the study and Madrid is similar. As can be gathered from observing Fig. 3, Bendt's Correlation predicts, to a large degree, the evolution of the daily clearness index values, K_t.

With the purpose of determining the good fit among the measured and calculated values, the Mean Bias Error (M.B.E.) and the Root Square Mean Error (R.S.M.E.) for the whole period analysed have been estimated, resulting in M.B.E. = −0.0111% and R.S.M.E. = 5.4%.

In order to detect ill-fitting periods (certain months), and possible compensations (positive and negative discrepancies), the mean error has been calculated month by month, M.A.P.E., is defined as is given in Appendix A.

In Fig. 4 estimated monthly values of M.A.P.E of all years and cities are shown. 8.71% is the mean value observed. In Table 2, a mean value of error found in the whole study is illustrated. From this, we can conclude that for Valencia and Madrid,

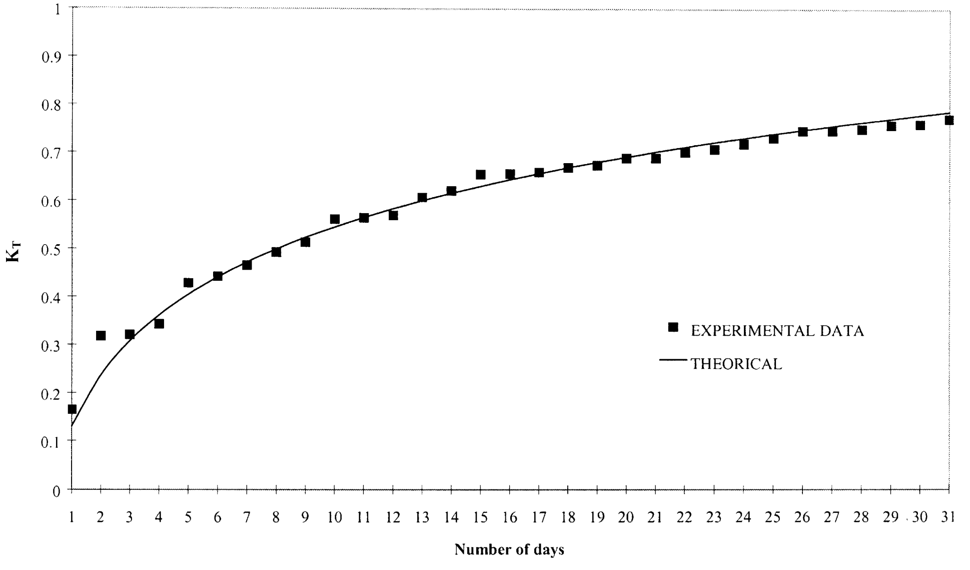

Fig. 3. Correlation of Bendt et al. for March 1998 and measured data for Valencia.

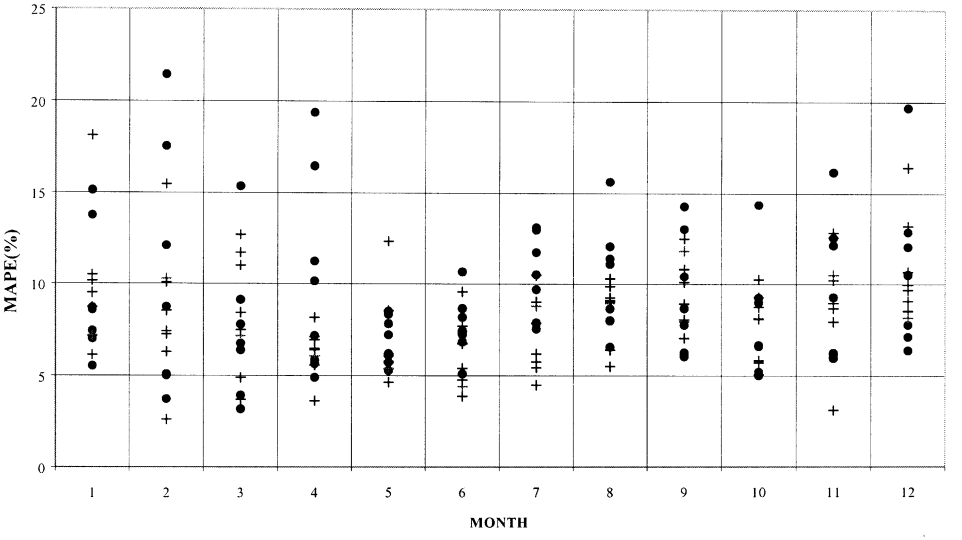

Fig. 4. M.A.P.E. estimated monthly values for all years. (● Valencia. + Madrid).

Table 2
Mean Value of Error. Mean Bias Error, Root Square Mean Error and Mean Absolute Percentage Error

City	Mean bias error	Root square mean error	Mean absolute percentage error
Valencia	−0.0001	0.001	9.152
Madrid	0.0002	0.0009	8.257

in the period considered, there is no significant difference between the values of the observed (model 1) and expected (measured values) daily clearness indices.

4.2. Study of the sequence of clearness index values

The analysis of these data was carried out using stochastic models that is, Time Series and in particular the ARIMA model (Box-Jenkins [20], Uriel [21]) and to do so the software package STATGRAPHICS® v4.0 was used. It should also be emphasized that this technique does not allow for missing days or moments and consequently, a complete series has been used for this study.

To determine the most suitable ARIMA model, two procedures were used, one using the study of Autocorrelation / Partial Correlation Correlograms and the other using different tests (amongst which the importance of the Box-Pierce Test and the Ljung-Box Test should be emphasized) (Box and Jenkins [20] and Uriel [21]).

First of all, in Fig. 5a and b, the Correlograms for the year 1998 in the city of Valencia (other years being similar) can be seen. Similar correlograms were also obtained for Madrid. If we observe this figure, the Autocorrelation Coefficient for the first lag is significant with a decay process for the Partial Correlation Coefficients. This form of evolution can be modeled with an ARIMA(0,1,1) time series model.

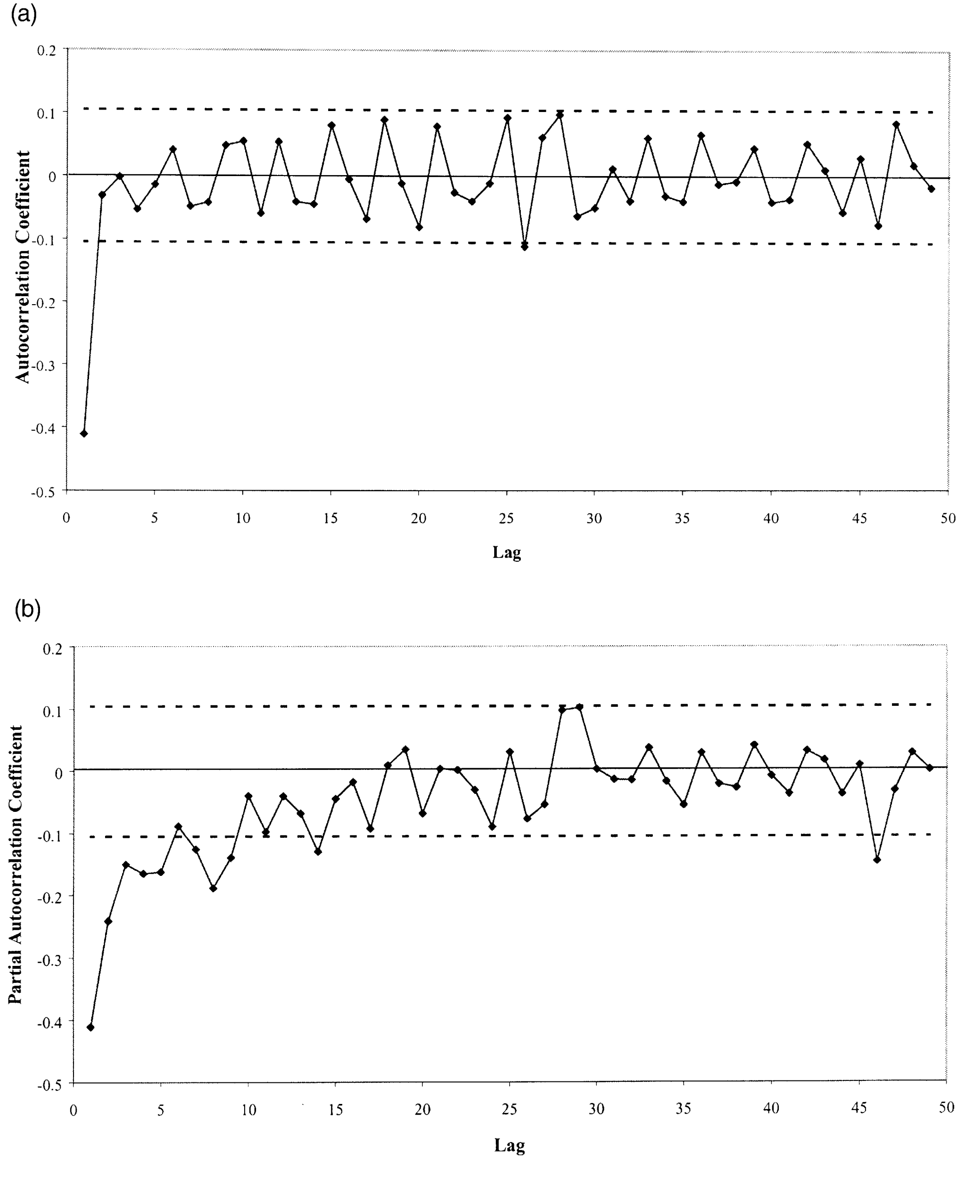

Fig. 5. Correlograms of Autocorrelation (ACF) (a) and Partial Correlation (PACF) (b) for the year 1998 for Valencia.

Numerous models were considered for determining the random sequence of days depending on the intervals used, since the random model can be based on a percentage remembered from what happened the day or days before together with a random component (and even a reminder of the random component used on previous days). In this respect, it is appropriate to underline the importance of the ARIMA(0,1,1), ARIMA(0.1,2), ARIMA(1,1,0), ARIMA(2,1,0) and the ARIMA(1,1,1).

In Table 3, we can see the statistical parameters of the residuals of the models studied, as well as the *P*-value of Box-Pierce Test, which would allow eliminating

Table 3
Statistical parameters of the residuals of the different studied stochastic models. (∗ Model presents some parameter that allows to eliminate the pattern)

Year	Sthocastic model	Mean	Standard deviation	*P*-value Box-Pierce test
1991	ARIMA(0,1,1)	−0.0005	0.155	0.336
	ARIMA(0,1,2)	−0.0003	0.153	0.378
	ARIMA(1,1,0)∗	−0.0003	0.164	0.0055
	ARIMA(2,1,0)	−0.0003	0.158	0.114
	ARIMA(1,1,1)	−0.0005	0.15	0.306
1992	ARIMA(0,1,1)	0.0014	0.156	0.766
	ARIMA(0,1,2)	0.0015	0.144	0.783
	ARIMA(1,1,0)	0.0004	0.154	0.115
	ARIMA(2,1,0)	0.0007	0.148	0.784
	ARIMA(1,1,1)	0.0015	0.144	0.807
1993	ARIMA(0,1,1)	0.0003	0.151	0.228
	ARIMA(0,1,2)	−0.0029	0.151	0.541
	ARIMA(1,1,0)∗	−0.00036	0.178	0.0021
	ARIMA(2,1,0)∗	−0.00027	0.168	0.0358
	ARIMA(1,1,1)	−0.0019	0.15	0.663
1994	ARIMA(0,1,1)	0.0014	0.163	0.339
	ARIMA(0,1,2)	−0.0022	0.16	0.932
	ARIMA(1,1,0)∗	0.0002	0.188	0.00007
	ARIMA(2,1,0)∗	0.0003	0.18	0.00298
	ARIMA(1,1,1)	0.0005	0.16	0.915
1995	ARIMA(0,1,1)	−0.0026	0.157	0.21
	ARIMA(0,1,2)	−0.0031	0.156	0.336
	ARIMA(1,1,0)∗	−0.0008	0.178	0.000023
	ARIMA(2,1,0)∗	−0.001	0.172	0.000481
	ARIMA(1,1,1)	−0.0034	0.156	0.377
1996	ARIMA(0,1,1)	0.0003	0.158	0.6376
	ARIMA(0,1,2)	0.0007	0.154	0.414
	ARIMA(1,1,0)∗	0.0002	0.169	0.0108
	ARIMA(2,1,0)∗	0.0001	0.166	0.00784
	ARIMA(1,1,1)	0.0018	0.153	0.701
1997	ARIMA(0,1,1)	−0.0024	0.169	0.0711
	ARIMA(0,1,2)	−0.0017	0.161	0.886
	ARIMA(1,1,0)	−0.0009	0.179	0.1089
	ARIMA(2,1,0)	−0.0013	0.174	0.347
	ARIMA(1,1,1)	−0.0065	0.159	0.523

all models not significant for this study. All ARIMA models with a *P*-value lower than 5% are not significant and they can be discarded: ARIMA(2,1,0) and ARIMA(1,1,0). ARIMA(1,1,1) is the most suitable model, since it is the model with a larger Box-Pierce *P*-value and a lower error mean and it shows little or no residue coefficient correlation. In the light of the results, we can conclude that the ARIMA(1,1,1) model can be accepted.

The ARIMA(1,1,1) model establishes that the daily clearness index value (K_t) depends on the value established for the day before (K_{t-1}), with a certain reminder (Φ_1) of the difference between values K_{t-1} and K_{t-2}, plus a normal random variable (a_t) and its reminder (θ_1) of the moment before. This random variable (a_t) gives a specific mean (m) and standard deviation (σ).

$$K_t = K_{t-1} + \phi_1 (K_{t-1} - K_{t-2}) + a_t - \theta_1 a_{t-1} \tag{10}$$

In Table 4, we can find the coefficients (Φ_1), (θ_1) and (σ) for Valencia and Madrid. Erasing all the extreme values, we can conclude that all these coefficients can be summarised for all the years (Table 5), rendering the mean coefficient for each city. But in order to obtain a mean year, we can consider a mean value 0.2955 for (Φ_1), 0.9305 for (θ_1), the random variable being of null mean and standard deviation σ = 0.151.

5. Comparing models for Valencia

In Table 6 we compare the results of models 1, 2 and 3, as obtainec for Valencia using data for the period 1998–1999.

In the proposed model 3, the sequence of daily clearness index values is determined according to the ARIMA(1,1,1) model, Eq. (10), using the monthly average clearness index value as a starting point. For this, a normal random series of numbers of null mean and standard deviation $\sigma = 0.151$, (series a_t) are generated, and together with parameters Φ_1 (0.2955) and θ_1 (0.9305) and using expression (10), they produce a sequence of daily clearness index values. These indexes are arranged in ascending order and the resulting sequence of days is taken. Then, making use of Bendt's Correlation, the daily clearness index values are determined and finally organised with the sequence of days produced by the ARIMA (1,1,1) model.

The monthly average clearness index values that result from implementing each model in question can be seen in Table 6. It is clear that model 1, model 3 and the measured values present almost identical values, differing considerably from those obtained for model 2.

In Fig. 6, the organised clearness indexes for May 98 are represented, (those for model 3 and 1 are the same). In Table 7, the annual error values, M.B.E. (Mear Bias Error), R.S.M.E. (Root Square Mean Error), M.A.P.E. (Mean Absolute Percentage Error), M.P.E. (Mean Percentage Error) and R.S.P.E. (Root Square Percentage Error) are illustrated.

We can statistically show by means of the *P*-value of the Box-Pierce Test, if each

Table 4
Coefficients of ARIMA(1, 1, 1) for Valencia and Madrid

Year	VALENCIA—ARIMA(1,1,1)		Year	MADRID—ARIMA(1,1,1)	
1991	$\Phi_1 = 0.452 \pm 2{\cdot}0.05$	$\sigma = 0.15$	1978	$\Phi_1 = 0.333 \pm 2{\cdot}0.0538$	$\sigma = 0.153$
	$\theta_1 = 0.976 \pm 2{\cdot}0.014$			$\theta_1 = 0.945 \pm 2{\cdot}0.0155$	
1992	$\Phi_1 = 0.148 \pm 2{\cdot}0.07$	$\sigma = 0.144$	1979	$\Phi_1 = 0.335 \pm 2{\cdot}0.0595$	$\sigma = 0.151$
	$\theta_1 = 0.8 \pm 2{\cdot}0.045$			$\theta_1 = 0.918 \pm 2{\cdot}0.0269$	
1993	$\Phi_1 = 0.151 \pm 2{\cdot}0.052$	$\sigma = 0.15$	1980	$\Phi_1 = 0.322 \pm 2{\cdot}0.054$	$\sigma = 0.147$
	$\theta_1 = 0.984 \pm 2{\cdot}0.006$			$\theta_1 = 0.959 \pm 2{\cdot}0.0175$	
1994	$\Phi_1 = 0.235 \pm 2{\cdot}0.053$	$\sigma = 0.159$	1981	$\Phi_1 = 0.26 \pm 2{\cdot}0.057$	$\sigma = 0.14$
	$\theta_1 = 0.974 \pm 2{\cdot}0.0106$			$\theta_1 = 0.927 \pm 2{\cdot}0.0227$	
1995	$\Phi_1 = 0.157 \pm 2{\cdot}10.0565$	$\sigma = 0.156$	1982	$\Phi_1 = 0.36 \pm 2{\cdot}0.0557$	$\sigma = 0.144$
	$\theta_1 = 0.935 \pm 2{\cdot}0.02$			$\theta_1 = 0.934 \pm 2{\cdot}0.0223$	
1996	$\Phi_1 = 0.325 \pm 2{\cdot}0.057$	$\sigma = 0.153$	1983	$\Phi_1 = 0.141 \pm 2{\cdot}0.0705$	$\sigma = 0.124$
	$\theta_1 = 0.929 \pm 2{\cdot}0.022$			$\theta_1 = 0.803 \pm 2{\cdot}0.044$	
1997	$\Phi_1 = 0.406 \pm 2{\cdot}0.0493$	$\sigma = 0.16$	1984	$\Phi_1 = 0.256 \pm 2{\cdot}0.0595$	$\sigma = 0.155$
	$\theta_1 = 0.982 \pm 2{\cdot}0.0094$			$\theta_1 = 0.907 \pm 0{\cdot}0265$	

Table 5
Coefficients of the model ARIMA(1, 1, 1) in an average year for locations of Valencia and Madrid

ARIMA(p,d,q) model	Parameters	Estimated parameters	Standard deviation of white noise σ_a
Valencia			
ARIMA(1,1,1): $K_t = K_{t-1} + \Phi_1 (K_{t-1} - K_{t-2}) + a_t - \theta_1 a_{t-1}$	Φ_1	0.28	0.154
	θ_1	0.946	
Madrid			
ARIMA(1,1,1): $K_t = K_{t-1} + \Phi_1 (K_{t-1} - K_{t-2}) + a_t - \theta_1 \alpha_{t-1}$	Φ_1	0.311	0.148
	θ_1	0.915	

Table 6
Value of the monthly-average daily clearness index for Year 98–99 resultant of applying the different models

	January	February	March	April	May	June	July	August	September	October	November	December
Measured	0.5764	0.6096	0.60936	0.6284	0.54672	0.6518	0.6823	0.64936	0.5809	0.6554	0.5736	0.5594
Model 1 and 3	**0.57585**	**0.6085**	**0.6083**	**0.6273**	**0.5472**	**0.6515**	**0.68384**	**0.64892**	**0.5803**	**0.65525**	**0.5732**	**0.55945**
Model 2	**0.5531**	**0.4918**	**0.6078**	**0.5776**	**0.5327**	**0.66221**	**0.67833**	**0.68552**	**0.4671**	**0.58057**	**0.52738**	**0.50206**

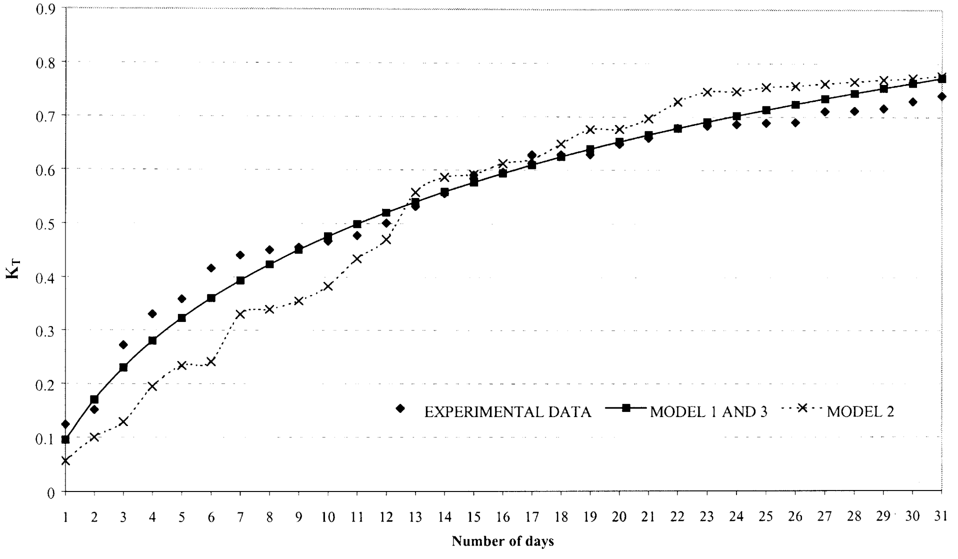

Fig. 6. Correlation of Bendt et al. for May 1998. Comparison of models 1, 2 and 3, and measured data for Valencia.

Table 7
Annual errors of the different models

	Models 1 and 3	Model 2
M.B.E.	−0.0003	−0.038
R.S.M.E.	0.0092	0.0189
M.A.P.E.	9.89%	16.03%
M.P.E.	2.14%	−8.9%
R.S.P.E.	1.6%	2.63%

model fits to the generated data through the models raised and the measured values, we can see that model 3 is the only one with an accurate fit. (Table 8)

From this comparison, we can deduce that the daily clearness index values produced by model 3 and by model 1 come closer to reality than those produced by

Table 8
Statistic *P*-value of Box-Pierce Test for the validation

Experimental Data	0.721
Model 1	0.0
Model 2	$6.217.10^{-24}$
Model 3	0.1463

model 2. In Fig. 7a and b, the Autocorrelation and Partial Correlation Correlograms of model 3 are represented, Fig. 8a and b are those produced by model 1, and Fig. 9a and b, are those produced by model 2. After a straightforward comparison, we can see that the model 2 and model 3 follow the measured data (similar values) and

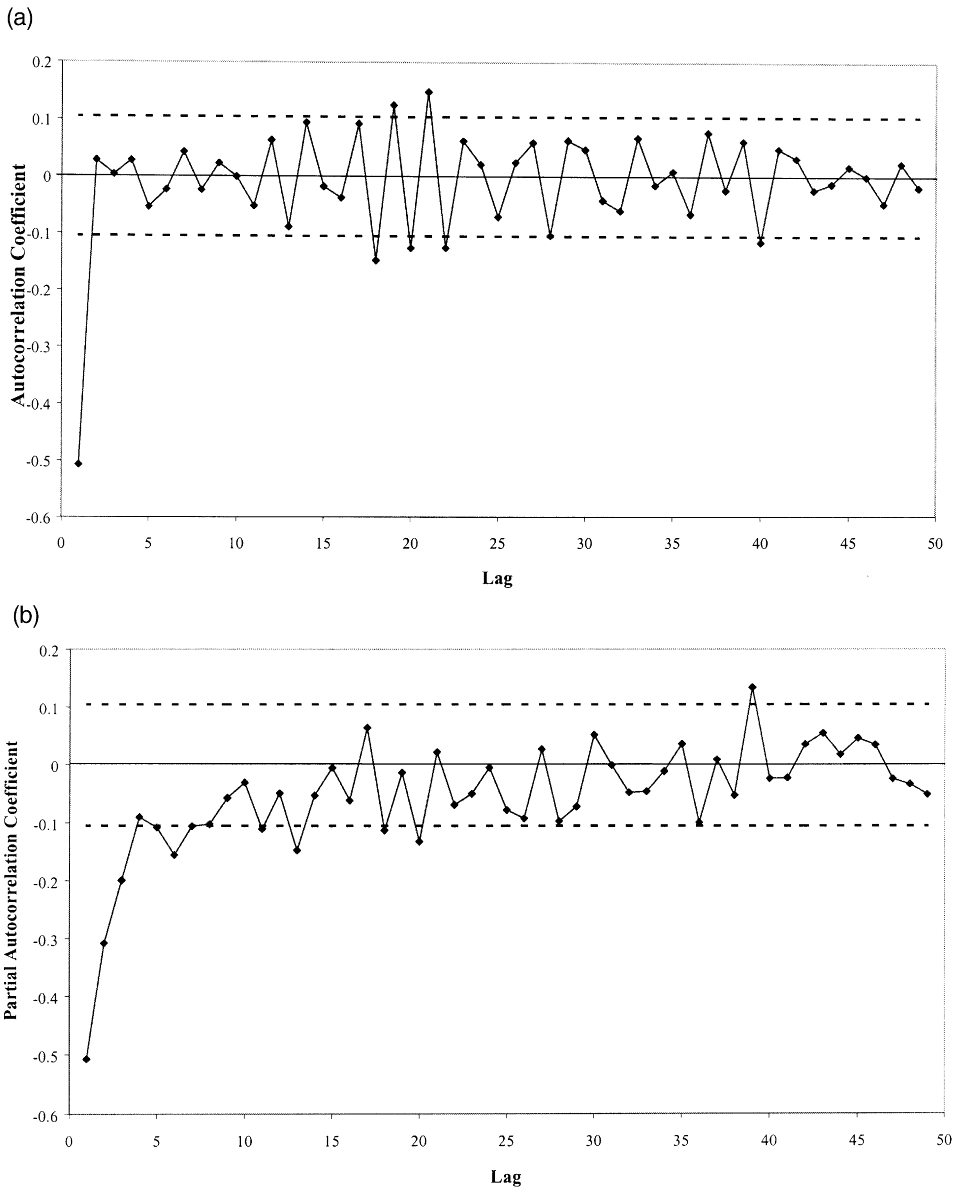

Fig. 7. Correlograms ACF (a) and PACF (b) for the period of validation year 1998–1999 according to model 3.

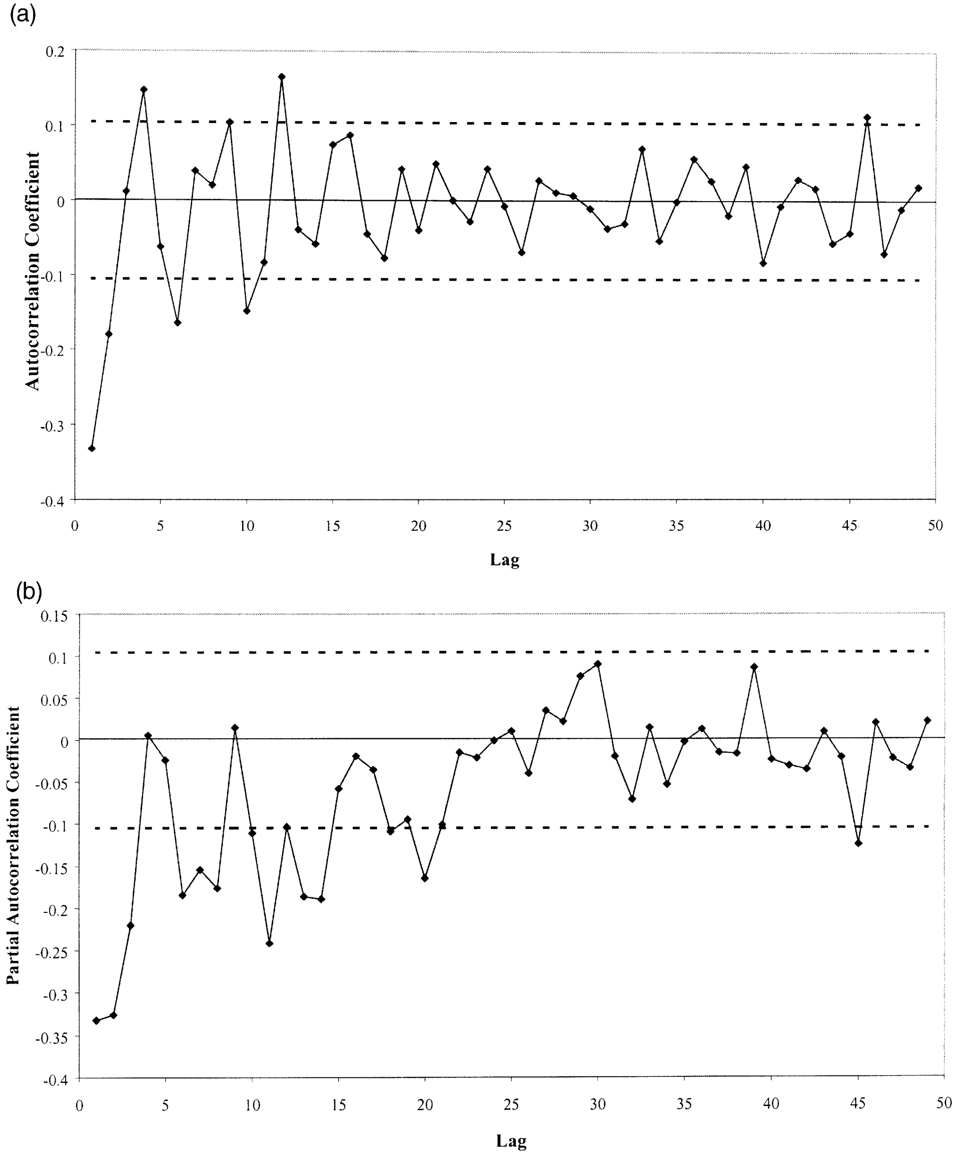

Fig. 8. Correlograms ACF (a) and PACF (b) for the period of validation year 1998–1999 according to model 1.

it is model 1 (used by TRNSYS) the one deviating from it (this is clearly due to the fixed establishment of the daily sequence). Therefore, the proposed method, model 3, is more efficient when it comes to generating the daily clearness index.

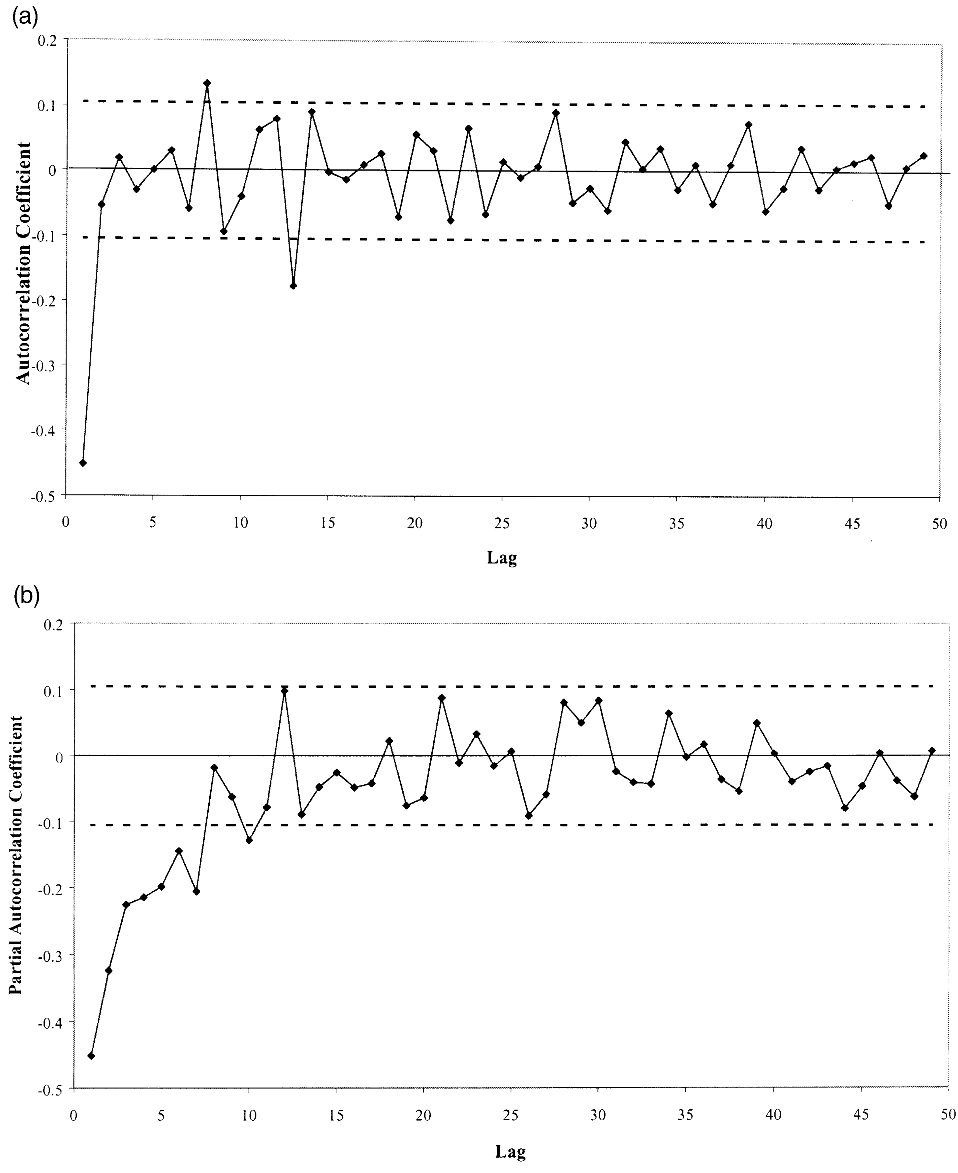

Fig. 9. Correlograms ACF (a) and PACF (b) for the period of validation year 1998–1999 according to model 2.

6. Conclusion

A methodology has been developed to obtain synthetic data from the daily clearness index (K_t) starting from the monthly average daily clearness index ($\bar{K}_t$). The methodology is based on obtaining the clearness index values and the sequence of days using Bendt's correlation an ARIMA(1,1,1) Time Series model, respectively.

The parameters that define the ARIMA(1,1,1) model, have been obtained for Valencia and Madrid, together with the combined correlation used in both locations. The validity of this model has been tested with data measured outside the period of time used for generating the model. The precision of the proposed method, once contrasted with the existing ones, has been proved to be higher. Finally, it should be pointed out that our study detected several Time Series models for the random sequence of days, such as ARIMA (0,1,1) and ARIMA (0,1,2) models, which could be equally valid (although less accurate).

Appendix A: Calculating errors

The equations of errors used in the analysis of the data were:

Mean Bias Error $M.B.E. = \frac{1}{DAYS}\Sigma_1^{DAYS}(K_{t_{OBSERVED}} - K_{t_{EXPECTED}})$

Root Square Mean Error $R.S.M.E. = \sqrt{\frac{1}{DAYS}\Sigma_1^{DAYS}(K_{t_{THEORICAL}} - K_{t_{REAL}})^2}$

Mean Percentage Error

$$M.P.E. = \frac{1}{DAYS}\Sigma_1^{DAYS}\left(\frac{K_{t_{OBSERVED}} - K_{t_{EXPECTED}}}{K_{t_{EXPECTED}}}\right) X\ 100$$

Root Square Percentage Error

$$R.S.P.E. = \sqrt{\frac{1}{DAYS}\frac{\Sigma_1^{DAYS}(K_{t_{OBSERVED}} - K_{t_{EXPECTED}})^2}{K_{t_{EXPECTED}}}}\ X\ 100$$

Mean Absolute Percentage Error

$$M.A.P.E. = \frac{1}{DAYS}\Sigma_1^{DAYS}\left(\frac{ABS(K_{t_{OBSERVED}} - K_{t_{EXPECTED}})}{K_{t_{EXPECTED}}}\right) X\ 100$$

Appendix B: Validating the model

Several studies have contributed to the validation of the ARIMA model (p,d,q) using the Autocorrelation and Partial Correlation Correlograms of the residues and so have several Tests amongst which the Box-Pierce and the Ljung-Box Tests stand out.

Firstly in order to see whether a random ARIMA model (p,d,q) is suitable, the following requirements should be fulfilled:Residues from the estimated model close to white noise behaviour:

- Null mean
- Constant variance

- Stationary, invertible model.
- Statistically significant correlation coefficients and moving average.
- The coefficients of the model are only slightly correlated with each other.
- The degree of precision is high in comparison to other alternative models.

The Box-Pierce coefficient consists of applying the formula:

$$Q = M \sum_{i=1}^{K} r_i^2$$

Another equally valid coefficient is the Ljung-Box coefficient:

$$Q* = M(M+2) \sum_{i=1}^{K} \frac{r_i^2}{M-i}$$

where M is the sample size, r_i is the correlation coefficient of the residues for day i. If $Q* < \chi^2_{M-p-q}$ is true the independence of the residues hypothesis is accepted.If, on the contrary, $Q* > \chi^2_{M-p-q}$ this initial hypothesis is rejected.

Using computers, hypothesis testing can be completed in a far more rational way. Some programs offer, along with the test statistic, the critical level of significance, CLS ($\alpha*$), associated with the test statistic:

$$\alpha* = \mathrm{Pr}ob\ [\chi^2 > Q*]$$

Once $\alpha*$ has been determined we know that the hypothesis is rejected at all α levels of significance when $\alpha > \alpha*$; on the contrary, the hypothesis is accepted when $\alpha < \alpha*$. The CLS is an indicator of the level of admissibility of the null hypothesis. The greater the CLS, the more confidence we can place in the null hypothesis.

In hypothesis testing a level of significance α is generally fixed in advance—for example, $\alpha = 5\%$—on the basis of which the value of the variable is found in the theoretical distribution tables of the statistic used, taking the degrees of freedom that correspond. So in this particular case, designated by $\chi^2_{M-p-q}(\alpha)$ to the value χ^2 of the tables for an α level of significance and M-p-q degrees of freedom.

References

[1] Klein SA, Beckman WA. TRNSYS v14.2 (1996). A transient simulation program.

[2] Petrakis M, Kambezidis HD, Lykoudis S, Adamopoulos AD, Kassomenos P, Michaelides IM et al. Generation of a “typical meteorological year” for Nicosia, Cyprus. Renewable Energy 2000;13:381–8.

[3] Pissimanis D, Karras G, Notaridou V, Gavra K. The generation of a “typical meteorological year” for the city of Athens. Solar Energy 2000;40:405–11.

[4] Aguiar RJ, Collares-Pereira M, Conde JP. Simple procedure for generating sequences of daily radiation values using a library of Markov transition matrices. Solar Energy 1988;40:269–79.

[5] Aguiar RJ, Collares-Pereira M. TAG: a Time-dependent, Autoregressive, Gaussian model for generating synthetic hourly radiation. Solar Energy 1992;49:167–74.

[6] Amato U, Andretta A, Bartoli B, Coluzzi B, Cuomo V. Markov Processes and Fourier analysis as a tool to describe and simulate daily solar irradiance. Solar Energy 1986;37:179–94.
[7] Graham VA, Hollands KGT, Unny TE. A time series model for K_t with application to global synthetic weather generation. Solar Energy 1988;40:83–92.
[8] Graham VA, Hollands KGT. A method to generate synthetic hourly solar radiation globally. Solar Energy 1990;44:333–41.
[9] Gordon JM, Reddy TA. Time Series analysis of daily horizontal solar radiation. Solar Energy 1988;41:215–26.
[10] Knight KM, Klein SA, Duffie JA. A methodology for the synthesis of hourly weather data. Solar Energy 1991;46:109–20.
[11] Mora-López LL, Sidrach de Cardona M. Characterization and simulation of hourly exposure series of global radiation. Solar Energy 1997;60:257–70.
[12] Mora-López LL, Sidrach de Cardona M. Multiplicative ARMA models to generate hourly series of global irradiation. Solar Energy 1998;63:283–91.
[13] Zeroual A, Ankrim M. Stochastic modelling of daily global solar radiation measured in Marrakech, Morocco. Renewable Energy 1995;6:787–93.
[14] Bendt P, Collares-Pereira M, Rabl A. The frequency distribution of daily insolation values. Solar Energy 1981;27:1–5.
[15] Hollands KG, Huget RG. A probability density function for the clearness index, with applications. Solar Energy 1983;30:195–209.
[16] Liu BY, Jordan RC. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. Solar Energy 1960;4:1–19.
[17] Herzog ME. Thesis, Estimation of hourly and monthly average daily insolation on tilted surfaces, Trinity University, 1985.
[18] Abdulla SH, Klein SA, Beckman WA. A new correlation for prediction of the frequency distribution of daily solar radiation. American Solar Energy Society (2000). "http://sel.me.wisc.edu/INFO/Publications/selpub.htm#meetings"
[19] Gordon JM, Reddy TA. Time Series analysis of hourly horizontal solar radiation. Solar Energy 1988;41:423–9.
[20] Box GEP. Jenkins. G.M., Time series analysis: forecasting and control. Holden-Day Inc., San Francisco, CA, 1970.
[21] Uriel Ezequiel. (1985). Análisis de Series Temporales. Modelos Temporales. Ediciones Paraninfo.