Electricity price forecasting with confidence-interval estimation through an extended ARIMA approach

M. Zhou, Z. Yan, Y.X. Ni, G. Li and Y. Nie

Abstract: Accurate electricity price forecasting is a crucial issue concerned by market participants either for developing bidding strategies or for making investment decisions. Due to the complicated factors affecting electricity prices, accurate price forecasting turns out to be very difficult. The autoregressive integrated moving average (ARIMA) approach has been extended to make hourly market clearing price (MCP) forecasting in electricity spot markets with error correction and confidence interval estimation. The ARIMA model used for forecasting price and the method to implement price forecasting are presented first. Then the ARIMA approach is extended to include error correction for improving accuracy of price forecasting. Moreover, the confidence interval of the forecasted prices is estimated assuming the residual errors are in gaussian or uniform distribution. Hourly MCP forecasting of the Californian Power Market is used as a computer example, and the comparison with conventional ARIMA approach is given. Computer test results show clearly that the suggested extended ARIMA approach for spot price forecasting is very effective with satisfactory accuracy. It can work under very worse market conditions with high price volatility.

1 Introduction

In power markets, accurate electricity price forecasting [1–15] is a crucial issue for all market participants. According to forecast electricity prices, producers can develop bidding strategies to maximise profits and minimise risks, while consumers can allocate purchases between long-term bilateral contracts and spot markets. Electricity prices are also an essential reference for investors for optimal portfolio management.

However, due to the complicated factors affecting electricity prices, accurate price forecasting turns out to be very difficult. As we all know, power plant construction requires large investment and a relatively long period, which leads to power-market entry barriers. Therefore the electricity markets are apt to be oligopolistic, in which power suppliers make strategic bidding and/or withhold capacities to raise the price so as to maximise their own profits. In addition, various factors such as transmission congestion, maintenance of generation units, variations of fuel costs or water supply and load fluctuations etc. will also affect the electricity price noticeably and increase the complexity of accurate price forecasting.

Due to the complexity of factors affecting electricity prices conventional techniques for load forecasting [16, 17] cannot be used directly for satisfactory price forecasting, especially for cases where drastic fluctuations of prices exist. Hence practical approaches for electricity price forecasting are investigated, which can be roughly classified into two categories. One is based on market simulators, in which all operation costs of generation are considered with generation limits and transmission constraints taken into account. This type of approach generally makes a large number of hypotheses and is more suitable for forecasting long-term prices. Another type of approach is based on statistical analysis. It is assumed that historical prices reflect the characteristics of the electricity prices to be forecast. If market rules and operating conditions are kept relatively stable this type of approach is suitable for short-term price forecasting.

Much work has been done on electricity market price forecasting. References [1, 2] analyse the main factors influencing electricity prices and compare situations of typical power markets around the world. It is found that large differences exist in different markets. Reference [3] provides a comprehensive review on various issues and techniques for forecasting daily loads and electricity prices in the new competitive power markets. In [4–7, 10], neural network approaches are proposed to forecast short-term electricity prices. Time-series model-based approaches are studied to predict market clearing prices (MCP) in [9, 12–14]. Attention is paid to confidence-interval estimation [10] and the probability distribution of forecast prices [11]. Fuzzy logic chaotic theory are also used for predicting electricity prices in [11]. Electricity prices are highly random and the predicted prices will always have errors. It is highly desirable to know not only the forecast prices but also the confidence interval of the forecast price since the confidence interval provides a range covering the true value at a specified probability [18, 19]. Therefore estimation of the confidence interval is another task of significance in price forecasting.

In this paper stochastic process theory [20, 21] has been applied to electricity price forecasting. Electricity prices are considered as a nonstationary stochastic time-series and the

IEE Proceedings online no. 20045131

doi:10.1049/ip-gtd:20045131

Paper first received 7th August 2004 and in final revised form 10th September 2005

M. Zhou, G. Li and Y. Nie are with Key Laboratory of Power System Protection and Dynamic Security Monitoring and Control under Ministry of Education, North China Electric Power University, Baoding, Hebei, Province, 071003, China

Z. Yan and Y.X. Ni are with Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, SAR, P.R. China

E-mail: zhouming@ncepu.edu.cn

autoregressive integrated moving average (ARIMA) model and corresponding solution method will be extended to make price forecasting of electricity spot markets. The ARIMA approach has been successfully applied to quite a few areas [22, 25]. It is superior in modelling nonstationary time-series and forecasting. Based on real market records and the corresponding forecast electricity prices, the model error time-series can be calculated with its features filtered out for utilisation in adjusting the forecast prices and further improving accuracy. Because of the highly volatile electricity prices, this paper also presents an approach for estimating confidence interval based on remaining error time-series. The historical data of Californian Power Market [26] are used for computer test. It is shown that the suggested ARIMA model and method adopted for price forecasting are of lower order, easy to implement and with satisfactory accuracy. The accuracy of forecast price with error correction is very satisfactory even for cases where large price volatilities exist. An attractive feature of the suggested approach is that by successively correcting the predicted prices with forecast errors it can control error tolerance of forecasting to certain degree. Moreover, the confidence interval estimation provides desired reliability information of the forecasted prices.

2 ARIMA model approach for electricity price forecasting

According to stochastic process theory [20, 21], the time evolution of market clearing price (MCP) in electricity spot markets can be considered as a stochastic time-series with equal time intervals. It is well known that stochastic time-series can be generally divided into two families, i.e. stationary and nonstationary stochastic processes. The ARMA model is suitable to describe a stationary stochastic process $X(t)$ with its mean EX_t and variance DX_t as constants and its autocorrelation function $R(\tau)=E[X_t, X_{t+\tau}]$ relevant only to τ and not to t. Under certain conditions, ARMA can be reduced to autoregressive (AR) or moving-average (MA) models. On the other hand, the ARIMA model is superior to describe a nonstationary stochastic processes $P(t)$. The electricity MCP time-series generally belongs to a nonstationary stochastic process. Hence the electricity price forecasting issue can be resolved through the ARIMA model approach, i.e. to evaluate the parameters in the appropriate ARIMA model established for MCP time series based on historical records and then use the ARIMA model obtained for future price forecasting. The ARIMA model parameters can be updated along with the time.

2.1 ARIMA model

Suppose that price series $\{P_t\}=\{P_1, P_2, \ldots, P_n\}$ is a nonstationary stochastic process denoted as $P(t)$, which is composed of three components and described by

$$P(t) = f(t) + g(t) + X(t) \tag{1}$$

where $f(t)$ is a nonperiodic trend component (reflecting price growth, etc.), $g(t)$ a periodical component (representing periodic price changes), and $X(t)$ a random noise component (considering random fluctuation of prices); $X(t)$ is usually represented as a stationary time-series. Equation (1) shows that through extracting (or eliminating) the periodic component and nonperiodic trend component from a nonstationary stochastic process, the remaining stationary time-series can be obtained. The separation processes are as follows.

2.1.1 Difference operator to eliminate trend component: A common practice of eliminating the trend component in (1) is to apply difference operations [20]. By applying the first-order difference operator ∇ to the original series $\{P_t\}=\{P_1, P_2, \ldots, P_n\}$ we get a new time-series $\{\nabla P_t\}=\{P_2-P_1, P_3-P_2, \ldots, P_n-P_{n-1}\}$, which satisfies

$$\nabla P_t = P_t - P_{t-1} = (1-B)P_t \tag{2}$$

where B is the back-shift operator. It is clear that $BP_t=P_{t-1}$ and $\nabla=1-B$. If there is no periodic component in $P(t)$, then the difference operator ∇ can be applied repeatedly until a stationary process is arrived. The dth-order difference operator is defined as ∇^d: $\nabla^d P_t=\nabla(\nabla^{d-1}P_t)$, with $\nabla^0 P_t=P_t$. And the kth-order back-shift operator can be defined similarly as B^k: $B^k P_t=P_{t-k}$. It is easy to show that the first-order difference operator can eliminate linear trend component, and the nth-order difference operator aims at eliminating nth-order trend component. In practice the order of trend components can often be taken as one or two [26]. For short-term electricity price forecasting we limit the order of trend components to three.

2.1.2 Periodic difference operator to eliminate periodic component: If a daily average price series has one periodic component, say with seven days as a period, by the operation of $\{(1-B^7)P_t\}$ or more generally $\{(1-\theta_7 B^7)P_t\}$ where θ_7 is a proper constant, the original time-series can be converted to a stationary series with this periodic component filtered out. Spot-market clearing-price time-series usually include periodic components with one day, one week and one month as periods.

After these two types of operations, nonperiodic trend components $f(t)$ and periodic components $g(t)$ are eliminated from $P(t)$ and we obtain the remaining stationary time-series of $X(t)=P(t)-f(t)-g(t)$. The proposed ARIMA model for electricity price forecasting is a general form to implement the operations mentioned. For a complex nonstationary electricity price series $\{P_t\}$, a general $ARIMA(P, Q, D)$ model is given by

$$\Phi(B^s)\nabla_s^D P_t = \Theta(B^s)A_t \tag{3}$$

where s is the period of P_t, D the order of growth trend, and

$$\begin{aligned}\nabla_s &= 1-B^s,\ \nabla_s^D = (1-B^s)^D\\ \Phi(B^s) &= 1-\Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_P B^{Ps}\\ \Theta(B^s) &= 1-\Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs}\end{aligned} \tag{4}$$

with $\Phi_1, \Phi_2, \ldots, \Phi_P$, and $\Theta_1, \Theta_2, \ldots, \Theta_Q$ as constants; P and Q are the orders of $\Phi(B^s)$ and $\Theta(B^s)$. If $D=0$, (3) should be replaced by

$$\Phi(B^s)P_t = c + \Theta(B^s)A_t \tag{5}$$

where A_t is another $ARIMA(p, q, d)$ model taking the following form:

$$\varphi(B)\nabla^d A_t = \theta(B)\varepsilon_t \tag{6}$$

with

$$\varphi(B) = 1-\varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$$

$$\theta(B) = 1-\theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

where $\varphi_1, \varphi_2, \ldots, \varphi_p$ and $\theta_1, \theta_2, \ldots, \theta_q$ are constants, p and q are orders of $\varphi(B)$ and $\theta(B)$; d the order of growth trend of A_t; and ε_t a white-noise stochastic series with mean 0 and variance σ^2. Combining (3) and (6), yields [21]

$$\varphi(B)\Phi(B^s)\nabla^d\nabla_s^D P_t = \theta(B)\Theta(B^s)\varepsilon_t \tag{7}$$

Equation (7) is called a multiplied seasonal ARIMA model with the order of $(p, q, d) \times (P, Q, D)$, where the transformation of $\nabla^d \nabla_s^D P_t$ results in a stationary time-series with nonperiodic trend components and periodic components filtered out.

2.2 ARIMA based price forecasting

Suppose the price series $\{P_t,\ t \in T\}$ can be described as (7) with $D > 0$. Denote $X_t = \nabla^d \nabla_s^D P_t$; X_t is a stationary series which is actually an $ARMA(p', q')$ series with the form

$$\varphi'(B) X_t = \theta'(B) \varepsilon_t \tag{8}$$

where $\varphi'(B) = \varphi(B)\Phi(B^s)$ and $\theta'(B) = \theta(B)\Theta(B^s)$ have the same format as $\varphi(B)$ and $\theta(B)$ in (6) but with orders as p' and q'. When $q' = 0$ or $p' = 0$ it will reduce to an autoregressive model of $AR(p')$ or a moving average model of $MA(q')$, respectively.

According to stochastic process theory, any $ARMA(p, q)$ time-series (see (8) with superscripts omitted) can make lth-step forecast based on the difference equation according to the criterion of *minimum prediction error expectation* [21]

$$\tilde{X}_k(l) - \sum_{j=1}^{l-1} \varphi_j \tilde{X}_k(l-j) = \sum_{j=1}^{p} \varphi_j X_{k-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{k-j+1} \tag{9}$$

where X_i is a real stationary time-series element at t_i obtained based on historical data $P(t)$ and after the two operations mentioned, therefore all $X_i (i \le k)$ are known; and $\tilde{X}_k(l) = \tilde{X}(k+l)$ is the lth time-step forecast value from known X_k; $\tilde{X}_k(j),\ j = 1, 2, \ldots, l$ can be calculated successively from (9) when $X_i\ (i \le k)$ are known. After $\tilde{X}_k(l)$ has been forecast, based on the relation of (see (7) and (8))

$$\tilde{X}_k(l) = \nabla^d \nabla_s^D \tilde{P}_k(l), \quad k > d + D,\ l \ge 0 \tag{10}$$

we make price forecasting without difficulty through

$$\tilde{P}_k(l) = \psi^{-1}(B) \tilde{X}_k(l) \tag{11}$$

where $\psi(B) = \nabla^d \nabla_s^D = (1-B)^d (1-B^s)^D$. As for the coefficients in the functions of $\varphi(B)$ and $\theta(B)$, we determine them by an regressive analysis approach based on the least-square error criterion using a large number of historical MCP records without difficulty.

3 Extended ARIMA model approach with accuracy improvement by predicted error correction

To improve accuracy in forecasting the ARIMA approach has been extended to make further accuracy improvement through successive error correction. The approach of error correction follows. For convenience of illustration, the time domain involved in electricity price forecasting is denoted by $\{D = D^- + D^+\}$, where D^- is the historical time domain with relevant electricity prices known and denoted as $\{P^-(t),\ t \in D^-\}$ and D^+ the 'future' time domain for price forecasting with corresponding electricity prices unknown and denoted as $\{P^+(t),\ t \in D^+\}$. We use D^- domain data to establish the ARIMA model used for D^+ domain forecasting.

Assume an ARIMA model M for price forecasting has been established according to Section 2 with proper orders of (P, D, Q) and parameters known from analysing autocorrelation function (ACF) and partial correlation function (PCF) over a large number of historical prices in D^-. Using model M we also make price 'forecasting' in domain D^- and get a *forecast* price series in D^- as $\tilde{P}^-(t)$ which leads to an error series $\{E_1^-(t),\ t \in D^-\}$: $E_1^-(t) = P^-(t) - \tilde{P}^-(t)$ that reflects the accuracy of model M in forecasting. This error series also forms a stochastic time-series that can be used to establish an ARIMA or ARMA model M_{E1} for error forecasting. Based on price forecasting model M and the corresponding error forecasting models $M_{Ei}\ (i = 1, 2, \ldots)$ (subscript i means the ith error-correction model), the following procedures of successive error correction can be applied to improve price forecasting accuracy.

(i) Forecast the price $\tilde{P}^-(t)$ at time $t\ (t \in D^-)$ based on model M (see (8)–(11)), and get corresponding error series $E_1^-(t),\ t \in D^-$, which is used to establish ARIMA or ARMA model M_{E1}.

(ii) Use the model M_{E1} to forecast the model error $\tilde{E}_1^-(t)$ at domain D^-.

(iii) Modify the forecasted price series at $t \in D^-$ by

$$\tilde{P}_1^-(t) = \tilde{P}^-(t) + \tilde{E}_1^-(t), \quad t \in D^- \tag{12}$$

(iv) Estimate the confidence interval (CI) of the forecasted price $\tilde{P}_1^-(t)$ (the technique is illustrated in Section 4), and check whether the range of CI reaches the desired accuracy. If not, calculate the residual error time-series after the correction $E_2(t) = P^-(t) - \tilde{P}_1^-(t),\ t \in D^-$ and go to step (v) for further error correction successively; otherwise go to step (viii).

(v) Set up the residual error forecasting model M_{E2} for $E_2(t)$ and predict the residual error $\tilde{E}_2^-(t)$.

(vi) Make further error correction to the forecast price through

$$\begin{aligned}\tilde{P}_2^-(t) &= \tilde{P}^-(t) + \sum_{i=1}^{2} \tilde{E}_i^-(t) \\ &= \tilde{P}_1^-(t) + \tilde{E}_2^-(t), \quad t \in D^-\end{aligned} \tag{13}$$

(vii) Compute new CI and check again whether the range of CI corresponding to $\tilde{E}_2^-(t)$ is acceptable. If not, calculate the current residual error $E_3(t) = P^-(t) - \tilde{P}_2^-(t)$ and repeat steps (v)-(vii) for further error correction; otherwise go to step (viii).

(viii) Integrate ARIMA model M and models $M_{Ei}\ (i = 1, 2, \ldots, I)$ to form the finally used MCP forecasting model in the domain $t \in D^+$ i.e. to forecast $(\tilde{P}^+(t),\ t \in D^+)$ through

$$\tilde{P}^{*+}(t) = \tilde{P}^+(t) + \sum_{i=1}^{I} \tilde{E}_i^+(t), \quad t \in D^+ \tag{14}$$

The entire forecasting process is shown in Fig. 1.

Extending the conventional ARIMA approach through error corrections can apparently improve the accuracy of forecasting and limit the CI range to a certain extent, which yields a compromise of computation time and accuracy in price forecasting. That is a very attractive feature of the proposed approach.

4 Confidence interval estimation based on forecast error

For highly volatile electricity prices it is desired to know not only the forecast price but also its confidence interval (CI). This provides a range with true MCP located inside the range at a given probability (or confidence degree CD). For a given or desired confidence degree, the smaller the range

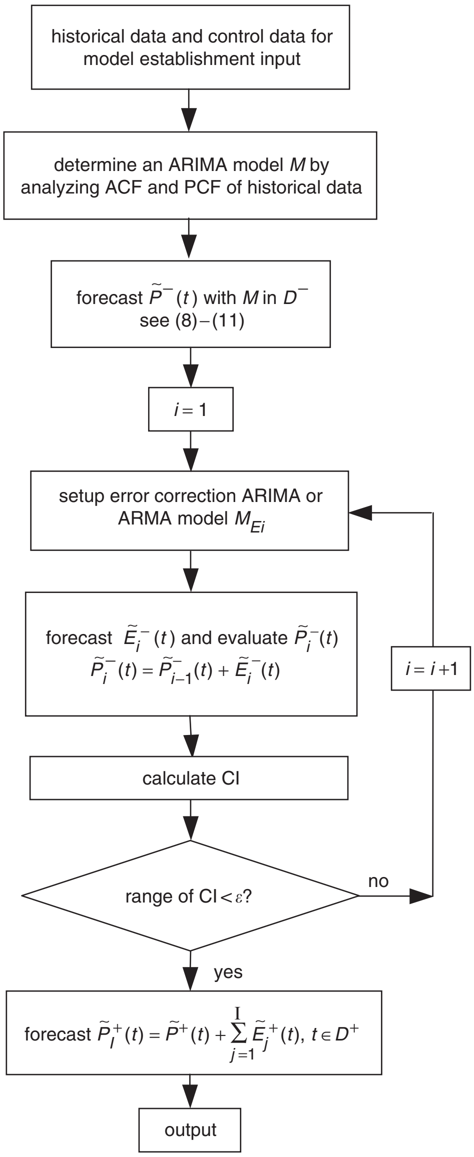

Fig. 1 *Flowchart for model establishment and price forecasting*

of confidence interval is, the higher the accuracy of the forecast price will be.

Generally MCP prediction errors may come from model mismatches and data noise. It is reasonable to assume that in our case the model mismatch dominates the errors with historical data correctly recorded without noise. Based on Section 3, we know that the residual errors of model M at domain D^- have been used to forecast errors and make error corrections. This makes it possible to use the forecast residual errors to compute the confidence interval. The residual errors of corresponding models are defined as (Fig. 2):

$$\begin{aligned}
\text{for } M: \quad & E_{r0}^-(t) = P^-(t) - \tilde{P}^-(t) = E_1^-(t) \approx \tilde{E}_1^-(t) \\
\text{for } (M+M_{E1}): \quad & E_{r1}^-(t) = P^-(t) - \tilde{P}^-(t) - \tilde{E}_1^-(t) \\
& \qquad = E_2^-(t) \approx \tilde{E}_2^-(t) \\
\text{for } (M+M_{E1}+M_{E2}): \quad & \cdots\cdots
\end{aligned} \tag{15}$$

Suppose that the statistical features of errors can be extended from D^- to D^+, then the forecast error series $\tilde{E}_i(t)$ ($i = 1, 2, \ldots, t \in D^- + D^+$) can be viewed as the error samples of the corresponding model to compute the confidence interval. The details are given as follows.

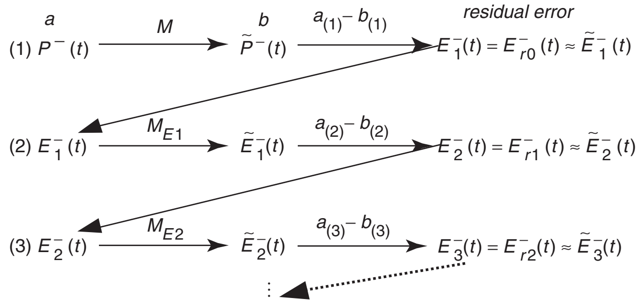

Fig. 2 *Relationship between residual and forecast errors*

Assume that the price series $\tilde{P}^+(t)$ and error series $\tilde{E}_i^+(t)$ ($t = 1, 2, \ldots, n$, and $i = 1, 2, \ldots, I$) have been forecast. Taking $CD = (1-\alpha)$ ($0 < \alpha < 1$), the range of confidence interval CI corresponding to the ith forecasting error $\tilde{E}_i^+(t)$ is denoted as $CI_{CD,i} = [L_i, U_i]$, where L_i and U_i are the lower and upper bounds of CI. Two steps are required to determine $CI_{CD,i}$. The first step is to determine the distribution function or probability density function (PDF) of the population (i.e. the error sampling set of $\tilde{E}_i^+(t)$); and the second step is to calculate the CI range $CI_{CD,i} = [L_i, U_i]$ based on the error sampling set and PDF used. The true MCP will locate inside the range at a probability of ($CD \times 100\%$).

4.1 Determine distribution function of population of $\tilde{E}_i^+(t)$

The determination of the distribution function of the population of $\tilde{E}_i^+(t)$ is a matter of mathematical statistics, based on large numbers for simplicity we assume the following lemma holds.

Lemma 1: If the error correction is effective enough, and the residual error sample set is large enough (e.g. $n \to \infty$), the distribution function of the population of $\tilde{E}_i^+(t)$ can be described as a gaussian distribution. However, if the sample space is not large enough, or the error correction is not effective enough, the distribution function of the population of $\tilde{E}_i^+(t)$ could be taken conservatively as uniform distribution.

Both gaussian and uniform distribution are tested for CI estimation. In general, for a well-established prediction model, it is appropriate to consider the model error as a gaussian distribution. However, it might lead to an optimistic CI estimation when the model error correction is not enough. On the other hand, the uniform distribution is relatively conservative and will lead to wider range of CI.

4.2 Estimate confidence interval

Assume a gaussian distribution is taken as the distribution function of the population of error samples $\{E_1, E_2, \ldots, E_n\}$ from $\tilde{E}_i^+(t)$, and $\tilde{P}^+(t)$ is the corresponding forecast prices series. For a continuous gaussian distribution error (Fig. 3), it is clear that if the required CD is 0.99, the corresponding CI range is $2d = 6\sigma$ (since $\int_{\mu-3\sigma}^{\mu+3\sigma} (PDF)dx \approx 0.99$). When $\mu \neq 0$ (μ is the mean of error $\tilde{E}_i^+(t)$), we have

$$CI_{CD=0.99} = [L_i, U_i] = [\tilde{P}^+(t) + \mu - d,\ \tilde{P}^+(t) + \mu + d]$$

which means for the real $\tilde{P}^+(t)$ the probability of its location in the range of $[L_i, U_i]$ is 99%. It is clear that the

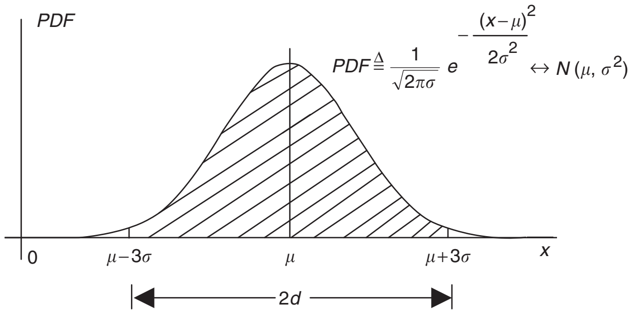

Fig. 3 *Gaussian probability distribution function*

larger the requested CD, the wider the CI range ($2d$) is. The effective way to make $\tilde{P}^+(t)$ have small CI range is to reduce residual error $\tilde{E}_i^+(t)$, which can be realised through error correction.

Usually $x \sim N(\mu, \sigma^2)$ distribution can be normalised to $x' \sim N(0, 1)$ through transformation of (16)

$$x' = \frac{(x-\mu)}{\sigma} \tag{16}$$

For the latter it is easy to check out the CI range ($2d$) (in the x′-domain) with respect to given CD from the standard Gauss distribution function table [21]. Therefore we map it back to the x-domain to get the real CI range ($2d$).

For a discrete error sample $\{E_1, E_2, \ldots, E_n\}$ we calculate its μ and σ^2, and then make the CI range ($2d$) estimation similar to a continuous gaussian distribution case.

For uniform distribution assumption it is easy to derive the CI range for a given CD as

$$\begin{aligned} 2d &= CD \times \left[\max(\tilde{E}_i^+) - \min(\tilde{E}_i^+)\right] \\ CI_{CD} &= [\tilde{P}^+(t) + \bar{E} - d, \tilde{P}^+(t) + \bar{E} + d] \end{aligned} \tag{17}$$

where $\bar{E}$ is taken as the mean of error samples $\{E_1, E_2, \ldots, E_n\}$.

5 Computer tests

5.1 ARIMA model for forecasting hourly MCP

The hourly MCP forecasting of Californian Power Market (CPM) is used as an example to verify the effectiveness of the extended ARIMA approach. The data used of Area SP15 in CPM is downloaded from the website [26]. For historical data of the CPM, through experienced analysis and computation of autocorrelation function and partial correlation function, the ARIMA model M for MCP forecasting is selected. Then through successive error correction by the ARMA model of M_{Ei} ($i=1, 2$), the accuracy of forecast prices can be further improved with residual errors used to estimate CI with respect to the given *CD*. The error-correction step allows the use of the simplified ARIMA model shown in (18)

$$(1-\phi_1 B - \phi_2 B^2)(1-B)(1-B^{24})P_t = (1-\theta_1 B)\varepsilon_t \tag{18}$$

where ε_t is a residual random error. The model (18) is a simple ARIMA model which reflects the main characteristics, e.g. the trend component ($1-B$, supposed to be a linear trend), the periodical component ($1-B^{24}$) with 24 hours as period, and the closer relationship with recent three hours' prices; ϕ_1, ϕ_2 and θ_1 are worked out by regressive analysis of the large amount of historical data.

5.2 Hourly MCP forecasting results of CPM

The price forecasting results of two different period of CPM are presented.

Case 1: The first case is to use the historical prices of 10 days from March 3 to 12 1999 to establish the forecasting model M as (19) and two error forecasting models M_{E1} and M_{E2} as (20) and (21).

$$\begin{aligned} P(t) = {} & 1.2413P(t-1) - 0.5774P(t-2) + 0.3361P(t-3) \\ & + P(t-24) - 1.2413P(t-25) + 0.5774P(t-26) \\ & - 0.3361P(t-27) - 0.6149\varepsilon(t-1) \end{aligned} \tag{19}$$

$$E_1(t) = 0.3804E_1(t-1) - 0.3114\varepsilon_{t1} \tag{20}$$

$$E_2(t) = 0.6544E_2(t-1) - 0.4563\varepsilon_{t2} \tag{21}$$

Equation (19) shows that the value at time t has a close relation with the latest three values at times $t-1$, $t-2$, $t-3$ as well as those values of previous day near the same daily time, i.e. $t-24$, $t-25$, $t-26$ and $t-27$, since the period is 24 hours and hourly MCPs are to be forecast. The coefficients in (19) referring to the correlation degree are obtained by regressive analysis.

The data of three days from March 13 to 15 are used to forecast the prices of the next day, March 16. The forecast results are shown in Fig. 4. The *time* axis in the Figure

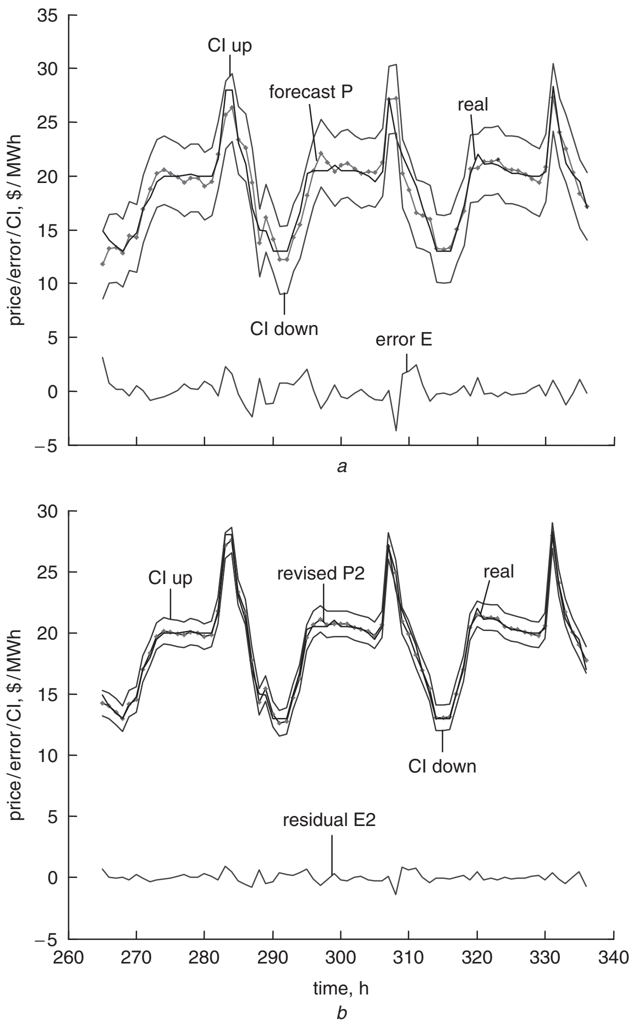

Fig. 4 *Computer results for case 1*
a Without error correction
b With error correction twice

Table 1: Error statistics for case 1

	MaxAE ($/MWh)	MaxRE (%)	Mean ($/MWh)	Variance ($/MWh)^2	MSE ($/MWh)^2	Mean-RE (%)	$2d$
E_0	3.6999	20.862	0.7737	1.1556	1.1419	4.12	4.21
E_{r1}	2.5485	12.93	0.5312	0.5342	0.5276	2.82	2.86
E_{r2}	1.3811	5.88	0.2760	0.1468	0.1447	1.45	1.51

E_0: no error correction; E_{r1}: with once error correction; E_{r2}: with twice error correction

shows the last three-day data of 14 days. In Fig. 4 'real' prices, 'predicted' prices by model M, and the prices of 'revised P2', i.e. prices after two error corrections are presented together with corresponding residual 'error' plots. The residual error series E (or $E2$) in Fig. 4 is used to work out the confidence interval $CI_i = [L_i, U_i]$ at $CD = 0.95$, also plotted in Fig. 4. In Fig. 4*a*, no errorcorrection is made and the error E is large, therefore uniform distribution function is used for CI estimation. We see that all 'real' prices locate inside the range of $CI_i = [L_i, U_i]$, but the CI range ($2d$) is very large which means the accuracy of forecasting is very poor. In Fig. 4*b*, after two error corrections the residual error ($E2$) decreases significantly, the revised values P2 almost coincide with the real values. Gaussian distribution is assumed and $CI_i = [L_i, U_i]$ at $CD = 0.95$ corresponding to $E2$ has been worked out with smaller range of ($2d$). It can be seen that the range of CI is very narrow, but the 'real' prices still lie well inside the CI range, which implies that the accuracy of forecasting improvement significantly.

Error statistics for case 1 are shown in Table 1. In Table 1, maximum absolute error (MaxAE), maximum relative error (MaxRE), mean, variance, mean square error (MSE) and mean relative error (Mean-RE) are calculated based on the following definitions:

MaxAE = max $\{abs(E_{ri}(t))\}$, $i = 0, 1, 2$, $t \in D$, time domain
MaxRE = $\max(E_{ri}(t)/P_t)$, P_t is the real price at time t
MSE = $\frac{1}{n}\sum_{t=1}^{n} E_{ri}^2(t)$, and mean-RE = $\frac{1}{n}\sum_{t=1}^{n} \frac{E_i(t)}{P_t}$, n is the number of prediction space with $n = 96$ in this example

Table 1 shows that the prediction accuracy has been improved significantly along with error corrections, and all indices in Table 1 support the suggested method very well, which show clearly the effectiveness of the proposed extended ARIMA approach.

Case 2: The second example is under the worse condition at the end of 2000 during the period of a so-called 'electricity crisis'. Ten-day historical data from Nov. 2 to 11, 2000 are used to set up the forecasting model and use the successive three-day data to forecast the prices of the next day, Nov. 15. The relevant forecast models are shown in (22), (23) and (24)

$$P(t) = 0.3655P(t-1) - 0.0733P(t-2) + 0.7078P(t-3) + P(t-24) - 0.3655P(t-25) + 0.0733P(t-26) - 0.7078P(t-27) - 0.3124\varepsilon(t-1) \qquad (22)$$

$$E_1(t) = 0.3444E_1(t-1) - 0.7043\varepsilon_{t1} \qquad (23)$$

$$E_2(t) = 0.5492E_1(t-1) - 0.6769\varepsilon_{t2} \qquad (24)$$

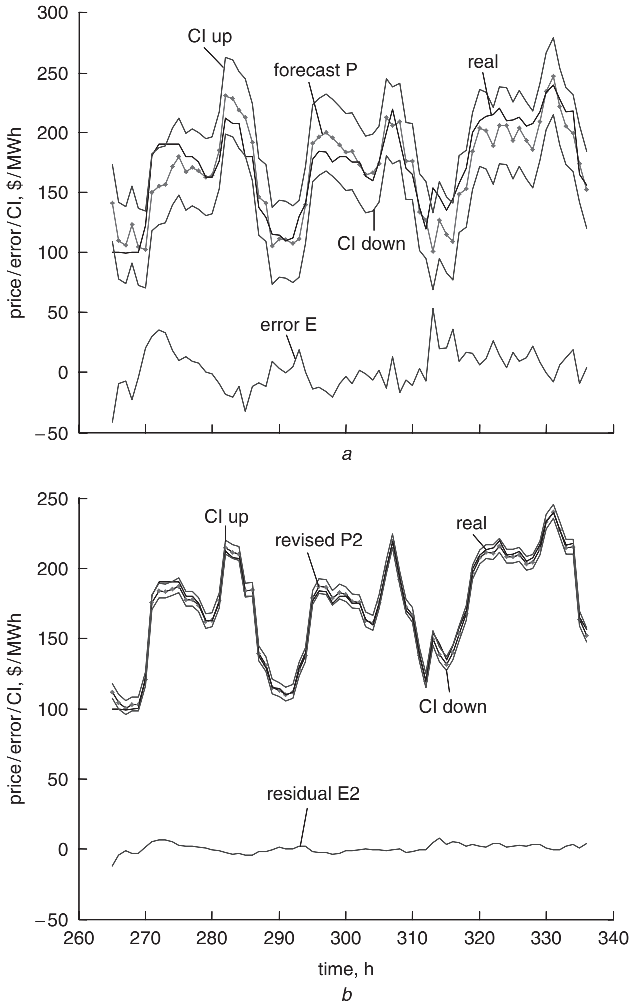

Fig. 5 *Computer results for case 2*
a Without error correction
b With twice error correction

Equation (22) has similar meaning as (19) except the coefficients are different.

The results are shown in Fig. 5 in the same way as Fig. 4 of case 1. Error statistics of case 2 are shown in Table 2 with the same index definitions as Table 1. It shows that the suggested method can still work well even in the worst market condition.

The real prices, the forecast ones by the conventional ARIMA method, and the forecast ones by the proposed extended ARIMA approach are listed in Tables 3 and 4 for March 16, 1999 (with normal daily MCPs) and Nov. 15, 2000 (with more changeable daily MCPs during energy crisis), respectively. Conventional ARIMA models refer to the models described by (19) and (22)

Table 2: Error statistics for case 2

	MaxAE ($/MWh)	MaxRE %	Mean ($/MWh)	Variance ($/MWh)^2	MSE ($/MWh)^2	Mean-RE (%)	$2d$
E_0	53.1744	41.07	13.2067	275.56	279.3133	8.08	65.1
E_{r1}	26.9260	26.93	6.3067	64.7542	65.8066	3.93	31.5
E_{r2}	12.1382	12.14	2.5059	10.2464	10.3470	1.57	12.5

E_0: no error correction; E_{r1}: with once error correction; E_{r2}: with twice error correction

Table 3: Comparison of extended ARIMA approach with conventional ARIMA method of 1999 in CPM

Time (hour)	Real price ($/MWh)	Conventional ARIMA		Extended ARIMA	
		price ($/MWh)	error (%)	price ($/MWh)	error (%)
1	15.0086	15.9746	6.4363	15.4461	2.9151
2	12.9983	13.2879	2.2280	13.0476	0.3796
3	12.9944	13.2067	1.6338	13.0439	0.3810
4	12.9958	13.3253	2.5354	13.0987	0.7921
5	14.9947	15.0812	0.5769	14.9995	0.0319
6	17.2455	16.7289	2.9956	17.0414	1.1833
7	20.2496	20.6636	2.0445	20.4410	0.9454
8	21.9922	20.7617	5.5952	21.5064	2.2092
9	21.0672	21.3668	1.4221	21.2667	0.9469
10	21.1756	21.3801	0.9657	21.2432	0.3191
11	20.9963	21.5449	2.6128	21.1856	0.9014
12	20.5524	20.6269	0.3625	20.5373	0.0735
13	20.2498	20.5800	1.6306	20.3613	0.5507
14	20.1943	20.4819	1.4242	20.2756	0.4025
15	20.0392	20.1900	0.7525	20.0704	0.1555
16	19.9920	19.7272	1.3245	19.8778	0.5713
17	19.9908	19.4191	2.8598	19.7940	0.9847
18	20.3654	20.7951	2.1100	20.5711	1.0100
19	28.3267	27.3297	3.5196	27.9285	1.4058
20	23.9917	24.0484	0.2363	24.0830	0.3807
21	21.2039	22.5308	6.2578	21.7070	2.3728
22	20.0694	20.3018	1.1580	20.0595	0.0492
23	19.4757	18.3538	5.7605	19.0239	2.3199
24	17.0309	17.1956	0.9671	17.0956	0.3799
average			2.3921		0.9026

without error correction, while the extended ARIMA models are based on (19)–(21) and (22)–(24) with two error corrections. It can be seen that error correction can evidently improve the accuracy of forecasting. Compared with other methods we should say that without error correction the MCP forecasting accuracy is similar to other conventional methods, however, with error correction the suggested approach is superior. In addition, the confidence interval can be derived based on it easily, which is an attractive feature of the new method.

6 Conclusions

An extended ARIMA approach to forecasting hourly MCP incorporating estimation of confidence interval has been proposed. The suggested error correction is very effective in improving the accuracy of price forecasting, which can also allow the use of a relatively simple or lower-order ARIMA model at the beginning. Since the residual error can be effectively reduced through error correction, gaussian and normal distribution assumptions can be used for CI estimation and the resultant CI range will be small. Computer tests on an hourly MCP forecast of CPM confirm the features mentioned and show that the proposed approach is effective not only in a normal power market condition, but also under a price-changeable period. Further research will be done to evaluate the worst MaxAE, MRE, MSE and Mean-RE according to the CI and other information available in the calculation, and to extend the method to forecast locational marginal price of power markets with congestion.

Table 4: Comparison of extended ARIMA approach with conventional ARIMA method of 2000 in CPM

Time (hour)	Real price ($/MWh)	Conventional ARIMA		Extended ARIMA	
		price ($/MWh)	error (%)	price ($/MWh)	error (%)
1	153.9767	100.8023	34.5341	149.7080	2.7723
2	145.9988	126.6263	13.2689	138.3589	5.2328
3	135.0593	114.5882	15.1571	131.5968	2.5637
4	145.0008	109.2044	24.6870	139.5764	3.7409
5	157.6268	148.4249	5.8378	153.0966	2.8740
6	170.0073	153.1039	9.9427	167.7879	1.3055
7	200.0127	183.9244	8.0436	196.7392	1.6366
8	209.7646	203.8095	2.8389	207.6386	1.0135
9	213.0038	201.3826	5.4559	211.4238	0.7418
10	214.7268	188.4494	12.2376	211.0205	1.7261
11	219.9142	205.6467	6.4878	215.8489	1.8486
12	209.7868	205.4293	2.0771	208.1175	0.7957
13	209.9954	193.3020	7.9494	208.0743	0.9148
14	212.2327	203.2075	4.2525	209.6121	1.2348
15	205.1474	193.1260	5.8599	203.2075	0.9456
16	207.5907	185.5265	10.6287	204.2824	1.5937
17	217.1714	209.0084	3.7588	214.1424	1.3947
18	233.5328	233.7761	0.1042	232.9135	0.2652
19	239.2509	246.8945	3.1948	240.0847	0.3485
20	227.3526	221.4309	2.6046	227.5974	0.1077
21	216.6697	204.7410	5.5055	214.8114	0.8576
22	218.0610	199.0957	8.6972	215.0342	1.3881
23	164.8094	173.9174	5.5264	163.7711	0.6300
24	156.2563	152.0682	2.6803	153.9219	1.4939
average			8.3888		1.5594

7 Acknowledgments

This work was supported in part by the Specialised Research Fund for the Doctoral Program of Higher Education under the Ministry of Education of the People's Republic of China (no. 20040079002), and the RGC of HK Government.

8 References

1 Benini, M., Marracci, M., Pelacchi, P., and Venturini, A.: 'Day-ahead market price volatility analysis in deregulated electricity markets'. Proc. of IEEE Power engineering society summer meeting, 2002, Vol. 3, pp. 1354–1359
2 Li, Y., and Flynn, P.C.: 'Power price in deregulated markets'. Proc. of IEEE Power engineering society general meeting, 2003, Vol. 2, pp. 874–879
3 Bunn, D.W.: 'Forecasting loads and prices in competitive power markets', *Proc. IEEE*, 2000, **88**, (2), pp. 163–169
4 Hong, Y.Y., and Hsiao, C.Y.: 'Locational marginal price forecasting in deregulated electricity markets using artificial intelligence', *IEE Proc. Gener. Transm. Distrib.*, 2002, **149**, (5), pp. 621–626
5 Szkuta, B.R., Sanabria, L.A., and Dillon, T.S.: 'Electricity price short-term forecasting using artificial neural networks', *IEEE Trans. Power Syst.*, 1999, **14**, (3), pp. 851–857
6 Gao, F., Guan, X.H, Cao, X.R., and Papalexopoulos, A.: 'Forecasting power market clearing price and quantity using a neural network method'. Proc. of IEEE Power engineering society summer meeting, 2000, Vol. 4, pp. 2183–2188
7 Li, Zhang, Luh, P.B., and Kasiviswanathan, K.: 'Energy clearing price prediction and confidence interval estimation with cascaded neural networks', *IEEE Trans. Power Syst.*, 2003, **18**, (1), pp. 99–105
8 Guo, J.-J., and Luh, P.B.: 'Market clearing price prediction using a committee machine with adaptive weighting coefficients'. Proc. of IEEE Power Engineering Society Winter Meeting, 2002, Vol. 1, pp. 77–82
9 Niimura, T., Ko, H.S., and Ozawa, K.: 'A day-ahead electricity price prediction based on a fuzzy-neuro autoregressive model in a deregulated electricity market'. Proc. of 2002 International Joint Conf. on Neural Networks, 2002, Vol. 2, pp. 1362–1366
10 Luh, P.B.: 'An integrated neural network method for market clearing price prediction and confidence interval estimation'. Proc. of 4th World Congress Control and Automation, 2002, Vol. 3, pp. 2045–2050
11 Zhang, P., and Guan, X.H.: 'Fuzzy modelling for electricity market price forecasting'. Proc. of 3rd World Congress on Intelligent Control and Automation, 2000, Vol. 3, pp. 2262–2266
12 Obradovic, Z., and Tomsovic, K.: 'Time-series methods for forecasting electricity market pricing'. Proc. of IEEE Power Engineering Society Summer Meeting 1999, Vol. 2, pp. 1264–1265
13 Ni, E., and Luh, P.B.: 'Forecasting power market clearing price and its discrete PDF using a Bayesian-based classification method'. Proc. of IEEE Power Engineering Society Winter Meeting, 2001, Vol. 3, pp. 1518–1523
14 Contreras, J., Espinola, R., Nogales, F.J., and Conejo, A.J.: 'ARIMA models to predict next-day electricity prices', *IEEE Trans. Power Syst.*, 2003, **18**, (3), pp. 1014–1020
15 Zhou, M., Yan, Z., Ni, Y.X., and Li, G.Y.: 'An ARIMA approach to forecasting electricity price with accuracy improvement by predicted errors'. Proc. of IEEE Power Engineering Society General Meeting, 2004
16 Ling, S.H., Leung, F.H.F., Lam, H.K., Lee, Y.S., and Tam, P.K.S.: 'A novel genetic-algorithm-based neural network for short-term load forecasting', *IEEE Trans. Ind. Electron.*, 2003, **50**, (4), pp. 793–799
17 Huang, S.J., and Shih, K.R.: 'Short-term load forecasting via ARMA model identification including nongaussian process considerations', *IEEE Trans. Power Syst.*, 2003, **18**, (2), pp. 673–679
18 Silva, P., Alvesda, and Moulin, L.S.: 'Confidence intervals for neural-network based short-term load forecasting', *IEEE Trans. Power Syst.*, 2000, **15**, (4), pp. 1191–1196
19 Papadopoulos, G., Peter, J.E., and Murray, A.F.: 'Confidence estimation methods for neural networks: a practical comparison', *IEEE Trans. Neural Netw.*, 2001, **12**, (6), pp. 1278–1287
20 Brockwell, P.J., and Davis, R.A.: 'Introduction to time series and forecasting' (Springer, New York, 2002)

21 Ma, Z.H.: 'Probability statistics and stochastic process' (Tsinghua University Press, Beijing, 2002)
22 Wang, H., and Leu, J.Y.: 'Stock market trend prediction using ARIMA-based neural networks'. Proc. of IEEE Int. Conf. on Neural Networks, 1996, Vol. 4, pp. 2160–2165
23 Kalpakis, D., Gada, and Puttagunta, V.: 'Distance measures for effective clustering of ARIMA time-series'. Proc. of IEEE Int. Conf. on Data Mining, 2001, pp. 273–280
24 Desouky, A., El, and Elkateb, M.M.: 'Hybrid adaptive techniques for electric-load forecast using ANN and ARIMA', *IEE Proc. Gener., Transm. Distrib.*, 2000, **147**, (4), pp. 213–217
25 Huang, S.J., and Shih K., R.: 'Short-term load forecasting via ARMA model identification including non-Gaussian process considerations', *IEEE Trans. Power Syst.*, 2003, **18**, (2), pp. 673–679
26 http://www.ucei.berkeley.edu/datamine/datamine.htm