ELSEVIER

Signal Processing 82 (2002) 1903–1915

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Modeling electricity loads in California: ARMA models with hyperbolic noise

J. Nowicka-Zagrajek[a,*,1], R. Weron[b,1]

[a] *Institute of Mathematics, Wrocław University of Technology, 50-370 Wrocław, Poland*
[b] *Hugo Steinhaus Center for Stochastic Methods, Wrocław University of Technology, 50-370 Wrocław, Poland*

Received 21 May 2001; received in revised form 8 May 2002

Abstract

In this paper we address the issue of modeling and forecasting electricity loads. We apply a two-step procedure to a series of system-wide loads from the California power market. First, we remove the weekly and annual seasonalities. Then, after analyzing properties of the deseasonalized data we fit an autoregressive moving average model. The obtained residuals seem to be independent but with tails heavier than Gaussian. It turns out that the hyperbolic distribution provides an excellent fit. As a justification for our approach we supply out-of-sample forecasts. As it turns out, our method performs significantly better than the one used by the California System Operator.

Keywords: Electricity load; ARMA model; Heavy tails; Hyperbolic distribution; Forecast

1. Introduction

The forecasting of energy demand has become one of the major fields of research in electrical engineering. The power industry requires forecasts with lead times that range from the short term (a few minutes, hours or days ahead) to the long term (up to 20 years ahead). Short-term forecasts, in particular, have become increasingly important since the rise of the competitive energy markets. During the last decade many countries have privatized and deregulated their power markets and electricity has become a commodity that can be sold and bought at market prices.

However, when dealing with the power market we have to bear in mind that electricity cannot simply be manufactured, transported and delivered at the press of a button. Electricity is non-storable (at least not economically), which causes demand and supply to be balanced on a knife edge. Relatively small changes in load or generation can cause large changes in price and all in a matter of hours, if not minutes. In this respect, there is no other market like it.

Load forecasting is vital to the whole power industry, however, it is a difficult task. Firstly, because the load time series exhibit seasonality—at the daily, weekly and annual timescales. Secondly, because there are many exogenous variables that should be considered, with weather conditions being the most

* Corresponding author.
E-mail address: nowicka@im.pwr.wroc.pl (J. Nowicka-Zagrajek).
[1] Research partially supported by KBN Grant PBZ-KBN 016/P03/99.

0165-1684/02/$ - see front matter
PII: S0165-1684(02)00318-3

influential. It is relatively easy to get forecasts with about 10% mean absolute percentage error (MAPE), however, the financial costs of the error are so high that research is aimed at reducing it even by a few percentage points.

Most forecasting models and methods have already been tried out on load forecasting, with varying degrees of success. They may be classified into two broad categories: classical (or statistical) approaches and artificial intelligence-based techniques.

The statistical methods forecast the current value of a variable by using a mathematical combination of the previous values of that variable and previous or current values of exogenous factors, specially weather and social variables. Some models of the first class suggested in recent papers are autoregressive (AR) models [39], dynamic linear [17] or non-linear [48] models, threshold AR models [30] and methods based on Kalman filtering [31,49]. Some of the second class are Box and Jenkins transfer function [24], ARMAX models [57], optimization techniques [58], non-parametric regression [15] and curve-fitting procedures [26]. Despite this large number of alternatives, however, the most popular models are still the linear regression ones [12,20,47,51,52]. These models are attractive because some physical interpretation may be attached to their components, allowing engineers and system operators to understand their behavior.

In recent times, much research has been carried out on the application of artificial intelligence techniques to the load forecasting problem. Expert systems have been used [46] and compared with the traditional methods [40]. Fuzzy inference [41] and fuzzy neural models [42,44] have also been applied. However, the models that have received the largest attention are the artificial neural networks (ANNs) [14,29,36,45,50]. Nevertheless, the reports on the performance of ANNs in forecasting have not entirely convinced the researchers and the practitioners alike and the skepticism may be partly justified [28]. Recent reviews and textbooks on forecasting argue that there is little evidence as yet that ANNs might outperform standard forecasting methods [16,38]. Reviews of ANN-based forecasting systems have concluded that much work still needs to be done before they are accepted as established forecasting techniques [23,59] and that they are promising but that "a significant portion of the ANN research in forecasting and prediction lacks validity" [1]. Two major shortcomings were found to detract from the credibility of the results [28]: the proposed ANN architectures were too large for the data at hand (the ANNs apparently overfitted

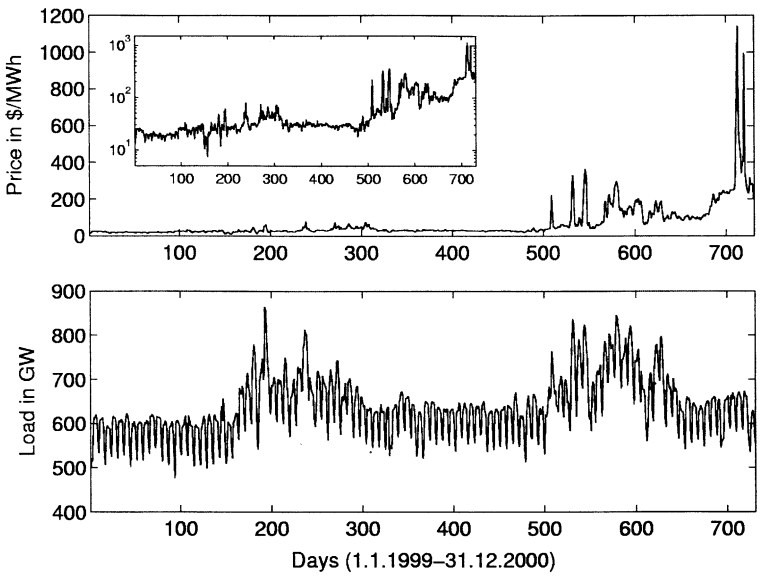

Fig. 1. California Power Exchange daily average market clearing prices (top panel) and California power market daily system-wide load (bottom panel) since January 1, 1999 until December 31, 2000. For clarity, the inset of the top panel displays the prices on a semi-logarithmic scale. The annual and weekly seasonalities of the system-wide load are clearly visible on the bottom panel.

the data) and the models were not systematically tested.

The soaring prices, see the top panel of Fig. 1, and San Francisco blackouts clearly showed that despite the rich literature and the huge number of proposed routines there still is a need for better methods of modeling electricity load dynamics [9,53]. In this paper, we propose a new technique and to justify it we test its out-of-sample forecasting capabilities. As it turns out, our method performs significantly better than the market benchmark—the official forecast of the California System Operator (CAISO).

In contrast to the previous statistical approaches, where the errors are assumed to be Gaussian (either iid $N(0,\sigma^2)$ or stationary autoregressive with Gaussian noise), we go one step further and, after removing the weekly and annual seasonalities, analyze the errors themselves. It turns out that the errors possess short-range correlations, squared errors are uncorrelated for lags greater than one and the distribution of errors exhibits heavy tails. As a consequence we model them using autoregressive moving average (ARMA) time series with hyperbolic noise. Note, that we cannot use ARCH/GARCH type models (which also are heavy-tailed), because we could do it only if the errors were uncorrelated and squared errors would exhibit long-range correlations.

2. Preparation of the data

The analyzed database was provided by the University of California Energy Institute (UCEI, www.ucei.org). Among other data it contains system-wide loads supplied by California's Independent (Transmission) System Operator. This is a time series containing the load for every hour of the period April 1, 1998–December 31, 2000. Due to a very strong daily cycle we have created a 1006 days long sequence of daily loads. Apart from the daily cycle, the time series displays weekly and annual seasonality, see the bottom panel of Fig. 1.

A well-known modeling technique that can be applied to data exhibiting either regular or dynamic periodicity with time increasing amplitudes consists of fitting—via a non-linear least-squares routine—a sum of sine (or cosine) waves having different amplitudes (increasing exponentially [43] or in a power-law fashion [2]), frequencies and/or phase angles. However, in our case the daily data span less than 3 years (in contrast to a 12-year monthly time series studied in [2]) and no significant change of amplitude can be observed.

Because common trend and seasonality removal techniques do not work well when the time series is only a few (and not complete, in our case ca. 2.8 annual cycles) cycles long, we restricted the analysis only to two full years of data, i.e. to the period January 1, 1999–December 31, 2000, and applied a new seasonality reduction technique [56].

The seasonality can be easily observed in the frequency domain by plotting a sample analog of the spectral density, i.e. the periodogram

$$I_n(\omega_k) = \frac{1}{n}\left|\sum_{t=1}^{n} x_t \exp\{-2\pi i(t-1)\omega_k\}\right|^2, \qquad (1)$$

where $\{x_1,\ldots,x_n\}$ is the vector of observations, $\omega_k = k/n$, $k = 1,\ldots,[n/2]$ and $[x]$ denotes the largest integer less then or equal to x. In the top panel of Fig. 2 we plotted the periodogram for the system-wide load. It shows well-defined peaks at frequencies $\omega_k = 0.1428$ and 2.7397×10^{-3} corresponding to cycles with periods of 7 and 365 days, respectively. The smaller peaks at $\omega_k = 0.2857$ and 0.4292 indicate periods of 3.5 and 2.33 days, respectively. Both peaks are the so-called harmonics (multiples of the 7-day period frequency) and indicate that the data exhibits a 7-day period but is not sinusoidal. The weekly period was also observed in lagged autocorrelation plots [55].

To remove the weekly cycle we used the moving average technique (see [11, p. 30]). For the vector of daily loads $\{x_1,\ldots,x_{731}\}$ the trend was first estimated by applying a moving average filter specially chosen to eliminate the weekly component and to dampen the noise

$$\hat{m}_t = \tfrac{1}{7}(x_{t-3} + \cdots + x_{t+3}), \qquad (2)$$

where $t = 4,\ldots,728$. Next, we estimated the seasonal component. For each $k = 1,\ldots,7$, the average w_k of the deviations $\{(x_{k+7j} - \hat{m}_{k+7j}),\ 4 \leqslant k + 7j \leqslant 728\}$ was computed. Since these average deviations do not necessarily sum to zero, we estimated the seasonal

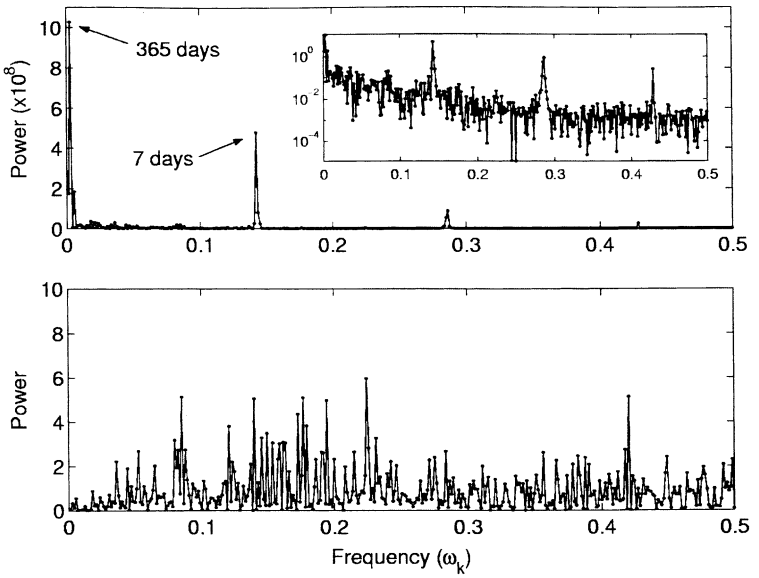

Fig. 2. Periodogram of the California power market daily system-wide load since January 1, 1999 until December 31, 2000 (top panel). The annual ($\omega_k = 2.7397 \times 10^{-3}$) and weekly ($\omega_k = 0.1428$) frequencies are clearly visible. The inset displays the periodogram on a semi-logarithmic scale. Periodogram of the load returns after removal of the weekly and annual cycles (bottom panel). No dominating frequency can be observed.

component s_k as

$$\hat{s}_k = w_k - \frac{1}{7}\sum_{i=1}^{7} w_i, \tag{3}$$

where $k = 1,\ldots,7$ and $\hat{s}_k = \hat{s}_{k-7}$ for $k > 7$. The deseasonalized (with respect to the 7-day cycle) data was then defined as

$$d_t = x_t - \hat{s}_t \quad \text{for } t = 1,\ldots,731. \tag{4}$$

Finally, we removed the trend from the deseasonalized data $\{d_t\}$ by taking logarithmic returns $r_t = \log(d_{t+1}/d_t)$, $t = 1,\ldots,730$.

After removing the weekly seasonality we were left with the annual cycle. Unfortunately, because of the short length of the time series (only 2 years), the method applied to the 7-day cycle could not be used to remove the annual seasonality. To overcome this we applied a new method which consists of the following [56]:

(i) calculate a 25-day rolling volatility [33]

$$v_t = \sqrt{\frac{1}{24}\sum_{i=0}^{24}(R_{t+i} - \bar{R}_t)^2},$$

where $\bar{R}_t = \frac{1}{25}\sum_{i=0}^{24} R_{t+i}$, (5)

for $t = 1,\ldots,730$ and a vector of returns $\{R_t\}$ such that $R_1 = R_2 = \cdots = R_{12} = r_1$, $R_{12+t} = r_t$ for $t = 1,\ldots,730$, and $R_{743} = R_{744} = \cdots = R_{754} = r_{730}$;

(ii) calculate the average volatility for 1 year, i.e. in our case

$$\bar{v}_t = \frac{v_t^{1999} + v_t^{2000}}{2}; \tag{6}$$

(iii) smooth the volatility by taking a 25-day moving average of $\bar{v}_t$;

(iv) finally, rescale the returns by dividing them by the smoothed annual volatility.

The obtained time series (see the top panel of Fig. 3) showed no apparent trend and seasonality (see the bottom panel of Fig. 2). Therefore, we treated it as a realization of a stationary process. Moreover, the dependence structure exhibited only short-range correlations. Both, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) rapidly tended to zero (see the bottom panels of Fig. 3), which

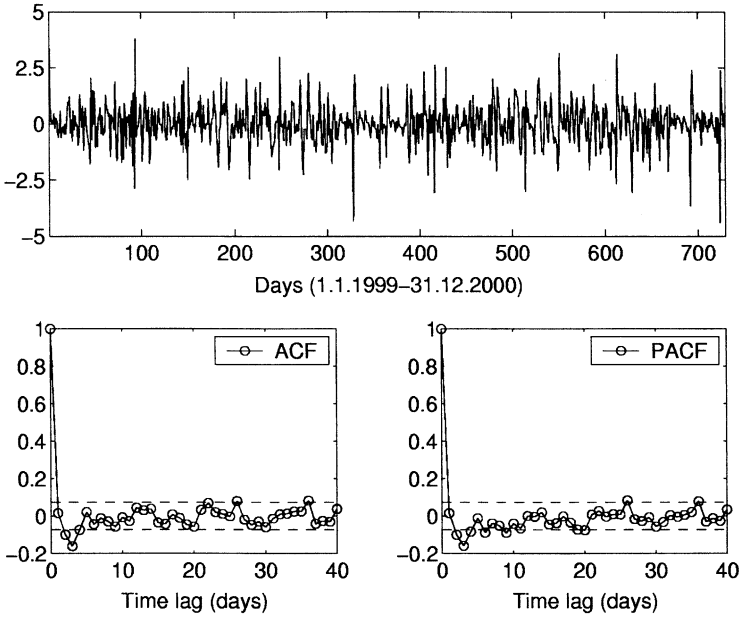

Fig. 3. Load returns after removal of the weekly and annual cycles (top panel). The ACF (bottom left panel) and PACF (bottom right panel) for the mean-corrected deseasonalized load returns. Dashed lines represent the bounds $\pm 1.96/\sqrt{730}$, i.e. the 95% confidence intervals of Gaussian white noise.

suggested that the deseasonalized load returns could be modeled by an ARMA-type process. As a further justification we tested the dependence of squared values of the time series. It turned out that they were uncorrelated for lags greater than one. These properties ruled out ARCH/GARCH-type models, which also are heavy-tailed, but exhibit long-range correlations for squared values and no correlations for the time series itself.

3. Modeling with ARMA processes

The mean-corrected (i.e. after removing the sample mean = 0.0010658) deseasonalized load returns were modeled by ARMA processes

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}, \quad t = 1, \ldots, n, \tag{7}$$

where (p, q) denote the order of the model and $\{Z_t\}$ is a sequence of independent, identically distributed variables with mean 0 and variance σ^2 (denoted by $\mathrm{iid}(0, \sigma^2)$ in the text).

The maximum likelihood estimators $\hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)$, $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_q)$ and $\hat{\sigma}^2$ of the parameters $\phi = (\phi_1, \ldots, \phi_p)$, $\theta = (\theta_1, \ldots, \theta_q)$ and σ^2, respectively, were obtained after a preliminary estimation via the Hannan–Rissanen method (see [11, p. 154]) using all 730 deseasonalized returns. The ML estimators used here are based on the Gaussian assumption. However, this does not exclude models with non-Gaussian noise since the large sample distribution of the estimators is the same for $\{Z_t\} \sim \mathrm{iid}(0, \sigma^2)$, regardless of whether or not $\{Z_t\}$ is Gaussian (see [10, Section 10.8]).

The parameter estimates and the model size (p, q) were selected to be those that minimize the bias-corrected version of the Akaike criterion, i.e. the AICC statistics (see [11, Section 5.5])

$$\mathrm{AICC} = -2 \ln L + \frac{2(p + q + 1)n}{n - p - q - 2}, \tag{8}$$

where L denotes the maximum likelihood function and $n = 730$. The optimization procedure led us to the

Table 1
AICC values for the best AR, MA and ARMA models

Model class	Best model	AICC value	Model class	Best model	AICC value
AR(·)	AR(11)	1965.226	MA(·)	MA(4)	1959.299
ARMA(1,·)	**ARMA(1,6)**	**1956.294**	ARMA(2,·)	ARMA(2,3)	1960.312
ARMA(3,·)	ARMA(3,3)	1962.037	ARMA(4,·)	ARMA(4,4)	1963.028
ARMA(5,·)	ARMA(5,3)	1962.045	ARMA(6,·)	ARMA(6,2)	1966.060
ARMA(7,·)	ARMA(7,3)	1967.511	ARMA(8,·)	ARMA(8,1)	1967.844
ARMA(9,·)	ARMA(9,1)	1965.556	ARMA(10,·)	ARMA(10,1)	1967.628
ARMA(11,·)	ARMA(11,1)	1968.005			

"·" denotes an integer from the interval [1,11].

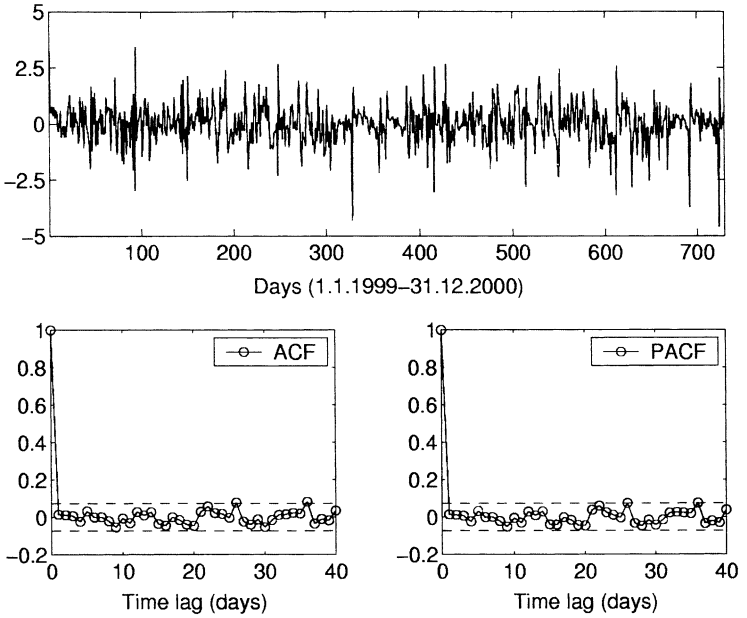

Fig. 4. The residuals obtained from the ARMA(1,6) model (top panel). The ACF (bottom left panel) and the PACF (bottom right panel) of the residuals. Dashed lines represent the bounds $\pm 1.96/\sqrt{730}$.

following ARMA(1,6) model (with $\theta_4 = \theta_5 = 0$):

$$X_t = 0.332776X_{t-1} + Z_t - 0.383245Z_{t-1} - 0.12908Z_{t-2} - 0.149307Z_{t-3} - 0.0531862Z_{t-6}, \quad (9)$$

where $t = 1,\ldots,730$ and $\{Z_t\} \sim \text{iid}(0, 0.838716)$. The value of the AICC criterion obtained for this model was AICC = 1956.294. For comparison, in Table 1 we present the AICC values for the best models in each class.

In order to check the goodness of fit of the model to the set of data we compared the observed values with the corresponding predicted values obtained from the fitted model. If the fitted model was appropriate, then the residuals

$$\hat{W}_t = \frac{X_t - \hat{X}_t(\hat{\phi},\hat{\theta})}{\sqrt{\varsigma_{t-1}(\hat{\phi},\hat{\theta})}}, \quad t = 1,\ldots,730, \quad (10)$$

where $\hat{X}_t(\hat{\phi},\hat{\theta})$ denotes the predicted value of X_t based on $X_1,\ldots,X_{t-1}$ and $\varsigma_{t-1} = E(X_t - \hat{X}_t)^2/\sigma^2$ should behave in a manner that is consistent with the model. In

Table 2
Test statistics and p-values for the residuals

Test	Test statistics value	p-value
Portmanteau	15.03	(0.7747)
Turning point	464	(0.0609)
Difference-sign	361	(0.6536)
Rank	131090	(0.5529)

our case this means that the properties of the residuals should reflect those of an iid noise sequence with mean 0 and variance σ^2.

The residuals obtained from the ARMA$(1,6)$ model fitted to the mean-corrected deseasonalized load returns are displayed in the top panel of Fig. 4. The graph gives no indication of a non-zero mean or non-constant variance. The sample ACF and PACF of the residuals fall between the bounds $\pm 1.96/\sqrt{730}$ indicating that there is no correlation in the series, see the bottom panels of Fig. 4. Recall that for large sample size n the sample autocorrelations of an iid sequence with finite variance are approximately iid with distribution $N(0, 1/n)$. Therefore, there is no reason to reject the fitted model on the basis of the ACF or PACF. However, we should not rely only on simple visual inspection techniques. For our results to be more statistically sound we performed several standard tests for randomness. The results of the portmanteau, turning point, difference-sign and rank tests are presented in Table 2. Short descriptions of all applied tests can be found in the appendix.

As we can see from Table 2, if we carry out the tests at commonly used 5% level, the tests do not detect any deviation from the iid behavior. Thus there is not sufficient evidence to reject the iid hypothesis. Moreover, the order $p = 0$ of the minimum AICC autoregressive model for the residuals also suggests the compatibility of the residuals with white noise, see the last paragraph of the appendix. Therefore, we may conclude that the ARMA$(1,6)$ model (defined by Eq. (9)) fits the mean-corrected deseasonalized load returns very well.

4. Distribution of the residuals

In the previous section we showed that the residuals are a realization of an iid$(0, \sigma^2)$ sequence. But what precisely is their distribution? The answer to this question is important, because if the noise distribution is known then stronger conclusions can be drawn when a model is fitted to the data. There are simple visual inspection techniques that enable us to check whether it is reasonable to assume that observations from an iid sequence are also Gaussian. The most widely used is the so-called normal probability plot, see Fig. 5. If the residuals were Gaussian then they would form a straight line (dotted line in the plot). Obviously, they are not Gaussian—the deviation from the line is apparent. This deviation suggests that the residuals have heavier tails. However, the tails are not Paretian as well. Observe that in the tails the distribution of the residuals does not comply with the solid line, which presents an α-stable law with parameters $\alpha = 1.7063$, $\sigma = 0.5116$, $\beta = -0.0809$, $\mu = 0.0083$ estimated from the data via the regression method [34]. This may indicate a finite variance of the empirical distribution, which is a desirable feature since in order to comply with the ARMA model assumptions the distribution of the residuals must have a finite second moment.

Moreover, if we plot the empirical probability density function (PDF)—to be more precise: a kernel estimator of the density—we can clearly see that, on the semi-logarithmic scale, the tails of the residuals' density form straight lines, see Fig. 6. In the class of heavy-tailed laws with finite variance and hyperbolic decay in the tails the hyperbolic distribution seems to be a natural candidate.

The hyperbolic law was introduced by Barndorff-Nielsen [3] in 1977 for modeling the grain size distribution of windblown sand [5,6]. Almost 20 years later it was found to provide an excellent fit to the distributions of daily returns of stocks from a number of leading German enterprises [18,35], giving way to its today's widespread use in stock price modeling [7] and market risk measurement [19]. The hyperbolic distribution or hyperbolic decay has also been fit to taxonomic systems [13], the size distribution of plant seeds [27], bibliometric and scientometric data [32], linguistics data [25] and impulse noises [54]. The name of the distribution is derived from the fact that its log-density forms a hyperbola. Recall that the log-density of the normal distribution is a parabola. Hence, the hyperbolic distribution provides the possibility of modeling heavy tails.

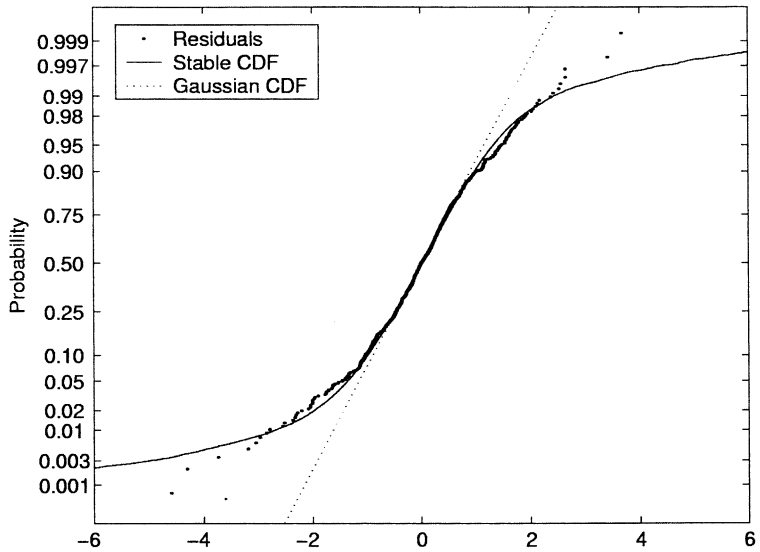

Fig. 5. The normal probability plot of the residuals obtained from the ARMA(1, 6) model. If the residuals were Gaussian then they would form a straight (dotted in the plot) line. If they were α-stable then they would comply with the curved solid line, which presents a stable law with parameters estimated from the data.

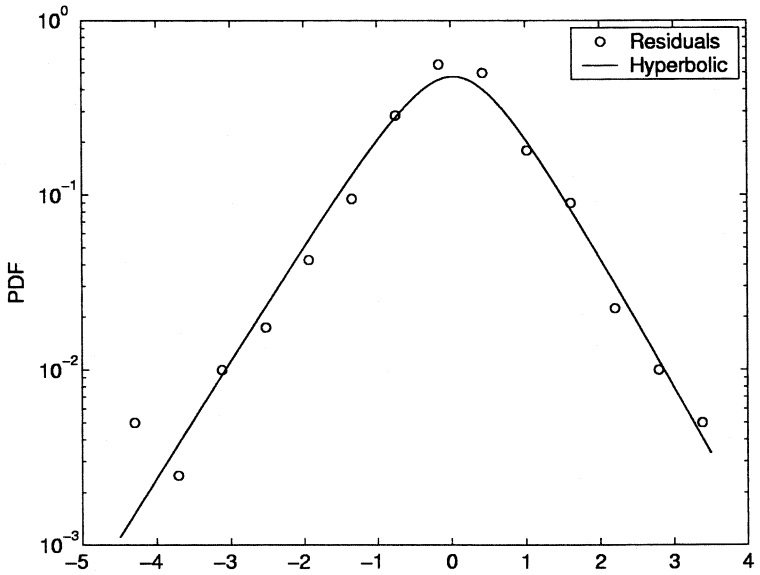

Fig. 6. The empirical probability density function (a kernel estimator of the density) and the approximating hyperbolic PDF on the semi-logarithmic scale.

The hyperbolic distribution is defined as a normal variance–mean mixture where the mixing distribution is the inverse Gaussian law [3,4], i.e. it is conditionally Gaussian (for other conditionally Gaussian models in the signal processing literature see e.g. [21,22]). More precisely, a random variable Z has the

hyperbolic distribution if $(Z|Y) \sim N(\mu + \beta Y, Y)$, where $Y \sim \mathrm{IG}(\chi, \psi)$, with the probability density function

$$f_Y(x) = \frac{(\psi/\chi)^{1/2}}{2K_1(\sqrt{\chi\psi})} \exp\left\{-\frac{1}{2}[\chi x^{-1} + \psi x]\right\}, \quad x > 0. \tag{11}$$

The normalizing constant $K_1(t) = 1/2 \int_0^\infty \exp\{-(1/2)t(x + \frac{1}{x})\}\,\mathrm{d}x$, $t > 0$, is the modified Bessel function with index 1. This means that $Z \sim \mathrm{Hyp}(\chi, \psi, \beta, \mu)$ can be represented in the form $Z = \mu + \beta Y + \sqrt{Y}N(0,1)$ with the characteristic function

$$\phi_Z(u) = \exp(\mathrm{i}u\mu) \int_0^\infty \exp\left\{\mathrm{i}\beta z u - \frac{1}{2} z u^2\right\} \mathrm{d}F_Y(z). \tag{12}$$

In what follows we adopt a more natural parameterization of the hyperbolic distribution with $\delta = \sqrt{\chi}$ and $\alpha = \sqrt{\psi + \beta^2}$. Then the probability density function of the hyperbolic $\mathrm{Hyp}(\alpha, \beta, \delta, \mu)$ law can be written as

$$f(x; \alpha, \beta, \delta, \mu) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} \exp\{-\alpha\sqrt{\delta^2 + (x - \mu)^2} + \beta(x - \mu)\}, \tag{13}$$

where $\delta > 0$ is the scale parameter, $\mu \in R$ is the location parameter and $0 \leqslant |\beta| < \alpha$. The latter two parameters—α and β—determine the shape, with α being responsible for the steepness and β for the skewness.

Given a sample of independent observations all four parameters can be estimated by the maximum likelihood method. In our studies we used the 'hyp' program [8] to obtain the following estimates:

$$\hat{\alpha} = 1.671304, \quad \hat{\beta} = -0.098790,$$

$$\hat{\delta} = 0.298285, \quad \hat{\mu} = 0.076975.$$

The empirical PDF together with the estimated hyperbolic PDF are presented in Fig. 6. The adjusted Kolmogorov statistics $K = \sqrt{n}\sup_x |F(x) - F_n(x)|$, where $F(x)$ is the theoretical and $F_n(x)$ is the empirical cummulative distribution function, returns the value $K = 1.5652$. This indicates that there is not sufficient evidence to reject the hypothesis of the hyberbolic distribution of the residuals at the 1% level. For comparison we fitted a Gaussian law to the residuals as well. In this case the adjusted Kolmogorov statistics returned $K = 1.8019$ causing us to reject the Gaussian hypothesis of the residuals at the same level.

5. Forecasting

It is not so surprising that a model will perform well when evaluated by its fit to the data set to which it was adjusted. The real test is whether it will be capable of also describing new data sets coming from the same process. A suggestive and attractive way of comparing different models is to evaluate their performance when applied to a data set to which none of them was adjusted. The standard measure of goodness of fit is the difference between actual and forecasted outputs. The disadvantage of this method is that we have to save part of the data set for the comparisons and therefore cannot use all available information to build the model.

In the previous sections we fit an ARMA(1,6) model to the deseasonalized system-wide load returns from the period January 1, 1999 to December 31, 2000. Now, we test the performance of the model on data from the two following months, i.e. from the period January 1 to February 28, 2001. For every day in the test period we run a day-ahead prediction by applying the model defined in Eq. (9) and using previous observations. The results are then "inverted" (the seasonality is added by following the steps of Section 2 in reverse order) and compared with the actual system-wide loads and the CAISO official day-ahead forecasts (see www.caiso.com).

The performance of the model is summarized in Table 3 (column 3). Looking at the mean square error (MSE) values for the whole test period we can observe that the CAISO forecast outperforms our model. However, this is only illusionary. Large value of the MSE for our model is caused by two extreme observations corresponding to January 1 and 2 (lags 1 and 2), see Fig. 7 and especially Fig. 8. The mean absolute error (MAE) places less weight to the extreme differences and here our model outperforms the CAISO forecast. The same result can be observed for the mean absolute percentage error ($\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} |(\hat{x}_i - x_i)/x_i| \times 100$).

Table 3
Comparison of forecasting capabilities

Forecasting approach			
Error	CAISO	ARMA(1,6)	Adaptive ARMA
January 1–*February* 28			
MSE	208.34	304.14	318.50
MAE	10.52	9.86	9.87
MAPE (%)	1.7799	1.6642	1.6682
January 3–*February* 28			
MSE	190.28	89.00	88.36
MAE	10.08	7.39	7.31
MAPE (%)	1.7087	1.2401	1.2282

Note the difference in performance for the period without New Year's Day.

Observe that the extreme differences in the ARMA(1,6) model correspond to the US national holidays—New Year's (lags 1 and 2) and Presidents' Day (lags 50 and 51). Obviously, our model cannot capture the holiday structure. However, this can be quite easily incorporated into it by simply subtracting a certain amount of GW for these holidays. When we compare the forecasting results for the period January 3–February 28, 2001 (i.e. without New Year's) our model is significantly better. The MSE is reduced by 53% and the MAE by 27%. The results would improve even more if we eliminated the Presidents' Day from the test period.

The forecasts can be still fine tuned by applying an adaptive scheme. In the fourth column of Table 3 we present the results of such an approach (we did not include these results in Figs. 7 and 8 because the differences are quite small and would be almost invisible). Instead of using a single model (defined in Eq. (9)), for every day in the test period we fitted the best ARMA model (of order $p,q \leqslant 6$) to the last 730 deseasonalized load returns and run a day-ahead prediction by applying the obtained models.

6. Conclusions

Short-term load forecasting plays an important role in power system operation and planning. Accurate load prediction saves costs by improving economic load dispatching, unit commitment, etc. At the same time it enhances the function of security control.

In this paper, we have proposed an efficient method for short-term load forecasting with heavy-tailed

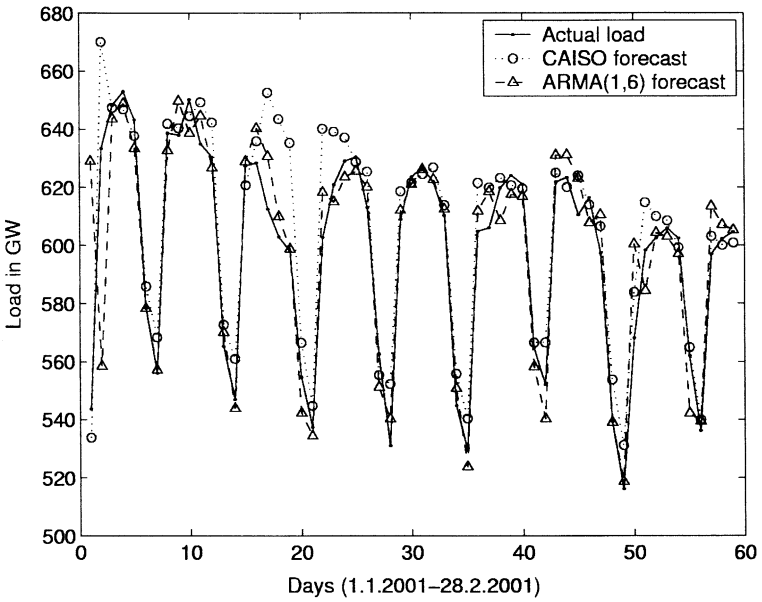

Fig. 7. California power market daily system-wide load, CAISO day-ahead forecast and ARMA(1,6)—see Eq. (9)—day-ahead forecast for the period January 1–February 28, 2001.

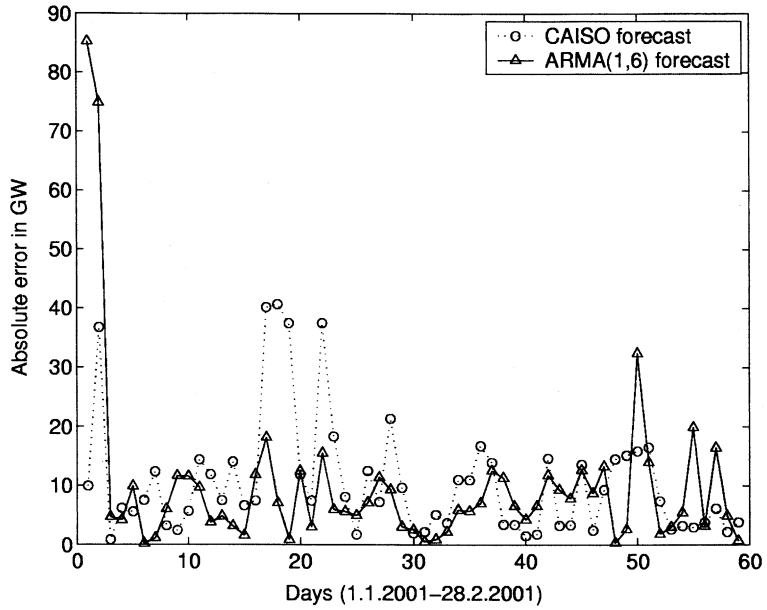

Fig. 8. Absolute errors for the CAISO and ARMA(1, 6) day-ahead forecasts of the system-wide load for the period January 1–February 28, 2001. Observe that the extreme differences in the ARMA(1, 6) model correspond to the US national holidays—New Year's (lags 1 and 2) and Presidents' Day (lags 50 and 51).

ARMA time series and a data-preprocessing technique. After applying a new seasonality removal technique we modeled the data with an ARMA process. However, in contrast to the previous statistical approaches, we did not assume a Gaussian distribution of the residuals. The performed analysis suggested the use of the hyperbolic distribution. As a consequence we modeled deseasonalized loads using ARMA time series with hyperbolic noise.

The proposed method was successfully applied to real data. A comparison was made between the proposed technique and the official forecasts of the California System Operator (CAISO). The effectiveness of the proposed method was demonstrated through a comparison of the real load data with short-term forecasted values. In terms of the mean absolute percentage error our approach yielded only a 1.2–1.25% difference, whereas the CAISO day-ahead forecasts returned a 1.7% error.

We strongly believe that our approach is a universal one and can be applied not only to the California power market system-wide load but also to other data sets displaying seasonalities and heavy tails. Moreover, the computational times are negligible and the method can be used in real time.

Acknowledgements

We gratefully acknowledge critical comments of two anonymous referees, which led to the improvement of the paper.

Appendix A. Tests for randomness (see [11, p. 164])

The portmanteau test. Instead of checking to see if each sample autocorrelation $\hat{\rho}(j)$ falls inside the bounds $\pm 1.96/\sqrt{n}$, where n is the sample size, it is possible to consider a single statistic introduced by Ljung and Box [37] $Q = n(n+2)\sum_{j=1}^{h} \hat{\rho}^2(j)/(n-j)$, whose distribution can be approximated by the χ^2 distribution with h degrees of freedom. A large value of Q suggests that the sample autocorrelations of the observations are too large for the data to be a sample

from an iid sequence. Therefore, we reject the iid hypothesis at level α if $Q > \chi^2_{1-\alpha}(h)$, where $\chi^2_{1-\alpha}$ is the $(1-\alpha)$ quantile of the χ^2 distribution with h degrees of freedom.

The turning point test. If $y_1, \ldots, y_n$ is a sequence of observations, we say that there is a turning point at time i $(1 < i < n)$ if $y_{i-1} < y_i$ and $y_i > y_{i+1}$ or if $y_{i-1} > y_i$ and $y_i < y_{i+1}$. In order to carry out a test of the iid hypothesis (for large n) we denote the number of turning points by T (T is approximately $N(\mu_T, \sigma_T^2)$, where $\mu_T = 2(n-2)/3$ and $\sigma_T^2 = (16n - 29)/90$) and we reject this hypothesis at level α if $|T - \mu_T|/\sigma_T > \Phi_{1-\alpha/2}$, where $\Phi_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. The large value of $T - \mu_T$ indicates that the series is fluctuating more rapidly than expected for an iid sequence; a value of $T - \mu_T$ much smaller than zero indicates a positive correlation between neighboring observations.

The difference-sign test. For this test we count the number S of values i such that $y_i > y_{i-1}$, $i = 2, \ldots, n$. For an iid sequence and for large n, S is approximately $N(\mu_S, \sigma_S^2)$, where $\mu_S = (n-1)/2$ and $\sigma_S^2 = (n+1)/12$. A large positive (or negative) value of $S - \mu_S$ indicates the presence of an increasing (or decreasing) trend in the data. We therefore reject the assumption of no trend in the data if $|S - \mu_S|/\sigma_S > \Phi_{1-\alpha/2}$.

The rank test. The rank test is particularly useful for detecting a linear trend in the data. We define P as the number of pairs (i, j) such that $y_j > y_i$ and $j > i$, $i = 1, \ldots, n-1$. For an iid sequence and for large n, P is approximately $N(\mu_P, \sigma_P^2)$, where $\mu_P = n(n-1)/4$ and $\sigma_P^2 = n(n-1)(2n+5)/72$. A large positive (negative) value of $P - \mu_P$ indicates the presence of an increasing (decreasing) trend in data. The iid hypothesis is therefore rejected at level α if $|P - \mu_P|/\sigma_P > \Phi_{1-\alpha/2}$.

The minimum AICC AR model test. A simple test for whiteness of a time series is to fit autoregressive models of orders $p = 0, 1, \ldots, p_{\max}$, for some large $p_{\max}$, and to record the value of p for which the AICC value attains the minimum. Compatibility of these observations with white noise is indicated by selection of the value $p = 0$.

References

[1] M. Adya, F. Collopy, How effective are neural networks at forecasting and prediction? A review and evaluation, J. Forecast. 17 (1998) 481–495.

[2] E.H. Barakat, Modeling of non-stationary time-series data, Part II, Dynamic periodic trends, Electr. Power Energy Systems 23 (2001) 63–68.

[3] O.E. Barndorff-Nielsen, Exponentially decreasing distributions for the logarithm of particle size, Proc. Roy. Soc. London A 353 (1977) 401–419.

[4] O.E. Barndorff-Nielsen, Hyperbolic distributions and distributions on hyperbolae, Scand. J. Statist. 5 (1978) 151–157.

[5] O.E. Barndorff-Nielsen, P. Blaesild, J.L. Jensen, M. Sørensen, The fascination of sand, in: A.C. Atkinson, S.E. Feinberg (Eds.), A Celebration of Statistics, Springer, New York, 1985, pp. 57–87.

[6] O.E. Barndorff-Nielsen, J.L. Jensen, M. Sørensen, Parametric modeling of turbulence, Philos. Trans. Roy. Soc. London A 332 (1990) 439–455.

[7] B.M. Bibby, M. Sørensen, A hyperbolic diffusion model for stock prices, Finance Stochastics 1 (1997) 25–41.

[8] P. Blaesild, M. Sørensen, 'Hyp'—a computer program for analyzing data by means of the hyperbolic distribution, Research Report No. 248, Dept. Theor. Statist., Univ. Aarhus, 1992.

[9] S. Borenstein, The trouble with electricity markets (and some solutions), POWER Working Paper PWP-081, University of California Energy Institute, www.ucei.org, 2001.

[10] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, 2nd Edition, Springer, New York, 1991.

[11] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, Springer, New York, 1996.

[12] D.W. Bunn, E.D. Farmer, Comparative Models for Electrical Load Forecasting, Wiley, New York, 1985.

[13] B. Burlando, The fractal dimension of taxonomic systems, J. Theor. Biol. 146 (1990) 99–114.

[14] W. Charytoniuk, M.S. Chen, Very short-term load forecasting using artificial neural networks, IEEE Trans. Power Systems 15 (1) (2000) 263–268.

[15] W. Charytoniuk, M.S. Chen, P. Van Olinda, Non-parametric regression based short-term load forecasting, IEEE Trans. Power Systems 13 (3) (1998) 725–730.

[16] C. Chatfield, Forecasting in the 1990s, in: Proceedings of the V Annual Conference of Portuguese Society of Statistics, Curia, 1997, pp. 57–63.

[17] A.P. Douglas, A.M. Breipohl, F.N. Lee, R. Adapa, The impact of temperature forecast uncertainty on Bayesian load forecasting, IEEE Trans. Power Systems 13 (4) (1998) 1507–1513.

[18] E. Eberlein, U. Keller, Hyperbolic distributions in finance, Bernoulli 1 (1995) 281–299.

[19] E. Eberlein, U. Keller, K. Prause, New insights into the smile, mispricing and value at risk: the hyperbolic model J. Bus. 71 (1998) 371–406.

[20] R. Engle, C. Mustafa, J. Rice, Modeling peak electricity demand, J. Forecast. 11 (1992) 241–251.

[21] S.J. Godsill, MCMC and EM-based methods for inference in heavy-tailed processes with alpha-stable innovations, in: Proceedings of the IEEE Signal Processing Workshop on Higher-order Statistics, June 1999, Caesarea, Israel.

[22] S.J. Godsill, E.E. Kuruoglu, Bayesian inference for time series with heavy-tailed symmetric alpha-stable noise processes, in: Proceedings of the Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics, June 1999, Washington DC, USA.
[23] W.L. Gorr, Research prospective on neural network forecasting, Int. J. Forecast. 10 (1994) 1–4.
[24] M.T. Hagan, S.M. Behr, The time series approach to short-term load forecasting, IEEE Trans. Power Systems 2 (3) (1987) 785–791.
[25] P. Harremoës, F. Topsoe, Zipf's law, hyperbolic distributions and entropy loss, Proceedings of the 2002 IEEE International Symposium on Information Theory, Lausanne, 2002.
[26] A. Harvey, S. Koopman, Forecasting hourly electricity demand using time-varying splines, J. Amer. Statist. Assoc. 88 (1993) 1228–1253.
[27] S.G. Hegde, R. Lokesha, K.N. Ganeshaiah, Seed size distribution in plants: an explanation based on fractal geometry Oikos 62 (1991) 100–101.
[28] H.S. Hippert, C.E. Pedreira, R.C. Souza, Neural networks for short-term load forecasting: a review and evaluation IEEE Trans. Power Systems 16 (1) (2001) 44–55.
[29] Y. Hsu, C. Yang, Electrical load forecasting, in: A. Murray (Ed.), Applications of Neural Networks, Kluwer, Boston, 1995, pp. 157–189.
[30] S.R. Huang, Short-term load forecasting using threshold autoregressive models, IEE Proc.—Gener. Transm. Distrib. 144 (5) (1997) 477–481.
[31] D.G. Infield, D.C. Hill, Optimal smoothing for trend removal short term electricity demand forecasting, IEEE Trans. Power Systems 13 (3) (1998) 1115–1120.
[32] L.E. Ivancheva, The non-Gaussian nature of bibliometric and scientometric distributions: a new approach to interpretation J. Amer. Soc. Inform. Sci. Technol. 52 (2001) 1100–1105.
[33] V. Kaminski, The challenge of pricing and risk managing electricity derivatives, in: P. Barber, ed., The US Power Market, Risk Books, London, 1997.
[34] I.A. Koutrouvelis, Regression-type estimation of the parameters of stable laws, J. Amer. Statist. Assoc. 75 (1980) 918–928.
[35] U. Küchler, K. Neumann, M. Sørensen, A. Streller, Stock returns and hyperbolic distributions, Math. Comput. Modelling 29 (1999) 1–15.
[36] K. Lee, Y. Cha, J. Park, Short-term load forecasting using an artificial neural network, IEEE Trans. Power Systems 7 (1992) 124–132.
[37] G.M. Ljung, G.E.P. Box, On a measure of lack of fit in time series models, Biometrica 65 (1978) 297–303.
[38] S. Makridakis, S.C. Wheelwright, R.J. Hyndman, Forecasting—Methods and Applications, 3rd Edition, Wiley, New York, 1998.
[39] G.A.N. Mbamalu, M.E. El-Hawary, Load forecasting suboptimal seasonal autoregressive models and iteratively reweighted least-squares estimation, IEEE Trans. Power Systems 8 (1) (1993) 343–348.
[40] I. Moghram, S. Rahman, Analysis and evaluation of five short-term load forecasting techniques, IEEE Trans. Power Systems 4 (1989) 1484–1491.
[41] H. Mori, K. Kobayashi, Optimal fuzzy inference for short-term load forecasting, IEEE Trans. Power Systems 11 (1) (1996) 390–396.
[42] K. Padmakumari, K.P. Mohandas, S. Thiruvengadam, Long term distribution demand forecasting using neuro fuzzy computations, Electr. Power Energy Systems 21 (1999) 315–322.
[43] S.M. Pandit, S.M. Wu, Time Series and System Analysis with Applications, Wiley, New York, 1983.
[44] S.E. Papadakis, J.B. Theocharis, S.J. Kiartzis, A.G. Bakirtzis, A novel approach to short-term load forecasting using fuzzy neural networks, IEEE Trans. Power Systems 13 (2) (1998) 480–492.
[45] T. Peng, N. Hubele, G. Karady, Advancement in the application of neural networks for short-term load forecasting, IEEE Trans. Power Systems 7 (1992) 250–257.
[46] S. Rahman, O. Hazim, A generalized knowledge-based short-term load forecasting technique, IEEE Trans. Power Systems 8 (2) (1993) 508–514.
[47] R. Ramanathan, R. Engle, C.W.J. Granger, F. Vahid-Araghi, C. Brace, Short-run forecasts of electricity loads and peaks, Int. J. Forecast. 13 (1997) 161–174.
[48] R. Sadownik, E.P. Barbosa, Short-term forecasting of industrial electricity consumption in Brasil, J. Forecast. 18 (1999) 215–224.
[49] S. Sargunaraj, D.P. Sen Gupta, S. Devi, Short-term load forecasting for demand side management, IEE Proc.—Gener. Transm. Distrib. 144 (1) (1997) 68–74.
[50] T. Senjyu, H. Takara, K. Uezato, T. Funabashi, One-hour-ahead load forecasting using neural network, IEEE Trans. Power Systems 17 (1) (2002) 113–118.
[51] M. Smith, Modeling and short-term forecasting of New South Wales electricity system load, J. Bus. Econom. Statist. 18 (2000) 465–478.
[52] S.A. Soliman, S. Persaud, K. El-Nagar, M.E. El-Hawary, Application of list absolute value parameter estimation based on linear programming to short-term load forecasting, Electr. Power Energy Systems 19 (3) (1997) 209–216.
[53] S. Stoft, The market flaw California overlooked, New York Times, January 2, 2001.
[54] K. Szechenyi, On the NEXT and impulse noise properties of subscriber loops, IEEE Globecom '89 Conference Record, Tokyo, 1989, pp. 1569–1573.
[55] R. Weron, Energy price risk management, Physica A 285 (2000) 127–134.
[56] R. Weron, B. Kozłowska, J. Nowicka-Zagrajek, Modeling electricity loads in California: a continuous-time approach Physica A 299 (2001) 344–350.
[57] H.T. Yang, C.M. Huang, A new short-term load forecasting approach using self-organizing fuzzy ARMAX models, IEEE Trans. Power Systems 13 (1) (1998) 217–225.
[58] Z. Yu, A temperature match based optimization method for daily load prediction considering DLC effect, IEEE Trans. Power Systems 11 (2) (1996) 728–733.
[59] G. Zhang, B.E. Patuwo, M.Y. Hu, Forecasting with artificial neural networks: the state of the art Int. J. Forecast. 14 (1998) 35–62.