Air Temperature Model Evaluation in the North Mediterranean Belt Area

JULIA BILBAO AND ARGIMIRO H. DE MIGUEL

Department of Applied Physics, University of Valladolid, Valladolid, Spain

HARRY D. KAMBEZIDIS

Institute of Environmental Research and Sustainable Development, National Observatory of Athens, Athens, Greece

(Manuscript received 3 July 2001, in final form 7 February 2002)

ABSTRACT

A comparative assessment of air temperature models, using hourly and daily air temperature measurements in 34 different stations in the north Mediterranean belt, is presented. Four air temperature models were used to estimate hourly and daily mean air temperature from daily maximum, daily minimum, and monthly mean air temperature. Root-mean-square error (rmse), scatter graphs, and cumulative frequency curves were used to determine the performance of each model. The best overall performance for estimating hourly air temperature from monthly mean values was presented by Erbs's model; the ''standard'' model gave the best performance for estimating daily mean air temperature from daily minimum and maximum air temperature values. The results show that the Erbs and standard models are the best for all stations used. A new Climatic Synthetic Time Series for the Mediterranean Belt Temperature Model (CLIMEDTEM) for estimating daily air temperature was developed by the authors with the help of available data banks, yielding a stochastic model that showed fits to the data with rmse values of 14%.

1. Introduction

The performance of all solar-energy systems is dependent upon solar radiation, ambient temperature, humidity, and wind speed. These variables are neither completely random nor deterministic and can best be described as random functions of time. The analysis of solar-energy systems and simulation methods in energy efficiency is inconvenienced by the random behavior of the weather. Information concerning hourly and daily air temperature values is required for most practical applications of solar energy in active and passive systems. Photovoltaic and thermal solar systems, building design, and thermal simulation performance analysis all require air temperature values. However, in different geographical areas these data are not available and must be estimated through models that use daily maximum, daily minimum, or monthly average air temperature values obtained from published data.

Temperature is one of the main meteorological variables measured by meteorological service networks. Nevertheless, daily and hourly time series, required for the most sophisticated studies and simulations, are expensive. Moreover, they often contain missing data, correspond to short recording periods, and, worst of all, are available for only very few sites world wide. In the European Union (EU), this situation is worse for the north Mediterranean belt area, where meteorological networks have low-timescale recording stations that are often hundreds of kilometers apart.

There are several existing models to predict daily mean, daily maximum, and daily minimum air temperature values, and some of them are based on data from stations in North America, Italy, Germany, or Spain. The following models are among the most prominent. Cuomo et al. (1986) studied and analyzed air temperature on a daily basis in the Italian climate. Amato et al. (1989) discussed stochastic–dynamic models for both air temperature and solar irradiance daily time series in the Italian climate. Hernández et al. (1991) developed stochastic models for the prediction of daily minimum air temperatures. Macchiato et al. (1993) analyzed cold and hot air temperatures observed at 50 stations in southern Italy. Some of the existing models have also been developed for northern latitudes with high albedos and cold air masses. For instance, Heinemann et al. (1996) developed an algorithm for the synthesis of hourly ambient temperature time series that takes into account a monthly average daily temperature pattern. The above-mentioned works developed models for predicting air temperature values; however, it is hard to find studies that compare the performance of different temperature models in order to obtain the best models for

Corresponding author address: Dr. Julia Bilbao, Department of Applied Physics, University of Valladolid, 47005, Valladolid, Spain. E-mail: juliab@fa1.uva.es

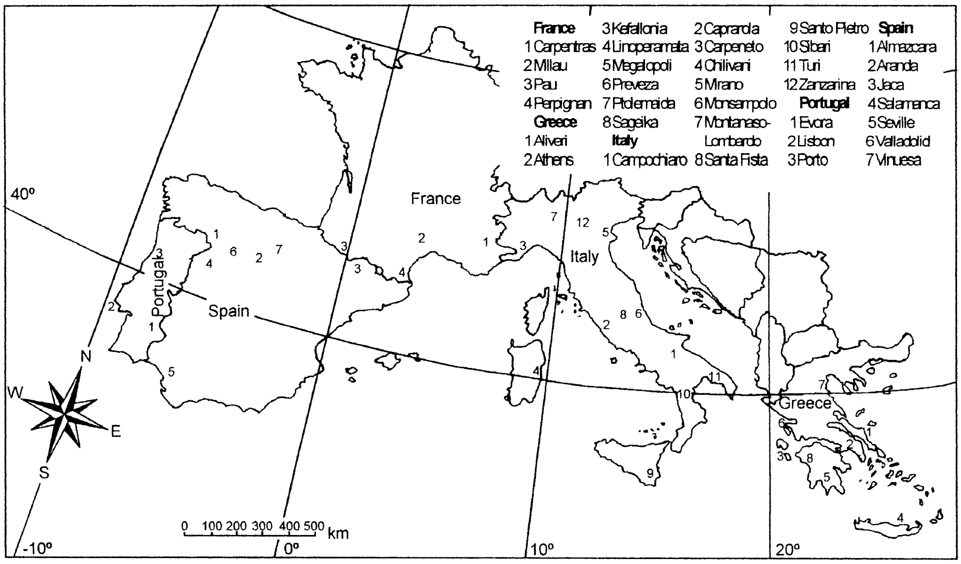

FIG. 1. Map of the selected measuring stations in each country of the Mediterranean belt area.

different places. In the framework of the ''JOULE III'' project on Climatic Synthetic Time Series for the Mediterranean Belt (CLIMED), a Mediterranean dataset was assembled by the participating institutions with the purpose of simulating meteorological variables.

The aim of this paper is to survey a number of air temperature models and to validate a selected subset based on the Mediterranean dataset in order to select the one(s) most suited to predict hourly and daily air temperature values necessary for photovoltaic, thermal, and building-energy analysis. Model selection was made on the basis of model equation availability, necessary input variables, and ability to generate data from limited average values.

The models selected may be divided into the following groups:

1) hourly models developed from daily minimum and maximum air temperature values,
2) stochastic models that link hourly air temperature to monthly average air temperature values, and
3) daily global models that link daily mean temperature and daily minimum and maximum values.

After selection, the models were tested using datasets from 34 measuring stations in France, Greece, Italy, Portugal, and Spain. The datasets consist of hourly and daily mean air temperature and daily maximum and minimum measured temperature data. After the analysis of the existing models, a new daily stochastic model based on the European Mediterranean dataset was proposed. For testing the models, the measured and generated data were compared in some north Mediterranean locations. The performance of the selected models was assessed using statistical characteristics of measured and modeled data of stations considered to be typical of various climatic divisions of the north Mediterranean belt area. The most appropriate model for the different climatic zones was subsequently proposed.

The theoretical base of the selected models, the new proposed model, the performance of different models, and the recommended models for use in this area are given in the following sections.

2. Data collection

A large volume of data from various sites in the south of France, Greece, Italy, Portugal, and Spain was collected. Not all operating stations were used in this study because of redundant and poor data quality for some stations (Kambezidis and Adamopoulos 1997). The main criterion for selection of the stations was the completeness of the database and the period covered. Another issue in the selection procedure was to cover the various climatic zones encountered in each country as fully as possible.

After the selection process, 34 measuring stations with multiyear records (ranging from 4 to 15 yr) of hourly air temperature, located in the five countries mentioned above, were retained for the work. Figure 1 shows the stations' situation in each country and reflects the climatological characteristics, following Kambezidis and Adamopoulos (1997). It can be said that the Portugal stations, numbers 2 and 3, belong to the Atlantic maritime zone with full influence from weather coming from the open ocean, and station 1 belongs to the Atlantic semimaritime zone, combining features of Atlantic and continental climates. In Spain, station 3 belongs to the mountainous Pyrenean zone and the rest of the stations belong to continental climate with cold winters and sunny summers with high temperatures. Because of its latitude, station 5 (Seville) has a milder winter. Most stations in France belong to a transition zone from oce-

TABLE 1. Number of data and geographical parameters used at the meteorological stations in the north Mediterranean belt area (height is meters above sea level).

Country	Station	Years	Data No.	Lat (°N)	Lon (°)	Height (m)
France	Carpentras	1978–87	87 591	44.08	5.05 E	105
	Millau	1993–96	34 988	44.12	3.04 E	715
	Pau	1993–96	35 025	43.38	0.42 W	183
	Perpignan	1993–96	35 064	42.73	2.87 W	42
Greece	Aliveri	1990–94	36 752	38.39	24.05 E	40
	Athens	1980–94	131 496	37.97	23.72 E	107
	Kefallonia	1994–96	25 082	38.17	20.48 E	5
	Linoperamata	1991–94	26 259	35.34	25.56 E	5
	Megalopoli	1990–94	38 315	37.42	22.11 E	430
	Preveza	1993–96	19 630	38.96	20.75 E	6
	Ptolemaida	1991–94	32 324	40.49	21.72 E	640
	Sageika	1993–96	32 105	38.11	21.47 E	50
Italy	Campochiaro	1993–96	29 772	41.47	14.53 E	502
	Caprarola	1993–96	29 419	42.32	12.17 E	650
	Carpeneto	1993–96	26 428	44.67	8.60 E	329
	Chilivani	1993–96	30 620	40.62	8.93 E	216
	Mirano	1993–96	31 147	45.50	12.10 E	9
	Monsampolo	1993–96	31 346	42.88	13.78 E	43
	Montanaso-Lombardo	1993–96	30 312	45.32	9.45 E	83
	Santa Fista	1993–96	30 479	43.52	12.12 E	311
	Santo Pietro	1993–96	28 233	37.12	14.52 E	313
	Sibari	1993–96	26 849	39.73	16.45 E	10
	Turi	1993–96	29 969	40.92	17.00 E	230
	Zanzariva	1993–96	31 741	45.20	10.52 E	40
Portugal	Evora	1978–90	110 289	38.57	7.90 W	350
	Lisbon	1984–90	58 872	38.70	9.10 W	70
	Porto	1980–90	96 360	41.14	8.60 W	200
Spain	Almazcara	1992–95	31 125	42.59	6.50 W	584
	Aranda	1989–95	54 719	41.65	3.69 W	798
	Jaca	1993–96	30 993	42.58	0.53 W	820
	Salamanca	1989–98	82 056	40.96	5.66 W	782
	Seville	1989–96	65 020	37.38	6.00 W	15
	Valladolid	1988–98	64 974	41.64	4.77 W	735
	Vinuesa	1992–95	32 336	42.02	2.80 W	1326

anic to Mediterranean, but Pau (French station 3) has an oceanic climate with influence from the Atlantic Ocean. The stations selected in Greece are situated in the Mediterranean maritime (milder winters and summers) and Mediterranean terrestrial zone. The Italian stations are situated in the mountainous (Apennines) zone. Stations 1 (Campochiaro) and 8 (Santa Fista) have more frequent rainfall throughout the year. Stations 2, 10, 11, and 9 are situated in the Mediterranean maritime zone, and stations 3, 7, and 12 are situated in the continental zone (with cold winters and warm summers).

Table 1 shows the available number of data and the geographical characteristics of the stations, such as latitude, longitude, and altitude above sea level. Latitude ranges between 35.34°N for Linoperamata (Greece) and 45.50°N for Mirano (Italy). Altitude ranges between 5 m for Athens (Greece) and 1326 m for Vinuesa (Spain). This factor greatly influences the performance of air temperature models.

Some necessary quality-control tests were performed before the data were used, following Kambezidis and Adamopoulos (1997). The tests used follow European Commission criteria and were aimed at eliminating spurious data and inaccurate measurements resulting from the response error of the sensors. For the dry-bulb temperatures, the following limits were checked:

$$-20°\mathrm{C} \leq T(y, m, d, t) \leq +50°\mathrm{C},$$

where $T(y, m, d, t)$ is the hourly air temperature, y is the year, m is the month, d is the day, and t is the hour. The final hourly dataset contains more than a million data points.

An initial analysis of the data characteristics was made. From the data series, the daily mean air temperature values were standardized using Eq. (15) from section 3d, and the lag autocorrelation coefficients for lags between 1 and 7 were calculated. The results showed that the lag-1 autocorrelation coefficient for all stations differed substantially from 0 but that subsequent autocorrelation coefficients were all close to 0. Table 2 shows the station name and the lag-1 autocorrelation coefficient ρ obtained from the measured data. It can be seen that the obtained results are between 0.60 for Kefallonia (Greece) and 0.82 for Athens (Greece). The mountainous stations can be seen to obtain a smaller lag-1 autocorrelation coefficient value because of the high difference between air temperature values. Average values are between 0.72 and 0.75. Similar results have

TABLE 2. Lag-1 autocorrelation coefficient of the standardized measured and estimated daily air temperature values using CLIMEDTEM model.

Country	Station	Years	Autocorrelation coefficient ρ	ρ_{est}
France	Carpentras	1978–87	0.75	0.75
	Millau	1993–96	0.71	0.72
	Pau	1993–96	0.69	0.71
	Perpignan	1993–96	0.64	0.66
Greece	Aliveri	1990–94	0.74	0.75
	Athens	1980–94	0.82	0.82
	Kefallonia	1994–96	0.6	0.62
	Linoperamata	1991–94	0.58	0.59
	Megalopoli	1990–94	0.73	0.75
	Preveza	1993–96	0.65	0.66
	Ptolemaida	1991–94	0.76	0.77
	Sageika	1993–96	0.71	0.72
Italy	Campochiaro	1993–96	0.68	0.69
	Caprarola	1993–96	0.75	0.76
	Carpeneto	1993–96	0.72	0.73
	Chilivani	1993–96	0.67	0.68
	Mirano	1993–96	0.75	0.76
	Monsampolo	1993–96	0.75	0.76
	Montanaso-Lombardo	1993–96	0.75	0.76
	Santa Fista	1993–96	0.73	0.74
	Santo Pietro	1993–96	0.75	0.75
	Sibari	1993–96	0.67	0.68
	Turi	1993–96	0.72	0.73
	Zanzariva	1993–96	0.76	0.77
Portugal	Evora	1978–90	0.78	0.79
	Lisbon	1984–90	0.76	0.76
	Porto	1980–90	0.74	0.74
Spain	Almazcara	1992–95	0.71	0.71
	Aranda	1989–95	0.76	0.77
	Jaca	1993–96	0.64	0.65
	Salamanca	1989–98	0.78	0.80
	Seville	1989–96	0.80	0.81
	Valladolid	1976–78, 1982–99	0.75	0.76
	Vinuesa	1992–95	0.71	0.71

been obtained by Cuomo et al. (1986) and Klein (1987). The results indicate that air temperature changes slowly from day to day in the area.

3. Existing air temperature models

A set of models was selected as the most promising for further study. The criteria used for selecting models were (i) full availability of algorithms and numerical coefficients, (ii) use of input data that are either generally available or obtainable from available model cascades (from monthly to daily, from daily to hourly) (iii) that models are based on data for one or more Mediterranean sites, and (iv) the quality of the results reported by the original authors as well as those published in reviews. The models selected for the current work may be divided into four groups, as described below.

a. *Models that link hourly air temperature T(y, m, d, t) and the daily maximum and minimum air temperature values, $T_{min}(y, m, d)$ and $T_{max}(y, m, d)$*

In this kind of model, the hourly air temperature values are obtained from the daily maximum and minimum values. The following models were selected.

1) DOUBLE COSINE MODEL (1995)

A deterministic model to calculate the daily mean air temperature profile is the double cosine model, which is recommended by the National Meteorological Institute of Portugal (Aguiar 1996) and uses three sinusoidal segments to connect the times of occurrence of the daily maximum and minimum air temperatures. The model is given by the following expressions:

$$T(y, m, d, t) = T(y, m, d) - \cos\left[\frac{\pi(t_{\mathrm{Tmin}} - t)}{24 + t_{\mathrm{Tmin}} - t_{\mathrm{Tmax}}}\right]\frac{A_T(y, m, d)}{2} \quad 1 \le t \le t_{\mathrm{Tmin}}, \tag{1}$$

$$T(y, m, d, t) = T(y, m, d) + \cos\left[\frac{\pi(t_{\mathrm{Tmax}} - t)}{24 + t_{\mathrm{Tmax}} - t_{\mathrm{Tmin}}}\right]\frac{A_T(y, m, d)}{2} \quad t_{\mathrm{Tmin}} \le t \le t_{\mathrm{Tmax}}, \quad \text{and} \tag{2}$$

$$T(y, m, d, t) = T(y, m, d) - \cos\left[\frac{\pi(24 + t_{\mathrm{Tmin}} - t)}{24 + t_{\mathrm{Tmin}} - t_{\mathrm{Tmax}}}\right]\frac{A_T(y, m, d)}{2}$$

$$t_{\mathrm{Tmax}} \leq t \leq 24, \tag{3}$$

where $T(y, m, d)$ is the daily mean air temperature, $A_T(y, m, d)$ is the daily thermal amplitude (°C), t_{Tmin} is the hour at which the hourly minimum temperature occurs, t_{Tmax} is the hour at which the hourly maximum temperature occurs, and t is the hour of the day. To clarify the terms, there is an appendix at the end of the paper.

2) ERBS'S MODEL (1984)

The diurnal variation of the hourly monthly mean air temperature $T(y, m, t)$ has a location- and month-independent shape when it is standardized by subtracting the mean and dividing by the amplitude. From U.S. temperature data, the authors (Erbs et al. 1983) show that the average normalized diurnal temperature variation can be represented by

$$\begin{aligned} T(y, m, t) &= T(y, m) + A_T(y, m)[0.4632 \cos(\alpha - 3.805) \\ &\quad + 0.0984 \cos(2\alpha - 0.360) \\ &\quad + 0.0168 \cos(3\alpha - 0.822) \\ &\quad + 0.0138 \cos(4\alpha - 3.513)], \end{aligned} \tag{4}$$

where $\alpha = 2\pi(t - 1)/24$; t is the hour; $T(y, m)$ is the daily monthly mean air temperature; and $A_T(y, m)$ is the monthly mean thermal amplitude in degrees Celsius (difference between the monthly mean maximum and minimum air temperature values), which varies considerably with month and location. The term $A_T(y, m)$ is related to the monthly average clearness (Erbs 1984), which is defined as the ratio of the monthly global solar radiation on a horizontal surface to the monthly extraterrestrial solar radiation on a horizontal surface. The last two models can be adapted to work at the daily timescale, replacing the hourly monthly values by the hourly ones.

b. *Stochastic models that link hourly and monthly average temperature values*

Hollands et al. (1989) studied the effect of neglecting the random component in hourly temperature data for various solar heat systems. The results indicate that, for some systems, the extra complexities of including the random component of the hourly ambient temperature are unwarranted. Boland (1997) showed that the stochastic air temperature component is critical for evaluating heating and cooling loads for passive solar applications. As a result, there are systems for which its inclusion is important, and for this reason the following model is studied.

KNIGHT'S MODEL (1991)

Knight et al. (1991) proposed a model to generate hourly ambient temperature series with the random component included without introducing discontinuities between the last hour of one day and the first hour of the next day. The model enables hourly temperature values to be obtained from an autoregressive method. Erbs et al. (1983) and Knight et al. (1991) showed that the cumulative distribution of the normalized hourly temperature values could be represented by the following probability function:

$$F[h(m, d, t)] = \frac{1}{1 + \exp[-3.396h(m, d, t)]}, \tag{5}$$

where $h(m, d, t)$ is the normalized hourly air temperature, given by

$$h(m, d, t) = \frac{T(m, d, t) - T(m)}{\sigma(m)\sqrt{N_m/24}}, \tag{6}$$

where $T(m, d, t)$ is an hourly air temperature value of a representative year, $T(m)$ is the monthly mean air temperature of a representative year, N_m is the number of hours in the month, and $\sigma(m)$ is the standard deviation of the $T(m)$ with regard to its long-term mean value T_{yr}. Here, $\sigma(m)$, as a function of $T(m)$ and σ_{yr}, can be estimated by the following expression:

$$\sigma(m) = 1.45 - 0.0290T(m) + 0.0664\sigma_{\mathrm{yr}}, \tag{7}$$

where σ_{yr} is the standard deviation of the 12 $T(m)$s with regard to the long-term average temperature during the period.

Temperature simulation requires modeling variables whose probability structure changes with time. To take this question into account, Knight et al. (1991) transform the temperature values through their cumulative distribution function to a normally distributed variable χ, with mean of 0 and variance of 1. This transformed variable can then be represented by a first-order autoregressive model,

$$\chi_n = \phi_1\chi_{n-1} + \varepsilon_n, \tag{8}$$

where ϕ_1 is the lag-1 autocorrelation coefficient of the χ values, and ε is the random component.

Knight et al. (1991) propose the following transformation law:

$$F(h) = F'(\chi) = \frac{1}{\sqrt{2\pi}}\int_0^{\chi} \exp\left(-\frac{1}{2}z^2\right) dz, \tag{9}$$

where z is an auxiliary variable.

From this method, the transformation that relates hourly temperature values to the normalized value and the corresponding χ values is obtained.

From the long-term series of hourly temperature data, the standardized values were calculated, and from Eq. (9) the corresponding χ values and the lag-1 autocorrelation coefficient were evaluated. Each hour, a new χ value is generated according to a first-order autoregressive model, following Eq. (8).

The generated χ values are transformed to an hourly temperature by equating the cumulative distribution of χ,

$$F'(\chi) = \frac{1}{\sqrt{2\pi}} \int_0^{\chi} \exp\left(-\frac{1}{2}z^2\right) dz = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\chi}{\sqrt{2}}\right)\right], \quad (10)$$

and the temperature distribution function [Eq. (5)]. Solving the hourly air temperature, the following expression is obtained:

$$T(m, d, t) = T(m) - \frac{\sigma(m)\sqrt{\frac{N_m}{24}}}{3.396} \ln\left\{\frac{1}{0.5\left[1 + \operatorname{erf}\left(\frac{\chi}{2}\right)\right]} - 1\right\}, \quad (11)$$

where erf is the error function for the variable χ, $T(m)$ is the monthly mean air temperature of a representative year whose hourly values are being calculated, and N_m is the number of hours in the month. A complete description of the model is given by the authors (Knight et al. 1991).

c. *Models that link daily mean air temperature T(y, m, d) and daily maximum and minimum air temperature values, $T_{min}(y, m, d)$ and $T_{max}(y, m, d)$*

In this kind of model the daily mean air temperature is calculated from the daily maximum and minimum air temperature values.

STANDARD MODEL

The daily maximum and minimum air temperature series are normally available from the national meteorological services, and in this case the method used to calculate the daily mean air temperature is the "standard" method, given by the following equation:

$$T(y, m, d) = [T_{\max}(y, m, d) + T_{\min}(y, m, d)]/2. \quad (12)$$

This approximation would be exact if the daily mean air temperature profile were smooth and symmetrical, which is not true. For instance, the increase in temperature in the morning is steeper than its decrease in the afternoon and at night (Aguiar 1997).

d. *New proposed model*

Within each group of existing models there is considerable disagreement between results, depending upon which air temperature model is used, as can be seen in Bilbao et al. (1997). The differences may be the result of various methods of calculating the data, location dependence of the data, or insufficient data. Models are also sometimes based on data from only one location, often at higher latitudes than those of the majority of the Mediterranean stations. For these reasons a new daily model, the CLIMED Temperature Model (CLIMED-TEM), was developed using the combined data of Table 1 except those for Athens (Greece), Porto (Portugal), and Seville (Spain). These data were used to test the new model. The residual daily series histograms were analyzed in this study, and from the results it could be seen that the histograms are closely fitted by a Gaussian function distribution.

From the daily mean air temperature values $T(y, m, d)$, the average of the set of corresponding values for each allowable year,

$$\mu(m, d) = \frac{1}{\text{ny}} \sum_{y=1}^{\text{ny}} T(y, m, d), \quad (13)$$

was calculated together with the corresponding standard deviation,

$$\sigma(m, d) = \left\{\frac{\sum_{y=1}^{\text{ny}} [T(y, m, d) - \mu(m, d)]}{\text{ny}}\right\}^{0.5}, \quad (14)$$

where ny is the number of data years.

To eliminate the seasonal trends of the daily mean air temperature values, their standardized residuals with respect to the trend were evaluated using the following expression:

$$X(y, m, d) = \frac{T(y, m, d) - \mu(m, d)}{\sigma(m, d)}. \quad (15)$$

The standardized daily mean air temperature values $X(y, m, d)$ were compared with a standardized normal distribution, and the lag-2 and -3 correlation coefficients were calculated. From the results, it was obtained that daily mean air temperature values can be described well by a first-order autoregressive model, and, as a consequence, the standardized daily mean air temperature can be simulated by the following expression:

$$X(y, m, d) = \rho X(y, m, d - 1) + Y(y, m, d), \quad (16)$$

where ρ is the lag-1 autocorrelation coefficient of the daily mean air temperature, and $Y(y, m, d)$ is the independent stochastic variable with 0 mean and variance $(1 - \rho^2)$. When the lag-1 autocorrelation coefficient ρ and the distribution of $Y(y, m, d)$ are known, Eq. (16) enables us to build the sequence of $X(y, m, d)$, and then, using the following expression,

$$T(y, m, d) = \mu(m, d) + X(y, m, d)\sigma(m, d), \quad (17)$$

it is possible to evaluate synthetic values of $T(y, m, d)$ with the same statistical properties as the measured data series.

TABLE 3. Rmse statistical estimator values for air temperature computed from the Knight, Erbs, and double cosine models.

Model		Knight		Double cosine		Erbs	
Country	Station	Rmse (°C)	Rmse (%)	Rmse (°C)	Rmse (%)	Rmse (°C)	Rmse (%)
France	Carpentras	5.68	50.25	2.23	16.90	1.82	13.81
	Millau	5.16	67.70	1.83	17.40	1.49	14.22
	Pau	5.42	53.94	2.22	17.01	1.66	12.76
	Perpignan	4.81	30.99	2.19	14.09	1.79	11.52
Greece	Aliveri	5.11	29.94	1.69	9.88	1.38	8.09
	Athens	4.77	33.56	1.47	8.35	1.08	6.12
	Kefallonia	5.04	30.31	2.07	12.49	1.69	10.19
	Linoperamata	4.14	21.90	1.79	9.46	1.47	7.80
	Megalopoli	6.25	44.81	1.93	13.83	1.74	12.44
	Preveza	4.37	25.07	1.69	9.71	1.41	8.11
	Ptolemaida	6.04	47.35	1.60	12.52	1.36	10.64
	Sageika	5.64	30.92	1.86	10.17	1.46	8.00
Italy	Campochiaro	6.55	56.78	3.19	27.68	2.57	22.29
	Caprarola	4.85	39.21	1.87	15.16	1.51	12.26
	Carpeneto	4.65	35.48	1.44	11.02	1.22	9.31
	Chilivani	6.11	40.73	2.56	17.09	2.08	13.89
	Mirano	5.74	48.17	2.13	17.85	1.66	13.89
	Monsampolo	5.26	36.46	2.46	17.07	2.08	14.39
	Montanaso-Lombardo	5.25	39.77	1.45	10.95	1,22	9.20
	Santa Fista	6.33	53.12	2.48	20.78	1.98	16.61
	Santo Pietro	5.13	32.36	2.61	16.44	2.12	13.34
	Sibari	4.97	29.34	2.59	15.27	2.13	12.57
	Turi	5.43	37.87	2.45	17.10	2.04	14.26
	Zanzariva	5.53	43.64	1.97	15.58	1.54	12.16
Portugal	Evora	5.31	38.09	1.60	10.13	1.71	10.78
	Lisbon	4.32	32.81	1.45	8.83	1.28	7.77
	Porto	4.74	39.43	1.55	10.75	1.46	10.11
Spain	Almazcara	6.38	57.87	2.40	21.75	1.83	16.61
	Aranda	6.55	56.24	2.34	20.12	1.81	15.57
	Jaca	5.88	51.98	2.15	18.97	1.66	14.63
	Salamanca	6.12	57.64	1.81	15.25	1.45	12.25
	Seville	5.88	30.02	1.96	9.98	1.37	7.01
	Valladolid	6.13	48.02	1.73	13.59	1.60	12.55
	Vinuesa	6.37	69.76	2.94	32.17	2.18	23.87

4. Performance results and discussion

The accuracy of the five different models was evaluated and assessed by means of widely used statistics: rmse (absolute and relative values), scatterplots of the estimated values as a function of the measured ones (Kambezidis et al. 1994), and histograms and cumulative frequency distributions (Knight et al. 1991). The following expressions for rmse (°C and %: absolute and relative values) were used:

$$\text{rmse (°C)} = \left[\sum_{i=1}^{N} (T_{i,\text{est}} - T_{i,\text{me}})^2 \Big/ N\right]^{0.5} \quad \text{and} \qquad (18)$$

$$\text{rmse (\%)} = \frac{100}{\overline{T}_{\text{me}}}\left[\sum_{i=1}^{N} (T_{i,\text{est}} - T_{i,\text{me}})^2 \Big/ N\right]^{0.5}, \qquad (19)$$

where $T_{i,\text{est}}$ is the estimated temperature value, N is the number of data points, $T_{i,\text{me}}$ is the measured temperature value, and $\overline{T}_{\text{me}}$ is the long-term air temperature value. Average statistical estimator values were calculated (de Miguel et al. 2001) for each of the compared models and stations, and the following results were obtained from the comparison of daily and hourly simulated air temperature values with the measured datasets.

a. Models that link hourly air temperature T(y, m, d, t) with daily maximum and minimum air temperature and daily monthly mean air temperature values: The Knight, double cosine, and Erbs models

In this group, models that link hourly air temperature values with the daily maximum and minimum air temperature and monthly mean daily air temperature were tested. Table 3 shows the name of the model and the statistical estimators (absolute and relative rmse values) obtained for all selected stations.

Comparing the rmse absolute results, more scatter is observed in Knight's model (which can vary from 4.14° to 6.55°C depending on the station) than in the case of Erbs's model (which varies from 1.08° to 2.57°C). In comparing the rmse relative values, it can be said that Knight's model shows the highest values, ranging between 21.90% and 69.76%, and Erbs's model obtains the lowest values, between 6.12% and 23.87%. The lowest rmse is obtained for stations in both the Mediterranean and Atlantic maritime zones, for instance, Athens at 6.12% and Lisbon at 7.77%.

The highest rmse values for the Erbs model were obtained for higher-level stations belonging to mountain climatic zones, for instance, Jaca (14.63%) and Vinuesa (23.87%) in Spain and Campochiaro (22.29%) and Santa Fista (16.61%) in Italy. In Greece, the stations obtaining greatest rmse value were Megalopoli (12.44%) and Ptolemaida (10.64%), both of which belong to the terrestrial Mediterranean climatic zone. Millau, France, located in the Mediterranean transitional zone and the station with the highest latitude, obtained the highest rmse value. Perpignan obtained the best result in relation to Erbs's model, being as it is the nearest station to the sea and is located at a lower altitude than the other French stations.

In conclusion, it may be seen that Erbs's model obtained the best results. This is because it gives the lowest error values at all stations and better predicts the air temperature values at the stations that belong to maritime climatic zones, except in Italy where the model obtained the best results for stations belonging to the continental zone, as can be seen, for instance, in Carpeneto (9.31%) and Montanaso-Lombardo (9.20%).

Comparison of scatterplots with hourly estimated versus measured air temperatures are made at each station, and, in the interest of clarity, selected plots were chosen for discussion of the results. Plot selection was based on the minimum relative rmse value to pick the best model and on the maximum relative rmse value to pick the worst. The diagonal line represents the ideal match between the estimated and measured values. Figures 2a and 2b show examples of the estimated versus measured hourly air temperature values using Erbs model results. Each figure consists of two scatter graphs; Fig. 2a shows the best performing model and Fig. 2b the worst. In comparing the scatter graphs of the double cosine, Knight, and Erbs models, it appears that the Erbs model introduces better predictions than the other models because the parameter model is more dependent on the location.

Figures 3a and 3b show the comparison of cumulative frequency distribution curves, where long-term distribution was obtained with measured data. In comparing these results, it can be said that the Erbs, double cosine, and Knight models for Athens (Greece) and Valladolid (Spain), respectively, show a similar behavior. In both figures, the results from Knight's model differ from the long-term distribution for high and low temperature values. The Erbs model distribution is close to the long-term results, indicating good results.

From these results it can be said that Erbs's model gives the best overall results in predicting the hourly temperature, and the double cosine model performed better than Knight's model.

b. Daily air temperature models: Standard and CLIMEDTEM models

In this group, the standard and CLIMEDTEM models were tested. The standard model links daily mean air temperature values with the corresponding daily maximum and minimum air temperature values. The CLIMEDTEM model is a newly proposed stochastic model that links daily mean air temperature with monthly average temperature values. The difference between these models is the input data; the standard model needs maximum and minimum data, and CLIMEDTEM only needs monthly average temperature values.

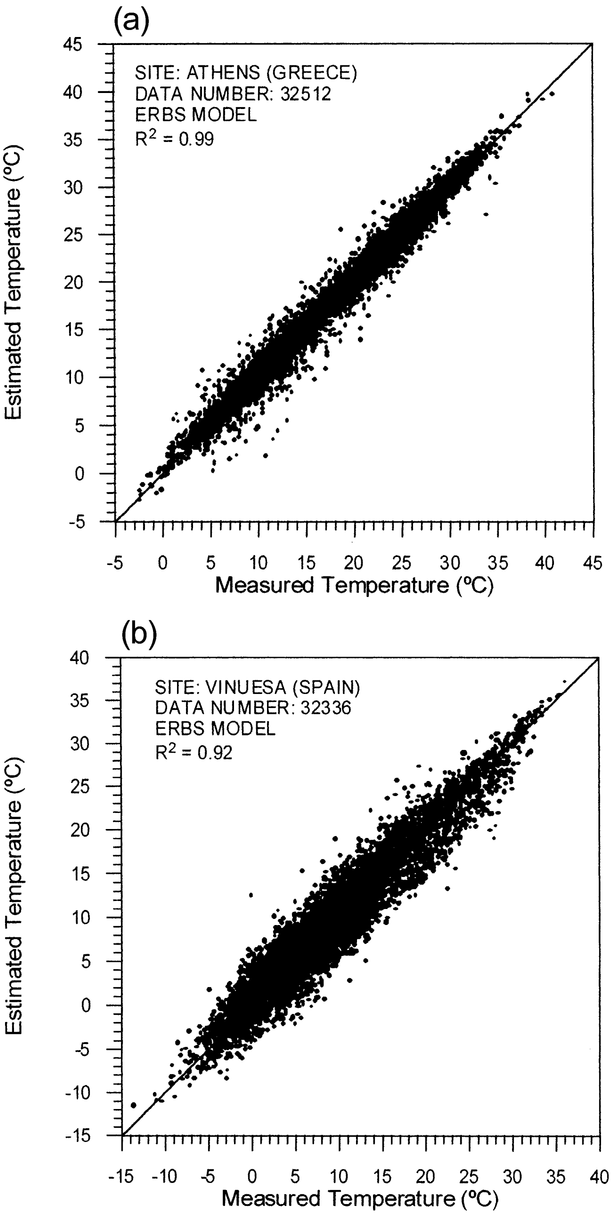

FIG. 2. Hourly air temperature values estimated by the Erbs model compared with measured values for (a) Athens, Greece, and (b) Vinuesa, Spain.

Table 4 shows the statistical estimators of the standard and CLIMEDTEM models. In comparing the rmse relative values, it can be observed that the standard model obtains smaller relative error values. Aliveri and Sag-

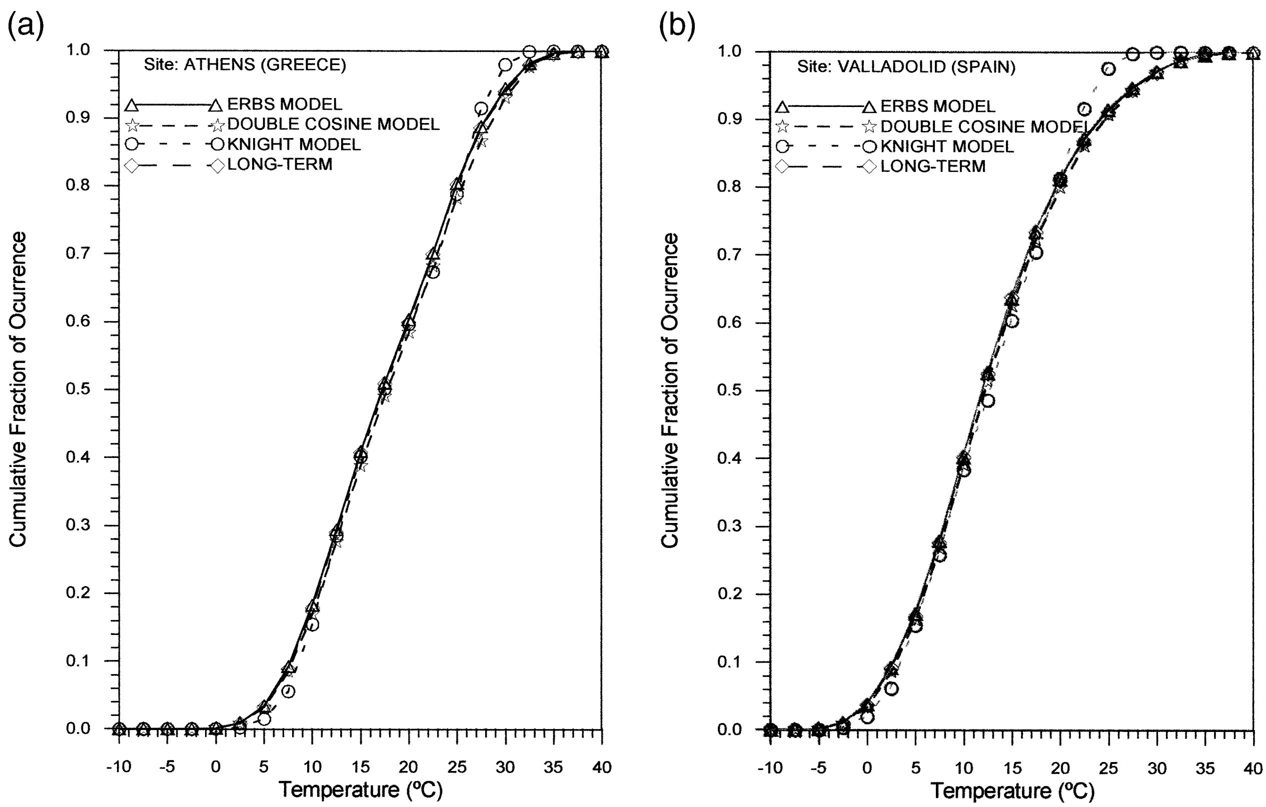

FIG. 3. Hourly air temperature cumulative probability distributions calculated from long-term data and estimated from the Erbs, double cosine, and Knight models for (a) Athens, Greece, and (b) Valladolid, Spain.

eika (Greece) obtain the best results and Vinuesa (Spain) the worst. The rmse absolute values vary between 0.49° and 1.31°C, depending on the station. Figures 4a and 4b show the estimated versus the measured values for the best and the worst stations, respectively. Figure 4a shows that, for the standard model, the scatter is small for Aliveri (Greece) in comparison with Fig. 4b, for Vinuesa (Spain), which has the worst result. It can also be observed that the simulated values are underestimated in Vinuesa (Spain).

Figure 5 shows the performance of the standard model and the comparison between estimated and measured values for Athens (Greece). It can be seen that estimated values are similar to measured ones and the model slightly underestimates the results.

The CLIMEDTEM model (Table 4) gives the lowest rmse values for the following stations: Linoperamata, Preveza, and Sageika in Greece; Sibari in Italy; and Seville in Spain. All these stations have a relatively low latitude and low altitude. In comparing the rmse absolute values, more scatter is observed from the CLIMEDTEM model (which varies from 4.65° to 1.17°C depending on the station) than on the standard model. Figures 6a and 6b show the estimated versus the measured values for the CLIMEDTEM model in Aliveri (Greece) and Vinuesa (Spain), respectively. The model overestimates the air temperature for Vinuesa.

Table 2 shows the lag-1 autocorrelation coefficient values for measured and CLIMEDTEM model estimated values ρ_{est}. The results are similar and agree with those obtained by Klein (1987), and from these results it can be said that the model performs well for the temperature values in the Mediterranean area. In comparing the standard and CLIMEDTEM models, it can be said that the standard model gives the best results because it is based on local temperature values of daily maximum and minimum temperature data, but, because the standard model needs long data series, it could be inconvenient to use in the Mediterranean area. The CLIMEDTEM model is an autoregressive model that only needs temperature data and variables on a big timescale that can be evaluated from a few climatological data values, as has been shown in previous sections. From the study and taking into account the results, it can be said that the CLIMEDTEM model gives good results for stations that have a relatively low altitude and latitude; CLIMEDTEM may be useful when monthly average temperature values are available, and the advantages are that it requires fewer input data than the other daily temperature model shown in the work.

In comparing Knight and CLIMEDTEM model results based on monthly temperature values, it can be said that these models may be useful when monthly average temperature data are available, which are values

TABLE 4. Rmse statistical estimator values for CLIMEDTEM and standard air temperature models.

Model		CLIMEDTEM		Standard	
Country	Station	Rmse (°C)	Rmse (%)	Rmse (°C)	Rmse (%)
France	Carpentras	1.17	31.71	0.72	5.54
	Millau	4.17	39.67	0.61	5.85
	Pau	3.82	29.29	0.75	5.84
	Perpignan	3.35	21.56	0.63	4.07
Greece	Aliveri	4.13	24.17	0.57	3.34
	Athens	3.77	21.38	0.66	3.74
	Kefallonia	2.83	17.01	0.75	4.53
	Linoperamata	2.62	13.85	0.65	3.44
	Megalopoli	3.50	25.10	0.71	5.14
	Preveza	2.49	14.30	0.63	3.64
	Ptolemaida	4.65	36.40	0.68	5.34
	Sageika	2.96	16.21	0.60	3.29
Italy	Campochiaro	3.89	33.76	0.95	8.33
	Caprarola	3.71	30.01	0.62	5.03
	Carpeneto	3.14	23.98	0.49	3.78
	Chilivani	3.24	21.60	0.74	5.00
	Mirano	3.27	27.39	0.70	5.85
	Monsampolo	3.15	21.82	0.64	4.44
	Montanaso-Lombardo	3.10	23.46	0.58	4.44
	Santa Fista	3.60	30.19	0.88	7.40
	Santo Pietro	3.13	19.74	0.69	4.34
	Sibari	3.01	17.74	0.64	3.77
	Turi	3.50	24.37	0.60	4.14
	Zanzariva	3.09	24.42	0.66	5.25
Portugal	Evora	4.28	27.01	0.56	3.59
	Lisbon	3.29	20.02	0.69	4.22
	Porto	3.64	25.30	0.72	5.02
Spain	Almazcara	4.02	36.39	0.96	8.73
	Aranda	4.23	36.23	0.84	7.15
	Jaca	3.96	34.87	0.74	6.55
	Salamanca	4.37	36.68	0.75	6.41
	Seville	3.63	18.51	0.79	4.08
	Valladolid	4.18	34.53	1.31	10.89
	Vinuesa	4.01	43.83	1.03	11.34

that can be obtained easily from different publications and meteorological atlases. In conclusion, the CLIMEDTEM model may be used in the Mediterranean belt area where monthly temperature values are available—for example, from temperature isoline maps—and the best results could be obtained in places with low altitude and latitude.

5. Conclusions

A dataset was assembled and used that is thought to be among the best available at this time in the north Mediterranean belt area. The data analysis performed has shown that the lag-1 autocorrelation coefficient is independent over time and its variations with location are negligible, at least for maritime stations. The different established models that calculate hourly and daily air temperature have been selected, run, and tested to decide which model is recommended, and a new model has also been proposed for the north Mediterranean area.

The models studied were classified into two groups. In the first group, selected models that calculate hourly air temperature from daily maximum and minimum air temperature values were tested. In the second group, one daily air temperature model was run and a new model was proposed.

The statistical estimators rmse (absolute and relative values), cumulative probability distributions, and scatterplots were used to indicate how closely the models agree with the data, and the variation of the rmse values with climatic zone has been studied. Among the first group of models, Erbs's model is recommended as the best for maritime climatic zones and for mountainous climatic zones in Italy, because it gives the best test results (small rmse values, together with the best scatterplot and cumulative frequency distribution) and provides good estimation for all data.

Among the models that calculate daily air temperature, the standard model is the most recommended, because it is the best in reproducing the statistical characteristics of data in all climatic zones.

It has been observed that a model's performance also depends on station altitude and climatic zone. In most cases, all models perform well at stations near sea level, although some previous studies show that altitude can be a dominant parameter (Macchiato et al. 1995). Stochastic models have also been tested with measured data from different stations in the Mediterranean area. The

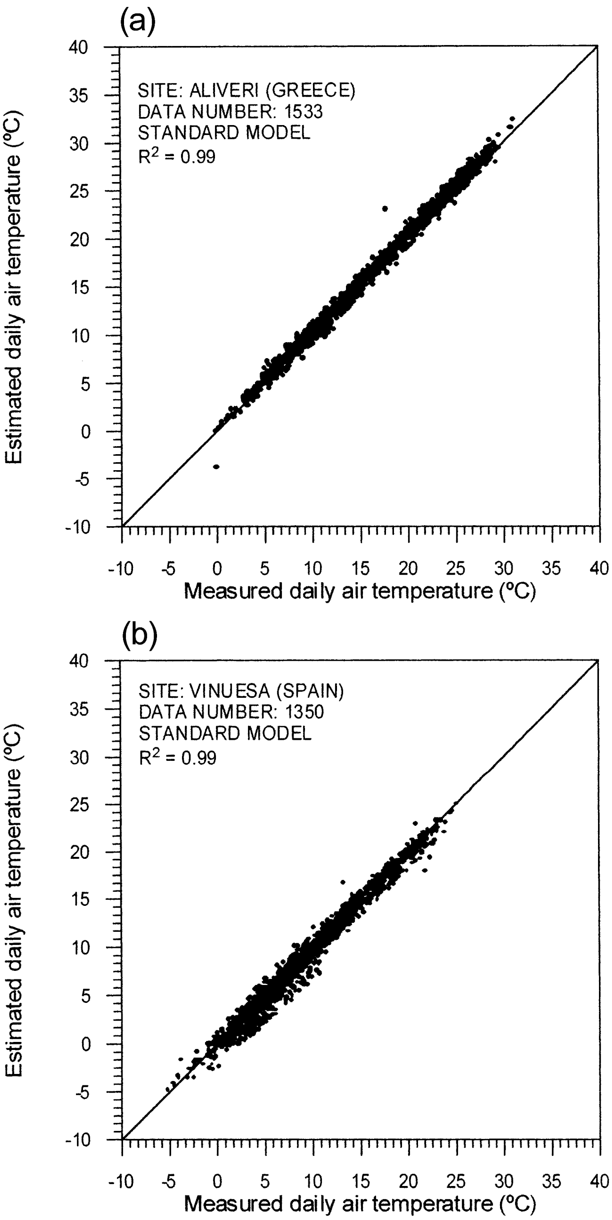

FIG. 4. Daily mean air temperature measured values for (a) Aliveri, Greece, and (b) Vinuesa, Spain, compared with estimated values obtained by the standard model.

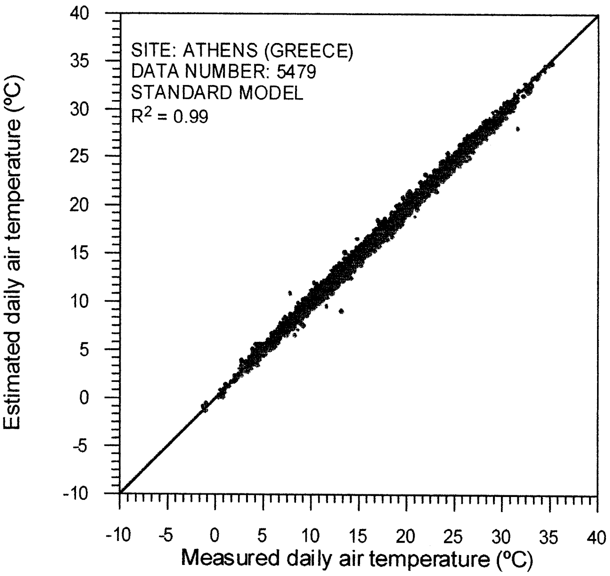

FIG. 5. Same as Fig. 4, but for Athens, Greece.

stochastic CLIMEDTEM model could be used for predicting daily air temperature at the southern maritime zones of the studied region and in low-latitude and -altitude cities, for instance, Seville (Spain). The advantages of the two stochastic models, Knight and CLIMEDTEM, are the limited necessary input data series in comparison with the analyzed models.

The study gives some new evidence that, for the Mediterranean area and by means of stochastic models, only limited temperature information will be needed as input to simulate data, and thus synthesized data might be obtained in many more locations.

The results of the paper can be used in different scientific areas such as solar climatology, renewable solar energy simulation and design, energy-efficiency studies, and solar-energy engineering applications, as well as in other scientific fields for which air temperature data, in different timescales, are required as input for important system simulations.

Acknowledgments. This investigation is a part of the JOULE III Project on Climatic Synthetic Time Series for the Mediterranean Belt (Contract JOR3-CT96-0042), known as CLIMED, for which the participating institutions were The Institute of Industrial Engineering (INETI) in Lisbon, Portugal, which was the coordinator, the National Observatory of Athens (NOA), Greece; the University of Valladolid (Spain); and, as a subcontractor to INETI, Energy Consulting of Aix en Provence, France. The authors thank the CLIMED Project Coordinator Dr. R. Aguiar for the project management. The authors gratefully acknowledge the financial support extended by the EU JOULE III Programme. The authors express their thanks to the national meteorological services in the countries involved for making the necessary data for this study available. The anonymous reviewers are also gratefully acknowledged for their useful comments and suggestions in improving the paper.

APPENDIX

Nomenclature

a. Latin symbols

$A_T(y, m, d)$	daily thermal amplitude (°C)
$A_T(y, m)$	monthly mean thermal amplitude (°C)

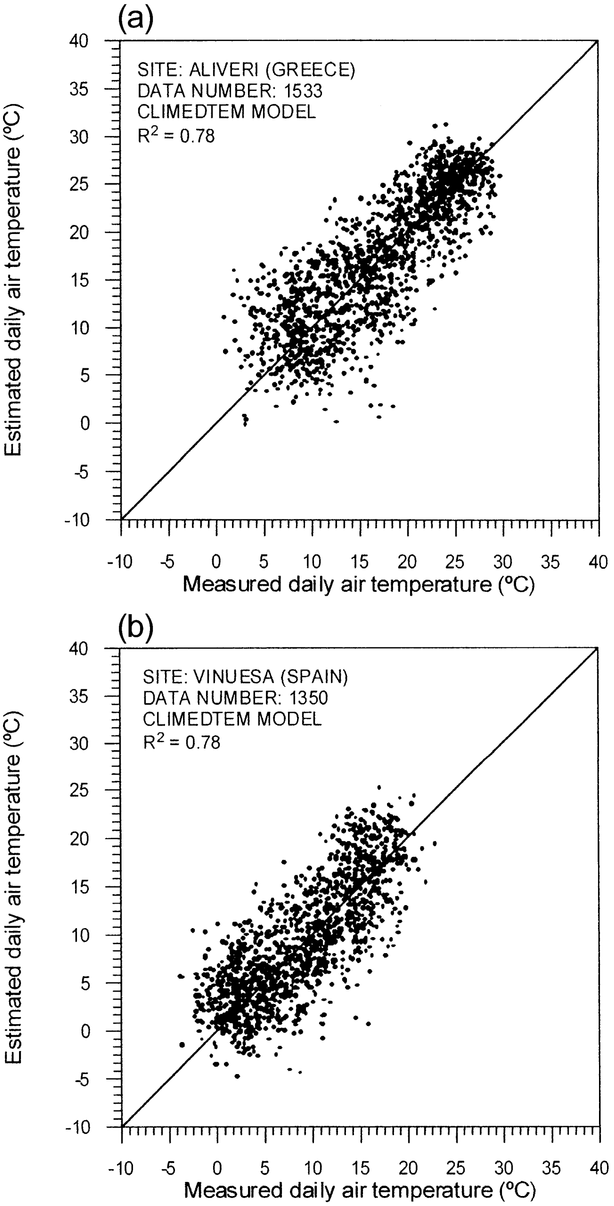

FIG. 6. Comparison between measured and estimated daily mean air temperature values by the CLIMEDTEM model for (a) Aliveri, Greece, and (b) Vinuesa, Spain.

$F[h(m, d, t)]$	cumulative distribution of hourly air temperature
$h(m, d, t)$	normalized hourly air temperature
N_m	No. of hours in the month
$Y(y, m, d)$	independent normal variable with 0 mean and $(1 - \rho^2)^{1/2}$ std dev
t	hour of the day
$T(y, m, d, t)$	hourly air temperature (°C)
$T(y, m, d)$	daily mean air temperature (°C)
$T_{\max}(y, m, d)$	daily max air temperature (°C)
$T_{\min}(y, m, d)$	daily min air temperature (°C)
$T(y, m, t)$	hourly monthly mean air temperature (°C)
$T(y, m)$	daily monthly mean air temperature (°C)
$T(m)$	monthly mean air temperature for a representative year (°C)
$T(m, d, t)$	hourly air temperature of a representative year (°C)
T_{yr}	yearly mean air temperature over the whole period (°C)
$T_{i,\mathrm{est}}$	estimated (daily or hourly) mean air temperature values (°C)
$T_{i,\mathrm{me}}$	measured (daily or hourly) mean air temperature values (°C)
$X(y, m, d)$	standardized daily mean air temperature (°C)
y	No. of the year
m	No. of the month
d	No. of the day

b. Greek symbols

ε	normally distributed random variable with 0 mean and variance of $(1 - \phi_1)^2$
ϕ_1	lag-1 autocorrelation coefficient in stochastic model
χ	normally distributed stochastic variable with 0 mean and variance of 1
ρ	lag-1 autocorrelation coefficient of standardized daily mean air temperature values
$\mu(m, d)$	long-term daily mean air temperature value over the corresponding set of values for each allowable year (°C)
$\sigma(m, d)$	daily temperature std dev with regard to long-term daily mean air temperature (°C)
$\sigma(m)$	std dev of a month's daily mean air temperature with regard to the long-term mean value for that mouth (°C)
σ_{yr}	std dev of the $T(m)$ values with regard to the yearly mean daily temperature (°C)

REFERENCES

Aguiar, R., 1996: Séries sintèticas de parâmetros meteorològicos (Synthetic series of meterological parameters). Ph.D. thesis. Lisbon University, Lisbon, Portugal, 525 pp.

——, 1997: Climatic synthetic series for the Mediterranean belt. Instituto Nacional de Engenharia E Tecnologia Industrial. Final CLIMED Project Rep., 101 pp.

Amato, U., V. Cuomo, F. Fontana, and F. C. Serio, 1989: Statistical predictability and parametric models of daily ambient temperature and solar irradiance: An analysis in the Italian climate. *J. Appl. Meteor.*, **28,** 711–721.

Bilbao, J., A. de Miguel, J. A. Medina, and J. J. López, 1997: Model performance tests. CLIMED Project Rep. to European Community DGXII, Applied Physics I Dept., Valladolid University, Spain, 183 pp.

Boland, J., 1997: The importance of the stochastic component of climatic variable in simulating the thermal behavior of domestic dwellings. *Sol. Energy,* **60,** 359–370.

Cuomo, V., F. Fontana, and C. Serio, 1986: Behaviour of ambient temperature on daily basis in Italian climate. *Rev. Phys. Appl.,* **21,** 211–218.

de Miguel, A., J. Bilbao, R. Aguiar, H. Kambezidis, and E. Negro, 2001: Diffuse solar irradiation model evaluation in the north Mediterranean belt area. *Sol. Energy,* **70,** 143–153.

Erbs, D. G., 1984: Models and applications for weather statistics related to building heating and cooling loads. Ph.D. thesis, Mechanical Engineering Dept., University of Wisconsin—Madison, 336 pp.

——, S. A. Klein, and W. A. Beckman, 1983: Estimation of degree-day and ambient temperature bin data from monthly-average temperatures. *ASHRAE J.,* **25,** 60–65.

Heinemann, D., C. Langer, and J. Schumacher, 1996: Synthesis of hourly ambient temperature time series correlated with solar radiation. *Proc. EuroSun'96 Conf.,* Freiburg, Germany, ISES-Europe, 1518–1523. [Available from Energy Department, Oldenburg University, D-26111 Oldenburg, Germany.]

Hernández, E., R. García, and M. T. Teso, 1991: Minimum temperature forecasting by stochastic techniques: An evidence of the heat island effect. *Mausam,* **41,** 161–166.

Hollands, G. T., L. T. D'Andrea, and I. D. Morrison, 1989: Effect of random fluctuations in ambient air temperature on solar system performance. *Sol. Energy,* **42,** 335–338.

Kambezidis, H. D., and A. D. Adamopoulos, 1997: Final data set. Third Progress Rep. of the CLIMED Project to the European Community. JOULE DG XII Programme, 21 pp. [Available from Institute of Environmental Research and Sustainable Development, National Observatory of Athens, Athens, Greece.]

——, B. E. Psiloglou, and C. Gueymard, 1994: Measurements and models for total solar irradiance on inclined surface in Athens, Greece. *Sol. Energy,* **53,** 177–185.

Klein, S. A., 1987: Scientific vs. correlation methods. *Proc. ISES Solar World Congress 1987,* Hamburg, Germany, ISES, 3109–3114.

Knight, K. M., S. A. Klein, and J. A. Duffie, 1991: A methodology for the synthesis of hourly weather data. *Sol. Energy,* **46,** 109–120.

Macchiato, M., C. Serio, V. Lapenna, and L. La Rotonda, 1993: Parametric time series analysis of cold and hot spells in daily temperature: An application in southern Italy. *J. Appl. Meteor.,* **32,** 1270–1281.

——, L. La Rotonda, V. Lapenna, and M. Ragosta, 1995: Time modelling and spatial clustering of daily ambient temperature: An application in southern Italy. *Environmetrics,* **6,** 31–53.