Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Energy 31 (2006) 208–227
ENERGY
www.elsevier.com/locate/energy

Models for mid-term electricity demand forecasting incorporating weather influences

S. Mirasgedis[a,*], Y. Sarafidis[a], E. Georgopoulou[a], D.P. Lalas[a], M. Moschovits[b], F. Karagiannis[b], D. Papakonstantinou[b]

[a]*National Observatory of Athens, Institute of Environmental Research and Sustainable Development, Lofos Nimfon, Thission, Athens 11810, Greece*
[b]*Public Power Corporation, Arahovis 32, Athens 10681, Greece*

Received 27 April 2004

Abstract

Electricity demand forecasting is becoming an essential tool for energy management, maintenance scheduling and investment decisions in the future liberalized energy markets and fluctuating fuel prices. To address these needs, appropriate forecasting tools for the electricity demand in Greece have been developed and tested. Electricity demand depends on economic variables and national circumstances as well as on climatic conditions. Following the analysis of the time series of electricity demand in the past decade, two statistical models have been developed, one providing daily and the other monthly demand predictions, to estimate medium term demand up to 12 months ahead, utilizing primitive (relative humidity) and derived (heating and cooling degree-days) meteorological parameters. Autoregressive structures were incorporated in both models, aiming at reducing serial correlation, which appears to bias the estimated effects of meteorological parameters on electricity demand. Both modeling approaches show a high predictive value with adjusted R^2 above 96%. Their advantages and disadvantages are discussed in this paper. The effect of the climatic conditions on the electricity demand is then further investigated via predictions under four different scenarios for the weather conditions of the coming year, which include both normal and recently observed extreme behavior.

1. Introduction

The experience of many utilities worldwide has illustrated the influence of weather in energy consumption and especially in electricity demand [1–3]. In the new liberalized environment of

* Corresponding author. Tel.: +30 210 349 0153; fax: +30 210 349 0159.
E-mail address: seba@meteo.noa.gr (S. Mirasgedis).

0360-5442/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.energy.2005.02.016

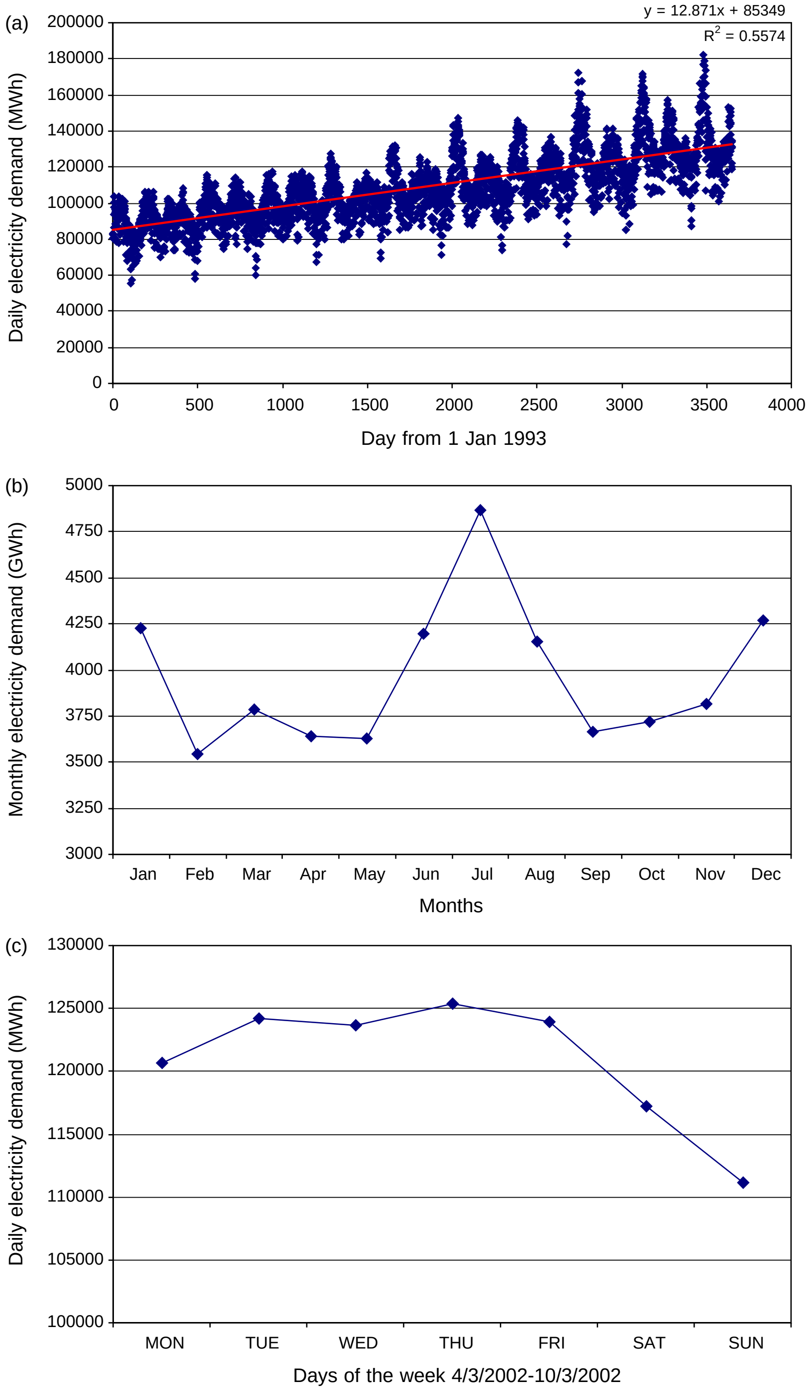

Fig. 1. Evolution of electricity demand for the Greek interconnected power system. (a) Daily electricity demand for the period 1/1/1993–31/12/2002. (b) Monthly electricity demand for the year 2002. (c) Daily electricity demand for the week 4/3/2002–10/3/2002.

electricity markets, it is important to understand and be able to predict the effects of natural variables on load in order to manage effectively the generation and supply of electricity. Knowing the load behavior in advance is crucial in planning, analysis and operation of power systems so as to assure an uninterrupted, reliable, secure and economic supply of electricity.

The effect of changes in natural conditions on electricity load can be significant. Figs. 1 and 2 present an initial exploration of the relationship between electricity demand and temperature for the Greek interconnected electricity system in the last decade. Specifically, in Fig. 1 the plot of electricity demand against time clearly reveals a seasonal and hence weather-sensitive fluctuation (with two peaks each year, in winter and summer), superimposed on a general overall increase in load, which is mainly attributed to the economic development (the GDP increased with an average annual rate of 2.7% in the last decade). Assuming a linear approximation of the demand growth due to economic reasons for the entire period under consideration, Fig. 2 shows a better representation of the relationship between electricity demand and temperature. Specifically, a normalized electricity demand obtained by dividing the actual observed electricity demand with the demand nominally attributed to the economic development and expressed by the linear trend line in Fig. 1 is plotted versus the mean daily temperature.

The variation of electricity demand with temperature is non-linear, increasing both for decreasing and increasing temperatures (reflecting mainly the use of electric heating appliances in winter and air conditioners in summer). For example, during summer, a rise in average daily temperature of 1 °C (say, from 24 to 25 °C) would result in an increase of about 2.6% in electricity consumption. The reference power system seems to be less sensitive to temperature fluctuations in winter, since a fall in average daily temperature of 1 °C (say, from 14 to 13 °C) would result in an increase of about 1.1% in electricity consumption. This is mainly attributed to the fact that final consumers can use a variety of energy sources for heating (e.g. diesel oil, natural gas, electricity, etc.) and practically only electricity for cooling. On the other hand at approximately 18.5 °C, the influence of temperature is minimized and electricity demand is inelastic to temperature changes [4].

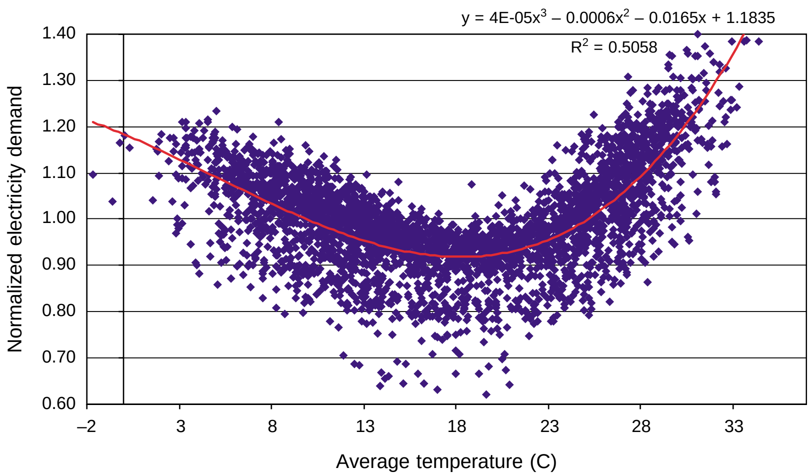

Fig. 2. Relationship between normalized electricity demand (obtained by dividing the actual observed electricity demand with the demand nominally attributed to the economic development) and temperature for the Greek interconnected power system. The cubic regression relationship fitted is based on all the days of the period 1/1/1993–31/12/2002.

The development of methods for appropriate normalization of actual load to given weather conditions and electricity forecasting taking into account specific weather scenarios is of importance to academic and commercial energy economists engaged in energy demand modeling. As a result, a significant number of papers exploring the role of weather variability and change on electricity demand have been published in the past, paying particular attention to short-term load forecasting (from some minutes up to 1 week). A survey of approaches to load forecasting can be found in [5]. The approaches include statistical models [1–3,6–9], neural networks [8,10–14], fuzzy logic [11] and expert systems [11,15]. A number of them [1,2,9,10,12,13] focused on mid- and long-term electricity demand and incorporated climate factors. Of those, some [10,12] only investigated yearly averaged time series and did not include seasonal let alone monthly variation and others either were not able to predict seasons with strong climatic variability [2] or were [1] tested only for a period up to 4 months ahead.

The goal of this study was to generate robust statistical models relating electricity demand to climate parameters with a view to provide electricity demand forecasts for up to 12 months. The analysis focused on the daily and monthly variability and the effects of possible dynamic patterns on the electricity demand—climate relationship. The main contribution of this study is the development of statistical models for mid-term electricity demand forecasting that incorporate climatic and economic factors in sufficient detail to both improve the forecast accuracy and bring out the relative contribution of the influence of various driving factors. This was motivated by the current increased attention to climate change and the evident increase of average yearly temperatures in the last decades as well as anticipated further advances in the decades ahead. The role of the time interval of the forecasting steps in the models developed was also analyzed in detail in the context of this study. It should also be noted that despite the fact that the relationship between electricity demand and climatic conditions in the Greek power sector was investigated in the past for short-term (i.e. hours) demand forecasting [15], this is the first study that focuses on a mid-term (i.e. months) horizon.

The structure of this paper is as follows: Section 2 presents the data used in the study, while Section 3 describes the methodological framework and the structure of the models developed. Section 4 evaluates the predictive power of the models providing forecasts for electricity consumption in the Greek interconnected electricity system for 12 months ahead. Finally, in Section 5, the main findings of the study are summarized and conclusions are drawn.

2. Data

The proposed energy models developed in this paper were applied to the historical data of electricity consumption available for the Greek interconnected power system from 1993 to 2002. This power system is responsible for supplying electricity to almost 10 million people in the mainland of Greece, including the two big urban centers of Athens and Thessaloniki. The data set used comprises hourly data of electric load (in MW) for the entire period under consideration. The daily electricity demand DE_t (in MW h) was estimated as the sum of the 24 values of hourly electricity loads, while the monthly electricity demand ME_t (in MW h) was estimated as the sum of the daily data calculated previously. The daily demand data are the total of all sectors of economic activity (industrial, commercial, residential and agriculture), as sectoral disaggregated data were not available for this time frequency.

Historical weather data required by the proposed models for the period 1993–2002 were obtained from a number of meteorological stations in representative locations in various regions of the country.

Taking into account that the available electricity consumption data used in the study correspond to the whole Greek mainland, while the meteorological parameters vary at different geographical regions (especially between northern and southern Greece), it is necessary to calculate weighted indices for the meteorological parameters, representative of the entire geographical region under consideration. In this study, mainland Greece has been subdivided into two parts, northern and southern. Meteorological data from two first class meteorological stations (so classified according to the criteria set by the World Meteorological Organization) in the main urban centers of Athens (central-southern) and Thessaloniki (northern) was used. The meteorological data from the two regions were weighted on the basis of electricity consumption in the corresponding geographical sub-regions during the last 5 years. A subdivision of the country into five separate climate regions (i.e. Athens, Thessaloniki, Patra, Ioannina and Larisa), did not significantly improve the results of the models (the estimated coefficient of determination was exactly the same) [4].

The experience of many utilities indicates that the weather elements, which influence electricity demand, consist of temperature, humidity, wind and precipitation in a decreasing order of importance [2]. In the context of this study the influence of a considerable number of meteorological parameters on electricity demand of the Greek interconnected system was analyzed, affirming that primary temperature and secondary humidity are the most significant of them, while for the Greek mainland, wind speed and direct solar radiation did not seem to be significant [4]. Therefore, in the present analysis we used the mean daily outdoor temperature (°C) and the mean daily relative humidity (%) as the basic meteorological parameters influencing the electricity demand. It should be noted that relative humidity and temperature are not strictly independent of the former incorporates both absolute humidity (or vapor pressure) and temperature. The selection of relative humidity as an independent variable in the developed models was based on the fact that energy consumers react to perceived comfort, which is better described by relative humidity (and temperature) rather than absolute humidity. The mean temperature (T_t) and relative humidity (HUM_t) of the day t have been calculated as the average of the corresponding hourly values.

As clearly depicted in Fig. 2, the non-linear influence of temperature on electricity demand allows the use of two temperature-derived functions (the heating degree-days and the cooling degree-days), thus separating the winter and summer data. Such a separation helps us build linear models, which keeps structures simple and helps to obtain easily utilizable results. The heating degree-days (HDD_t) and the cooling degree-days (CDD_t) of the day t are estimated on the basis of the following equations

$$\mathrm{HDD}_t = \max(T_{\mathrm{ref}} - T_t, 0) \tag{1}$$

$$\mathrm{CDD}_t = \max(T_t - T_{\mathrm{ref}}, 0) \tag{2}$$

where T_t is the weighted average temperature for the day t and T_{ref} is a reference temperature that should be adequately selected to separate the heat and cold branches of the demand-temperature relationship. In the context of this study, the reference temperature has been selected to be equal to 18.5 °C, which is the temperature at which, as shown in Fig. 2, the influence of temperature is minimized and electricity demand is inelastic to temperature changes. It should be mentioned that we also run the models for various reference temperatures in the interval 16–21 °C, without any discernable improvement in the results.

Similarly, the total number of heating degree-days ($MHDD_t$) and cooling degree-days ($MCDD_t$) for the month t are estimated from the daily heating and cooling degree-days on the basis of the following equations

$$\mathrm{MHDD}_t = \max\left(\sum_{i=1}^{m} \mathrm{HDD}_{t,i} - \sum_{i=1}^{m} \mathrm{CDD}_{t,i}, 0\right) \tag{3}$$

$$\mathrm{MCDD}_t = \max\left(\sum_{i=1}^{m} \mathrm{CDD}_{t,i} - \sum_{i=1}^{m} \mathrm{HDD}_{t,i}, 0\right) \tag{4}$$

where m is the number of days of the month t.

This approach seems to reflect more realistically the consumers' behavior, which follows the temperature fluctuations in using heating or cooling devices, but with an initial delay period. For the months that the mean temperature of all days is below or above the reference temperature (usually in winter and summer), the corresponding monthly heating and cooling degree-days are essentially estimated as the sum of the daily values. For months that the mean daily temperature fluctuates below and above the reference temperature (usually in autumn and spring), the monthly heating and cooling degree-days are estimated on the basis of the net difference of the sum of the daily heating and cooling degree-days. It should be noted that such overlaps between HDDs and CDDs occur only in 2–4 months per year, when the recorded average daily temperatures are slightly above or below the T_{ref}. With this modification, an attempt is made to model more realistically the behavior of consumers, who usually do not use air conditioning systems until the average daily temperature exceeds 19.5 °C (i.e. only one degree above the T_{ref}) or heating devices until the average daily temperature dips below 17.5 °C (i.e. only one degree below the T_{ref}).

3. Methodological framework

In the context of this study two multiple regression models were developed and comparatively evaluated, differentiated on the time basis of their structure. The former, hereafter named as the *daily model*, uses the day as the temporal basis of analysis. Daily average temperature was determined to estimate heating and cooling degree-days for each day of the sample, which were related together with average humidity to the daily electricity demand (dependent variable). The latter, hereafter named as the *monthly model*, uses the month as the basic unit for analysis. Monthly electricity demand (dependent variable) data were related to the total number of heating and cooling degree-days of the corresponding month. Both daily and monthly electricity demand data have been transformed by taking natural logarithms to reduce the impact of heteroskedasticity that could be present due to the large amount of data and its high temporal frequency [1]. The existence of heteroskedasticity violates the 'constant variance' assumption in regression analysis. In other words, the regression models assume that the residuals have the same variance throughout. As clearly depicted in Fig. 1(a), the annual seasonal variation of the reference electricity demand data increases as the level of the series increases, implying that existence of heteroscedasticity cannot be excluded which was the reason for utilizing an appropriate mathematical transformation as suggested in [16], namely a logarithmic one.

3.1. The daily model

The explanatory variables used in the daily model are defined as follows:

- The *time* (t) aiming at taking into account the strong long-term trend in the daily electricity demand, which is clearly depicted in Fig. 1. This trend is mainly related to social, economic, technological and demographic factors, which affect the economic development of the region under consideration. The trend shows a continuous growth pattern reflected in improved living standards, accelerating purchase of air conditioning equipment and population increase mostly because of immigration patterns. Assuming a linear approximation of the economic development in the model appeared to be an appropriate compromise between accuracy, as expressed by the correlation coefficient, and simplicity [2]. The incorporation of more complex modifications for representing the economic development did not significantly improve the results [4].
- The *heating* and *cooling degree-days* (HDD_t and CDD_t) of the corresponding day t that electricity demand is assumed.
- The *heating* and *cooling degree-days of the two previous days* (HDD_{t-1}, HDD_{t-2}, CDD_{t-1} and CDD_{t-2}) of the reference day t. This consideration suggests that the degree-day variables could have, in addition to a contemporary influence on the electricity demand, a lagged effect. It was assumed that the lagged effect can be relevant only over short time periods and thus the degree-days of the last two days were taken into account.
- The *average relative humidity* (HUM_t) for the reference day t.
- Six dummy variables (D_{it}) aiming at representing the significant *daily variability of the electricity demand*. The index i takes values in the interval [2,7] representing correspondingly all days in the week ($i=2$ for Tuesday, $i=3$ for Wednesday, …, $i=7$ for Sunday) except the base day of Monday. It should be noted that any day of the week could have been selected as the base day of the model without affecting the results of the analysis. Each dummy variable has only two allowable values, 0 or 1. For example, if the day t is Wednesday, then the variable D_3 takes the value 1 while all other variables D_i take the value 0. In case that the day t is Monday, then all the variables D_i take the value 0. In general $(s-1)$ dummy variables are used to denote s different periods in regression models because otherwise we would encounter the problem of multicollinearity, affecting the stability of the regression coefficients [16].
- Four dummy variables cover anomalous events in electricity demand related to *holidays or days near a holiday*. Electricity consumption, mainly in the industrial and services sectors, decrease appreciably during holidays. Thus, the following dummy variables have been introduced in the model: (a) CH_t, which takes the value of 1 if day t is a fixed holiday (e.g. Christmas) and 0 otherwise; (b) CH_{t-1}, which takes the value of 1 if day t corresponds to a day following a fixed holiday and 0 otherwise; (c) VH_t, which takes the value of 1 if day t is a moveable holiday (e.g. Easter) and 0 otherwise; and (d) PP_t, which takes the value of 1 if t is the last day of the year (the 31st of December) and 0 otherwise (this day is usually the day that the winter peak of the reference system is recorded).
- Eleven dummy variables (M_{jt}) in the model account for the *monthly seasonality of electricity demand* not related to the weather conditions. Monthly seasonality of electricity demand is influenced by a number of economic and social factors such as summer vacations, the seasonal character of some industrial activities (e.g. food industries), etc. The index j values in the

interval [2,12] representing correspondingly all months in a year ($j=2$ for February, $j=3$ for March,…, $j=12$ for December) except the base month of January (again the selection of January as the base month of the model is arbitrary). The variable M_{jt} takes the value of 1 if the t observation belongs to the month j and 0 otherwise. Again we use one dummy variable less than the number of periods (12 months per year) to avoid multicollinearity problems.

The equation of the daily model is finally given by

$$\log(\mathrm{DE}_t) = c + at + k\mathrm{HDD}_t + l\mathrm{CDD}_t + k_1\mathrm{HDD}_{t-1} + l_1\mathrm{CDD}_{t-1} + k_2\mathrm{HDD}_{t-2} + l_2\mathrm{CDD}_{t-2} + n\mathrm{HUM}_t + \sum_{i=2}^{7} d_i D_{it} + \sum_{j=2}^{12} f_j M_{jt} + p\mathrm{CH}_t + s\mathrm{CH}_{t-1} + q\mathrm{PP}_t + r\mathrm{VH}_t + e_t \quad (5)$$

where c, α, k, l, k_1, l_1, k_2, l_2, n, d_i (i from 2 to 7), f_j (j from 2 to 12), p, s, q and r are the coefficients to be estimated from the regression analysis and e_t is the residual term.

Using $n=3652$ daily observations (covering the period 1/1/1993 up to 31/12/2002), the daily multiple linear regression coefficients were estimated and the results are presented in Table 1. The regression run gave an adjusted R^2 of 94.6% showing that the model has a high predictive power. According to the results of the analysis, the constant c, and the time coefficient α of the linear time variation are very significant underlying that the level of economic activity primarily affects the total electricity demand. Among the weather variables, the temperature-derived parameters are also significant while relative humidity seems to be of minor importance. The estimated coefficients for CDDs are larger than that of HDDs, showing that the reference electric system is more sensitive to the high temperatures. Sunday and Saturday are the days of the week with the least demand, while Tuesday and Thursday are the days presenting the highest electricity consumption. This latter observation is partially attributed to the fact that on Tuesday and Thursday, store closing hours extend to 20:00 h (as opposed to 15:00 h on Monday and Wednesday), which clearly affects the total electricity demand. The effect of the extended hours on Friday is mitigated by the weekend exodus. Examining monthly variability of electricity consumption, one notes that all the estimated coefficients are significant, with July, November and December presenting a relatively higher electricity consumption compared to the rest. From the dummy variables related to holidays or days near a holiday, only the last day of the year seems to have no significance.

3.2. The monthly model

The explanatory variables used in the monthly model are defined as follows:

- ☐ The *time* (t) again as an approximation of the long-term trend in electricity demand attributed mainly to economic development.
- ☐ The total number of *monthly heating and cooling degree-days* (MHDD_t and MCDD_t) estimated for the month t.

- ☐ As in the daily model eleven dummy variables (M_{jt}), which are aimed at representing the *monthly seasonality of electricity demand* and are not related to the weather conditions.
- ☐ A new variable H_t represents the *number of non-working days* (Sundays, Saturdays and holidays) during the month t. As already mentioned previously electricity consumption in the industrial and services sectors is expected to be lower during these days.

The equation of the monthly model is given by

$$\log(\mathrm{ME}_t) = c + at + k\mathrm{MHDD}_t + l\mathrm{MCDD}_t + \sum_{j=2}^{12} f_j M_{jt} + pH_t + e_t \tag{6}$$

where c, α, k, l, f_j (j from 2 to 12) and p are the coefficients to be estimated from the regression analysis and e_t the residual term.

Using data for the same period 1/1/1993–31/12/2002, the coefficients in Eq. (6) were obtained and the results are also presented in Table 1. The adjusted R^2 for the monthly model was estimated at 98.5% showing a very high predictive power. The constant, the time, the weather variables and the number of non-working days are highly significant. As in the daily model, the estimated coefficient for MCDDs is larger than the MHDDs that show the strong influence of weather conditions in electricity demand during summer. The results also show the importance of the monthly variability in the electricity consumption. The coefficients for July, October, November and December are not significant. This implies that electricity consumption in these months is similar to that of January. All the dummy variables for the other months are negative and significant, which shows that less electricity is consumed during the corresponding months than in January.

3.3. The problem of autocorrelation

Despite the fact that both multiple regression models developed in the context of this study are characterized by high R^2 values, they still show large serial correlation. The Durbin–Watson statistic (D–W), which is widely used for testing the serial correlation or the independence of residuals at lag 1 in regression models, is estimated at 0.963 for the daily model and at 1.469 for the monthly model (Table 1). These figures are lower than 2, indicating that in both models positive autocorrelation is present (the D–W statistic ranges in value from 0 to 4, with an intermediate value of 2, which indicates the absence of autocorrelation at lag 1). According to econometric regression theory, if the residuals are not independent (or in other words the errors are serially correlated), the use of the F- and t-tests and confidence intervals is not strictly valid and the estimate of the coefficients may be unstable [16]. It is therefore obvious that the autocorrelation problem can lead to misleading results.

One of the common methods to reduce the serial correlation observed in econometric regression is the incorporation of an autoregressive structure in the error term [17,18]. This approach assumes that the current values of the error term (e_t) can be expressed as a finite number of previous errors. Mathematically it can be represented as follows

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \ldots + \phi_p e_{t-p} + \varepsilon_t \tag{7}$$

Table 1
Values of the coefficients of the daily and monthly multiple regression models, together with statistical results based on data for the period 1/1/1993–31/12/2002

Variable	Daily model		Monthly model	
	Coefficients	t-Statistic	Coefficients	t-Statistic
c	4.904	1529.898	3.410	315.974
t	5.113×10^{-05}	190.757	1.545×10^{-03}	68.789
HDD_t	3.296×10^{-03}	13.354		
CDD_t	5.474×10^{-03}	15.002		
HDD_{t-1}	6.603×10^{-04}	1.855		
CDD_{t-1}	1.229×10^{-03}	2.344		
HDD_{t-2}	2.260×10^{-03}	9.212		
CDD_{t-2}	2.383×10^{-03}	6.583		
$MHDD_t$			2.079×10^{-04}	9.657
$MCDD_t$			3.494×10^{-04}	9.246
HUM_t	-3.844×10^{-05}	−1.168		
D_{2t}	1.096×10^{-02}	10.497		
D_{3t}	1.055×10^{-02}	10.096		
D_{4t}	1.188×10^{-02}	11.363		
D_{5t}	1.009×10^{-02}	9.691		
D_{6t}	-1.578×10^{-02}	−15.163		
D_{7t}	-5.707×10^{-02}	−54.865		
M_{2t}	-4.456×10^{-03}	−3.165	-4.295×10^{-02}	−10.374
M_{3t}	-1.274×10^{-02}	−8.942	-9.975×10^{-03}	−2.464
M_{4t}	-1.801×10^{-02}	−10.676	-2.939×10^{-02}	−5.103
M_{5t}	-2.168×10^{-02}	−11.015	-1.680×10^{-02}	−2.114
M_{6t}	-4.104×10^{-03}	−1.745	-2.405×10^{-02}	−2.196
M_{7t}	1.685×10^{-02}	6.348	5.577×10^{-03}	.410
M_{8t}	-1.281×10^{-02}	−4.958	-2.380×10^{-02}	−1.841
M_{9t}	-2.313×10^{-02}	−10.843	-4.048×10^{-02}	−4.389
M_{10t}	-7.144×10^{-03}	−3.904	3.189×10^{-03}	.446
M_{11t}	3.086×10^{-03}	2.015	-7.903×10^{-03}	−1.544
M_{12t}	6.796×10^{-03}	4.943	6.364×10^{-03}	1.705
CH_t	-6.648×10^{-02}	−34.232		
CH_{t-1}	-2.410×10^{-02}	−12.415		
PP_t	-5.711×10^{-03}	−1.059		
VH_t	-8.294×10^{-02}	−33.259		
H_t			-2.871×10^{-03}	−3.484
Statistics				
Adjusted R^2	94.6%		98.5%	
Regression std error	0.0168		0.0081	
Durbin–Watson	0.963		1.469	

where $\phi_1, \phi_2,\ldots,\phi_p$ are constant parameters, p is the order of the autoregressive structure and ε_t is a white noise error term.

In order to reduce the autocorrelation observed in the models investigated here, alternative autoregressive structures were evaluated using common diagnostic tests. Specifically, the models developed were tested using both the Akaike Information Criterion and the Schwarz Criterion. A fourth-order autoregressive structure was eventually chosen to be introduced into the daily model while for the monthly model a first-order autoregressive process was selected.

Table 2 presents the results of the analysis for both autoregressive models under consideration. For the daily model, the adjusted R^2 increased to 96.4% (from 94.6%), which indicates an improvement of

Table 2
Values of the coefficients of the daily and monthly autoregressive models together with statistical results based on data for the period 1/1/1993–31/12/2002

Variable	Daily model		Monthly model	
	Coefficients	t-Statistic	Coefficients	t-Statistic
c	4.894	1150.787	3.415	340.174
t	5.100×10^{-05}	63.746	1.543×10^{-03}	53.148
HDD_t	2.789×10^{-03}	13.794		
CDD_t	4.897×10^{-03}	16.506		
HDD_{t-1}	1.194×10^{-03}	4.953		
CDD_{t-1}	1.975×10^{-03}	5.597		
HDD_{t-2}	1.422×10^{-03}	7.111		
CDD_{t-2}	1.174×10^{-03}	4.015		
$MHDD_t$			1.961×10^{-04}	9.384
$MCDD_t$			3.383×10^{-04}	8.381
HUM_t	1.690×10^{-04}	5.069		
D_{2t}	1.104×10^{-02}	15.785		
D_{3t}	1.079×10^{-02}	13.668		
D_{4t}	1.185×10^{-02}	15.311		
D_{5t}	1.008×10^{-02}	13.102		
D_{6t}	-1.585×10^{-02}	−20.201		
D_{7t}	-5.722×10^{-02}	−82.416		
M_{2t}	-2.490×10^{-03}	−0.881	-4.373×10^{-02}	−12.123
M_{3t}	-9.105×10^{-03}	−2.862	-1.102×10^{-02}	−2.808
M_{4t}	-1.359×10^{-02}	−3.964	-3.191×10^{-02}	−5.637
M_{5t}	-1.784×10^{-02}	−4.829	-1.974×10^{-02}	−2.521
M_{6t}	3.305×10^{-03}	0.810	-2.545×10^{-02}	−2.275
M_{7t}	1.758×10^{-02}	3.998	5.271×10^{-03}	0.375
M_{8t}	4.827×10^{-04}	0.112	-2.441×10^{-02}	−1.829
M_{9t}	-1.784×10^{-02}	−4.593	-4.263×10^{-02}	−4.600
M_{10t}	-8.091×10^{-03}	−2.274	7.600×10^{-07}	0.000
M_{11t}	-3.260×10^{-05}	−0.010	-9.829×10^{-03}	−1.977
M_{12t}	5.337×10^{-03}	1.775	6.067×10^{-03}	1.854
CH_t	-6.245×10^{-02}	−40.341		
CH_{t-1}	-1.905×10^{-02}	−12.529		
PP_t	1.326×10^{-02}	3.132		
VH_t	-7.943×10^{-02}	−33.520		
H_t			-2.910×10^{-03}	−4.169
AR(1)	4.631×10^{-01}	27.984	2.728×10^{-01}	2.829
AR(2)	6.795×10^{-02}	3.731		
AR(3)	1.018×10^{-01}	5.558		
AR(4)	9.465×10^{-02}	5.711		
Statistics				
Adjusted R^2	96.4%		98.8%	
Regression std error	0.0138		0.0078	

the model's predictive power. The adjusted R^2 for the monthly autoregressive model increased marginally, reaching a value of 98.8% (from 98.5%). The marginal improvement in the monthly model as compared to the daily one is clearly due to the loss of 'memory' on a monthly basis as compared to the daily basis. This is also evident in the much higher adjusted coefficient of determination (R^2) without strong autocorrelation of the monthly model, which can hardly be expected to improve further.

Inevitably the incorporation of autoregressive structures in both models affected the results presented in previous paragraphs. In the daily autoregressive model all the coefficients associated with weather variables now turn out to be significant, including relative humidity. However, the value of the coefficients of the variables HDD_t, CDD_t, HDD_{t-2} and CDD_{t-2} has decreased while the ones for the variables HDD_{t-1} and CDD_{t-1} have increased, indicating an increased significance of the previous day temperature on the total electricity demand. In any case the coefficient for CDD_t is still higher than the accompanying the HDD_t value. The dummy variables referred to the daily seasonality including those representing holidays or days near a holiday, are all significant, in contrast to almost half of the coefficients for the dummy variables related to the monthly variability, which implies that electricity consumption for the corresponding months (February, June, August, November and December) tends to be closer to that of the base month (January).

The strong autoregressive nature of the daily demand for electricity is clearly discernable from the resulting increase that the model predicts for a nominal 1 °C in the mean temperature of the day, assuming that its absolute value is above 18.5 °C. For this case the increase in demand is 1.1%, but if now a similar increase is also present for the two previous days of the day t, then the daily electricity demand increases by 1.9%. Similarly, if the mean temperature is below 18.5 °C and on day t there is a decrease of 1 °C, the resulting change in demand is 0.6%, while this increase reaches 1.3% in case that a 1 °C temperature decrease is also recorded for the two previous days. This seems to reflect a tendency of the population to retain the memory of the overall weather conditions for one and possibly two previous days, which is enhanced by the thermal capacity of the building stock, especially the one to be found in Greece where reinforced cement and brick is the dominant construction material.

In the monthly autoregressive model all the coefficients reflecting economic activity (constant, time, autoregressive structure, number of non-working days) and weather ($MHDD_t$ and $MCDD_t$) are significant. As in the daily model the coefficient for cooling degree-days is higher than that accompanying the heating degree-days value. It should be noted that the dummy variables for July, August, October, November and December are not significant.

4. Results and discussion

4.1. Evaluation of the predictive capability of the models

To evaluate the predictive capability of the models, the available data set of historical hourly electricity and weather data for the period 1/1/1993–31/12/2002, which were previously used to estimate the coefficients presented in Tables 1 and 2 were now split and the data for the period 1/1/1993–31/12/2001 were used in training the models as given in Eqs. (5)–(7). It should be noted that this training procedure gave slightly different coefficient values from those presented in Tables 1 and 2 as might be expected, since the 2002 data are now excluded from the training. The weather data for the year 2002 was then used to test the proposed models in forecasting electricity demand for this year. Table 3 shows the actual and forecast monthly electricity demands by both models.

For 2002 the proposed monthly autoregressive model follows the actual electricity demand very closely, whereas the daily autoregressive model is characterized by slightly higher deviations. Specifically, the monthly model forecasts the electricity demand with very high accuracy for the entire period under consideration. The maximum prediction error of 2.7% appears for the month of August. However, the estimated deviations between the actual and forecasted monthly electricity demands are less than 1% for 7 out of 12 months of the reference year, indicating the very high predictive power of the model. On an annual basis the projected electricity demand for the entire year 2002 is very close to the actual data.

The daily autoregressive model also forecasts the total electricity demand for 2002 with very high accuracy (see Table 3). However, examining the results obtained for each month separately, it should be noted that the model gives a maximum error of 4.6% (also in August). The number of months of the reference year with a deviation less than 1% between forecasted and actual data is only three as opposed to seven in the monthly model. The results clearly indicate that the proposed model forecasts the monthly electricity demand with less accuracy for the second half of the year (with the exception of December the estimated deviations for all other months of this period were above 1%). This is mainly attributed to the time increment of the model, which affects the capability of the autoregressive structure to adjust the results.

At this juncture, it is helpful to review the nature of the bias and how the autoregressive structure rectifies the error variance inflation in the forecasting process. For simplicity, it is assumed that the regression model at time, t, is

$$y_t = a + bx_t + e_t \tag{8a}$$

Table 3
Forecasting electricity demand for the year 2002 using the daily and monthly autoregressive models for the training set 1993–2001

Month	Actual electricity demand (GW h)	Daily model		Monthly model	
		Electricity demand (GW h)	Error (%)	Electricity demand (GW h)	Error (%)
January	4224	4143	−1.9	4236	0.3
February	3544	3591	1.3	3608	1.8
March	3785	3805	0.5	3767	−0.5
April	3638	3638	0.0	3571	−1.8
May	3630	3592	−1.1	3652	0.6
June	4198	4125	−1.7	4121	−1.8
July	4866	4683	−3.7	4799	−1.4
August	4156	4347	4.6	4270	2.7
September	3664	3734	1.9	3667	0.1
October	3721	3784	1.7	3734	0.4
November	3815	3859	1.1	3848	0.9
December	4266	4257	−0.2	4256	−0.3
Total	47,508	47,558	0.1	47,529	0.0

with

$$e_t = \rho e_{t-1} + \varepsilon_t \tag{8b}$$

where y_t is the independent variable, x_t the explanatory variable, a, b and ρ the coefficients of the regression and ε_t a white noise error term. The forecasting process is then described by the following Eqs.

$$y_{t+1} = a + bx_{t+1} + \rho e_t + \varepsilon_{t+1} \tag{9}$$

$$y_{t+2} = a + bx_{t+2} + \rho^2 e_t + \rho e_{t+1} + \varepsilon_{t+2}$$

$$y_{t+h} = a + bx_{t+h} + \rho^h e_t + \rho^{h-1} e_{t+1} + \ldots + \varepsilon_{t+h}$$

To predict h periods ahead, the best predictor in terms of the current error at the time t is $\rho^h e_t$. All the other predictors $\rho^{h-1} e_{t+1}$, $\rho^{h-2} e_{t+2} \ldots, \rho e_{t+h-1}$ should not be taken into account since the error terms $e_{t+1}, e_{t+2}, \ldots, e_{t+h-1}$ are unknown. As h increases, the amount of error incrementally added to the forecast exponentially attenuates until an asymptote is approximated [19].

It is obvious that the daily model, using the day as the basic unit for analysis, should forecast 365 periods ahead for estimating the future electricity demand up to a period of 1 year. As a result of the process described previously, the autoregressive structure affects the model estimates for future electricity demand for some periods ahead (usually a number of 100–150 covering the first 3–5 months), while its role is practically insignificant for the rest of the period. On the other hand, using the monthly autoregressive model for forecasting the electricity demand for a period of 1 year, the model needs to predict only 12 periods ahead. Thus, the role of autoregressive structure is important for the entire period under consideration improving the results of the analysis.

To gauge this limitation, especially of the daily model, the results of forecasts with training data for the period 1/1/1993–31/12/2001 were compared with results from an alternative approach where the daily model projects the electricity demand for the month m of the year 2002 using as training set all the data from 1/1/1993 up to the last day of the month $m-1$ of the year 2002 (12 runs were elaborated for projecting the monthly electricity demand for the entire 2002). The results are shown in Fig. 3. With this latter approach the role of the autoregressive structure is present for all the months of the year examined thus showing obvious improvements in the results of the model particularly during the second half of the year.

4.2. *Electricity demand forecasting in the Greek interconnected system for the year 2003*

In the proposed models, demand forecasts are direct functions of weather and other variables. A change in any of these variables will therefore affect the forecast; the extent of change is expected to depend on the degree of correlation and the contribution of these variables to the forecast. From the planning point of view, this is especially important as it gives the manager of the electrical system a feeling about the trustworthiness of the forecast and the influence of weather on the total electricity demand. Given the nature of the model, for medium term planning purposes, one needs to know rather detailed meteorological information for the year or years ahead, which at present are not routinely

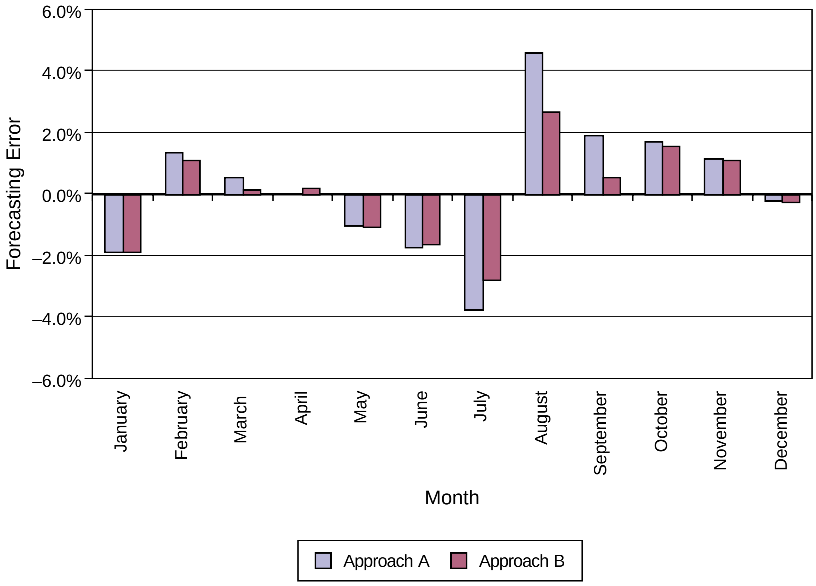

Fig. 3. The error of the demand forecasts for the year 2002 with the daily model using training data from the period 1/1/1993–31/12/2001 (approach A), and extending the period from 1/1/1993 up to the previous month of 2002 (approach B).

available. Yet, it is of considerable planning value to at least bracket the expected change in demand. Appropriate meteorological scenarios can be used to accomplish this.

To this end, four alternative scenarios (S1, S2, S3a and S3b) were developed to be used with the models, on the basis of different meteorological conditions, aiming at representing both typical and extreme potential weather conditions for the coming year 2003. The meteorological parameters of the two geographical sub-regions have been weighted on the basis of the electricity consumption in these two sub-regions during the last 5 years. Specifically:

- *Scenario S1* uses a typical meteorological year of the 22-year period 1977–1998, defined according to a modified Festa–Ratto methodology [20,21]. This methodology generates a set of 8760 hourly values of all relevant meteorological parameters that comprise a typical year for simulation purposes. The values are actual measured values from complete months, chosen from a set of the corresponding months for a period of at least 10 years. Each particular month, says March, is selected from all the same months of the years of the period (i.e. all 'Marches') by comparing the mean monthly values for the month of each year in the period with those of the mean monthly values of the set of the months in question for the whole period. The specific parameters to be used as selection criteria and especially their combination that have been proposed by the Festa–Ratto approach are both the mean and the standard deviation as well as the Kolmogorov–Smirnov parameter. This method seems to be one of the most suitable for predicting typical meteorological years for energy applications according to an evaluation carried out in [21]. This scenario should provide an estimation of the future electricity demand for the Greek interconnected system, assuming that weather conditions in 2003 will be closer to the 20-year average.

☐ *Scenario S2* uses a typical meteorological year (defined again according to the modified Festa–Rato methodology) of the 10-year period 1993–2002 and estimates the future electricity demand for the system assuming that weather conditions in 2003 will be similar to the average weather conditions of the last decade.

☐ *Scenarios S3a and S3b* explore the influence of extreme meteorological conditions on the future electricity demand of the reference power system. Specifically, *scenario S3a* uses a meteorological year developed on the basis of the coldest winter months and the warmest summer months of the last decade (e.g. the January with the maximum number of HDDs among all the other Januaries of the decade 1993–2002 was selected for developing this reference meteorological year). This scenario gives a measure of the maximum electricity demand that could be expected for the power system in 2003. *Scenario S3b* uses a reference meteorological year developed on the basis of the three most extreme meteorological years of the last decade. To this end the total number of HDDs and CDDs for each year of the period 1993–2002 were estimated and then the 3 years with the highest cumulative

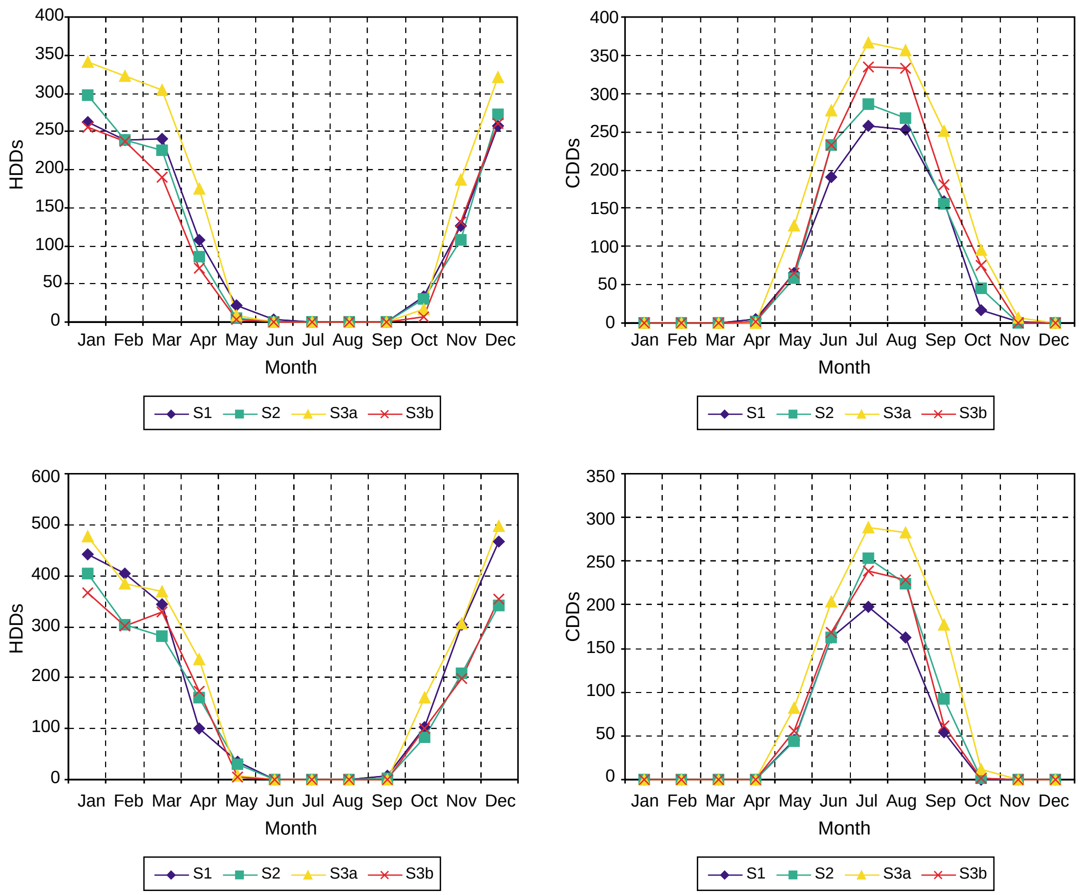

Fig. 4. Variation of the number of HDDs and CDDs on the basis of the four meteorological scenarios developed in the context of this study for the two major climatic sub-regions of Greece to study expected electricity demand.

sum of HDDs and CDDs were selected. For each day of the year, the three values of each meteorological parameter were averaged to produce the reference meteorological year.

Fig. 4 presents the variation of HDDs and CDDs in the two climatic sub-regions according to the four scenarios developed. Tables 4 and 5 show the results of electricity demand forecasts for the year 2003 using the autoregressive models developed in the context of this study.

As shown in Table 4, the daily model estimates a total annual increase of electricity demand between 4.3 and 8%, indicating a significant increase under all weather scenarios. The results of the scenarios S1, S2 and S3b are quite similar leading to an annual increase of electricity demand between 4.3 and 4.7%. On the other hand, it is obvious that the extreme weather conditions incorporated into scenario S3a lead to significantly higher electricity consumption for both heating and cooling purposes. The results of the monthly model are summarized in Table 5 and show an annual increase of electricity demand in the Greek interconnected system between 3.9 and 8.4%.

From the results of the analysis, it is obvious that weather variability affects significantly the total electricity demand and therefore the associated economic expenditures for electricity generation and consumption. The difference between the weather conditions of scenarios S1 (normal) and S3a (extreme) results in a change in the total electricity demand for the year 2003 between 1762 GW h (daily model) and 2135 GW h (monthly model). Assuming a typical average electricity cost of 40 €/MW h to the utility, the extreme weather scenario (S3a) leads to additional annual expenditures for electricity generation in the range of € 70–85 million, compared to the corresponding results of the scenario S1.

The results obtained from the two models are quite similar also on a monthly basis. Given the 4 weather scenarios developed, the models gave predictions for $4\times12=48$ periods (months). For this

Table 4
Forecasting electricity demand for 2003 using the daily autoregressive model and estimation of the percentage increase of electricity demand compared to 2002 levels, assuming temperature, humidity and other meteorological data as provided in the four typical meteorological years synthesized for this study

Month	Electricity demand in 2002	Forecasting electricity demand for 2003							
		S1		S2		S3a		S3b	
	GW h	GW h	(%)	GW h	(%)	GW h	(%)	GW h	(%)
January	4224	4254	0.7	4272	1.1	4360	3.2	4211	−0.3
February	3544	3915	10.5	3854	8.8	3979	12.3	3848	8.6
March	3785	4124	9.0	4063	7.3	4226	11.6	4057	7.2
April	3638	3690	1.4	3710	2.0	3832	5.3	3695	1.5
May	3630	3849	6.0	3785	4.3	3923	8.1	3810	4.9
June	4198	4175	−0.6	4254	1.3	4363	3.9	4255	1.4
July	4866	4646	−4.5	4752	−2.3	4929	1.3	4824	−0.9
August	4156	4393	5.7	4483	7.9	4679	12.6	4594	10.5
September	3664	3913	6.8	3942	7.6	4162	13.6	3958	8.0
October	3721	3926	5.5	3960	6.4	4022	8.1	3967	6.6
November	3815	4103	7.5	4026	5.5	4176	9.4	4054	6.3
December	4266	4541	6.4	4481	5.0	4639	8.7	4480	5.0
Total	47,508	49,528	4.3	49,582	4.4	51,290	8.0	49,753	4.7

Table 5
Forecasting electricity demand for 2003 using the monthly autoregressive model and estimation of the percentage increase of electricity demand compared to 2002 levels, using the four typical meteorological years synthesized for this study

Month	Electricity demand in 2002	Forecasting electricity demand for 2003							
		S1		S2		S3a		S3b	
	GW h	GW h	(%)	GW h	(%)	GW h	(%)	GW h	(%)
January	4224	4325	2.4	4345	2.9	4453	5.4	4266	1.0
February	3544	3925	10.7	3864	9.0	4011	13.2	3861	8.9
March	3785	4099	8.3	4041	6.8	4191	10.7	4026	6.4
April	3638	3672	0.9	3687	1.3	3832	5.3	3677	1.1
May	3630	3711	2.2	3732	2.8	3925	8.1	3786	4.3
June	4198	4145	−1.3	4247	1.2	4393	4.6	4252	1.3
July	4866	4707	−3.3	4851	−0.3	5098	4.8	4951	1.8
August	4156	4272	2.8	4377	5.3	4654	12.0	4533	9.1
September	3664	3878	5.8	3914	6.8	4209	14.9	3934	7.4
October	3721	3964	6.5	3914	5.2	3887	4.5	3917	5.3
November	3815	4115	7.8	4033	5.7	4185	9.7	4055	6.3
December	4266	4568	7.1	4501	5.5	4677	9.6	4493	5.3
Total	47,508	49,379	3.9	49,507	4.2	51,514	8.4	49,752	4.7

sample, a high correlation coefficient is obtained between the predicted values of the two models (98.8%), which shows a considerable degree of agreement.

5. Conclusions

Mid-term (i.e. monthly and yearly) electricity demand forecasting in power systems is a complicated task because it is affected directly or indirectly by various factors primarily associated with the economy and the weather. In this work, two multiple regression models, also incorporating autoregressive structures to reduce serial correlation, have been developed linking electricity demand directly to climatic conditions as well as seasonal activity patterns. The two models differ on the interval (day or month) used to march forward in time.

The models are shown to provide high accuracy forecasts for a number of months ahead, which extend well over a 1-year period, assuming that reasonably weather forecasts are available for this period. This is very useful in planning fuel procurement, scheduling unit maintenance, and imports. Specifically, the daily autoregressive model is capable of forecasting monthly electricity demand with maximum error of less than 4.6% for a year in advance and a maximum error of less than 2.8% for a month in advance. Even without accurate forecasts, planning guidance can be gained through the investigation of expected cooler or warmer than typical years. Such expected long-term variation forecasts are now becoming available from the meteorological services in a number of countries. The temperature of the day that electricity demand is projected, the temperature of the two previous days and the relative humidity have been found to be the most important weather parameters that affect electricity consumption in the Greek interconnected power system.

The monthly autoregressive model is capable of forecasting monthly electricity demand a year in advance with maximum error of less than 2.7% (again it is assumed that accurate weather forecasts are available to the modelers). The predictive power of the model is better compared to that of the daily model because of the greater effectiveness of the autoregressive contribution that stems from the much smaller number (12 rather than 365) of steps ahead required. However, it should be noted that it models the influence of weather in electricity demand in a more aggregated way and thus may not account well for the influence of unusual or extreme weather on electricity consumption.

The two models were finally used for estimating the increase of electricity demand in the reference power system for 2003 under alternative meteorological scenarios. A significant increase in electricity demand should be expected for 2003 compared to 2002, if the climatological conditions assumed in the scenarios utilized come to pass. Both models showed that the increase in electricity demand compared to the previous year would range from 3.9% in case of a moderate meteorological year that is close to the 22-year historical average up to 8.4% if extreme meteorological conditions were recorded both in summer and winter months. The influence of the extreme weather appears to amount to about 4–4.5% above expected demand increase.

References

[1] Pardo A, Meneu V, Valor E. Temperature and seasonality influences on Spanish electricity load. Energy Econ 2002;24: 55–70.
[2] Robinson P. Modeling utility load and temperature relationships for use with long-lead forecasts. J Appl Meteorol 1997; 36:591–8.
[3] Sailor D, Munoz R. Sensitivity of electricity and natural gas consumption to climate in the USA—methodology and results for eight states. Energy 1997;22(10):987–98.
[4] Lalas DP, Mirasgedis S, Sarafidis Y, Georgopoulou E, Adamopoulou A. Impacts of environmental restrictions on the structure of the electricity sector. Technical report prepared by National Observatory of Athens for the Greek Public Power Corporation, Lofos Nimfon, Thission, 11810 Athens; 2003.
[5] Lotufo ADP, Minussi CR. Electric power systems load forecasting: a survey. Paper BPT99-028-25 accepted for presentation at the IEEE Power Tech'99 Conference, Budapest, Hungary; 1999.
[6] Barakat EH, Al Rashed SA. Analysis of the impact of dynamic events on future power requirements. Electric Power Syst Res 1996;37:137–42.
[7] Tripathy SC. Demand forecasting in a power system. Energy Convers Manage 1997;38(4):1475–81.
[8] Al-Saba T, El-Amin I. Artificial neural networks as applied to long-term demand forecasting. Artif Intell Eng 1999;13: 187–97.
[9] Sforna M. Searching for the electric load-weather temperature function by using the group method of data handling. Electric Power Syst Res 1995;32:1–9.
[10] Kermanshahi B, Iwamiya H. Up to year 2020 load forecasting using neural nets. Electrical Power Energy Syst 2000;24: 789–97.
[11] Tzafestas S, Tzafestas E. Computational intelligence techniques for short-term electric load forecasting. J Intell Robo Syst 2001;31:7–68.
[12] Ringwood JV, Bofell D, Murray FT. Forecasting electricity demand on short, medium and long time scales using neural networks. J Intell Rob Syst 2001;31:129–47.
[13] Islam S, Al-Alawi S, Ellithy K. Forecasting monthly electric load and energy for a fast growing utility using an artificial neural network. Electric Power Syst Res 1995;34:1–9.
[14] Chen CS, Tzeng YM, Hwang JC. The application of artificial neural networks to substation load forecasting. Electric Power Syst Res 1996;38:153–60.

[15] Petridis V, Kehagias A, Petrou L, Bakirtzis A, Kiartzis S, Panagiotou H, et al. A Bayesian multiple models combination method for time series prediction. J Intell Rob Syst 2001;31:69–89.
[16] Makridakis S, Wheelwright S, Hyndman R. Forecasting methods and applications. 3rd ed. New York: Wiley; 1998.
[17] Peirson J, Henley A. Electricity load and temperature: issues in dynamic specification. Energy Econ 1994;16:235–43.
[18] Taylor J, Buizza R. Using weather ensemble predictions in electricity demand forecasting. Int J Forecasting 2003;19: 57–70.
[19] Yaffee R, McGee M. Introduction to time series analysis and forecasting with applications of SAS and SPSS. New York: Academic Press; 2000.
[20] Festa R, Ratto CF, DeGol D. A procedure to obtain average daily values of meteorological parameters from monthly averages. Sol Energy 1988;40(4):309–13.
[21] Argiriou A, Lykoudis S, Kontoyiannidis S, Balaras C, Asimakopoulos D, Petrakis M, et al. Comparison of methodologies for TMY generation using 20 years data for Athens—Greece. Sol Energy 1999;66(1):35–45.