2007 IEEE International Conference on Control and Automation Guangzhou, CHINA - May 30 to June 1, 2007

WeD4.2

Medium Term Horizon Market Clearing Price and Load Forecasting With an Improved Dual Unscented Kalman Filter

A. Karsaz
Department of Electrical Engineering
Ferdowsi University
Mashhad, Iran
a_karsaz1@yahoo.com

H. Khaloozadeh
Department of Electrical Engineering
Kh. N. Toosi University
Tehran, Iran
h_khaloozadeh@yahoo.com

Abstract— **The deregulation of electric power supply industries has raised many challenging problems. One of the most important ones is forecasting the Market Clearing Price (MCP) of electricity. Decisions on various issues, such as to buy or sell electricity and to offer a transaction to the market, require accurate knowledge of the MCP. Another problem, which has also been an important issue of the traditional power systems, is load forecasting for both short and long terms.**
The extended kalman filter has been widely adopted for state estimation of nonlinear systems, machine learning applications and neural network training. In the EKF, the state distribution is approximated by the first-order linearization of the nonlinear system. Therefore this can introduce large errors in the load and price forecasting as two Chaotic, nonstationary and nonlinear time-series. The unscented Kalman filter (UKF), in contrast, achieves third-order accuracy, by using a minimal set of MCP and load sigma points. In this paper an improved dual unscented Kalman filter (DUKF), which estimate state and parameter simultaneously has been applied to the real New England power market. The numerical stability and more accurate predictions of our method is comparable to the EKF, and traditional neural network training methods. Remarkably, the computational complexity of the DUKF is the same order as that of the EKE. The obtained results show significant improvement in both price and load forecasting.

Keywords—Price prediction, load forecasting, deregulation, electric power market, artificial neural network, unscented filtering.

I. INTRODUCTION

The electric utility industries are undergoing a fundamental transformation from being regulated and monopolistic to becoming deregulated and competitive. This has created new issues in the operation and planning of power systems. From a supplier's viewpoint, predicting the market clearing price (MCP) of electricity is a major issue in the bidding process. This is also an important problem for the independent system operator (ISO) who is responsible for congestion management and secure operation of the system. Another important issue in the operation and planning of power systems is load forecasting.

The Kalman filter is a well-known method for recursive state estimation of linear dynamic systems, and is a minimum mean-square-error estimator. Through linearization, the extended Kalman filter (EKF) has been widely adopted for state estimation of nonlinear systems [1]. Classical (as time-series) and intelligent methods, such as artificial neural networks, have been used to predict the MCP and load forecasting [2]-[6]. For example, a modular general regression neural network (GRNN) is used to predict the next day 24 hours spot price or location marginal price (LMP) [7]. A single neural network, however, may misrepresent part of the input-output data mapping that could have been correctly represented by multiple networks. Using a "committee machine" composed of multiple networks can in principle alleviate such a difficulty [8]. A major challenge for using a committee machine is to combine the predictions of multiple networks properly. In this reference, the weighting coefficients for combining network predictions are chosen to be the probabilities that individual networks capture the true input-output relationship at that prediction instant.
As an alternative, a combination of two approaches, such as neural networks and fuzzy logic, have been used for forecasting the energy price [7] [9]. EKF_based neural network learning, and developed a novel method (called DEKF_UD) to reduce computation and improve numerical stability has been suggested [10]. The Baysian method is a minimum mean and square error predictor by using BP to minimize the cost function. The BP is a first-order algorithm, and suffers from slow convergence [11].

In this paper, a new approach based on UKF for both state and parameters estimation for MCP is introduced. The rest of the paper is organized as follows. Some background materials on price and load forecasting are described in Sections II and III. In Section IV, the UKF basic idea is briefly reviewed. The dual proposed method is described in Section V and the simulation results are presented and compared b the results obtained by other methodologies in Section VI. Section VII provides some relevant conclusions.

1-4244-0818-0/07/$20.00 ©2007 IEEE

II. LITERATURE REVIEW

There are several methods for load and price forecasting. They can be classified into the following categories.

1. Regression Methods

One approach to predict the market behavior is regression. The basic idea is to use the historical price, quantity, and other information such as load, temperature, fuel and other effective factors to predict the LMP or MCP. Assuming $x_1, x_2, ..., x_p$ are independent random variables, the output of a linear regression model, y, is obtained as:

$$y = b_0 + b_1 x_1 + ... + b_p x_p + \varepsilon \quad (1)$$

where $b_0, b_1, ..., b_p$ are unknown but Constants, and ε is a random variable with Gaussian distribution of mean 0 and variance σ^2, that is $\varepsilon = N(0, \sigma^2)$ [12].

2. Multi-layer Perceptrons (MLP)

The most widely used neural network is multi-layer perceptron (MLP). The training of MLP is commonly performed by the back-propagation (BP method) algorithm [13][14]. Although this algorithm is straightforward and usually provides good solutions, the results are not necessarily optimum. To overcome this problem, the genetic algorithm has been proposed for training MLP's. This is a more powerful technique and can provide optimum solution. However, it requires more calculation time and therefore, it is useful only for small or medium scale problems.

3. Hybrid different Neural Network Structures

Because of problems such as insufficient input-output data points and too many tunable parameters, a single neural network might misrepresent part of the nonlinear input-output relationship. This could have been more appropriately represented by using different neural networks. For example, radial-basis function (RBF) network is effective in exploiting local data characteristics, while MLP is good at capturing global data trends [8], [23]. Therefore, a committee machine consisting of different types of neural networks can in principle alleviate the misrepresentation of the input-output data relationship in a single network [8].

To obtain predictions by a committee machine, a well-known method is the ensemble averaging as depicted in Fig. 1. In this method, the predictions of neural networks are linearly combined based on a simple averaging or the statistics of historical prediction errors.

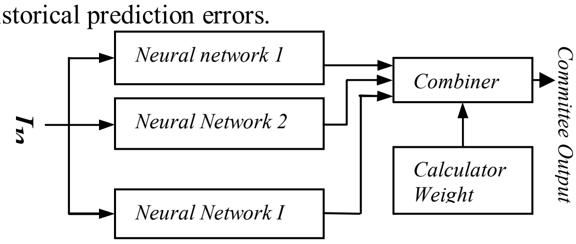

Fig. 1. Schematic of an ensemble-averaging committee machine with network predictions and weighting coefficients

4. EKF-based Neural Network method

Since EKF is a second-order learning algorithm, fast convergence with less accurate predictions is expected than traditional neural learning algorithms. In addition calculating Jacobian matrices can be very difficult and introduces numerous errors. Therefore for some cases it causes its estimation to diverge. In [10] presents a modified U-D factorization algorithm within the decoupled EKF (DEKF) framework to make the EKF more faster and numerical stability as compared to standard EKF. To providing predictions, EKF can also estimate **confidence interval (CI)** based on its innovation covariance matrix [10].

III. EKF PRICE FORECASTING

EKF has been used to train MLP networks by training weights of a network as the state of an unforced or forced nonlinear dynamic system. The training can be described as a state estimation problem with the following linear forced dynamic and nonlinear observation equations [10]:

$$\begin{aligned} P(n+1) &= f[W(n+1), L(n-l_j), P(n-p_j)] + q(n) \\ W(n+1) &= W(n) + G(n)U(n) \\ Z_P(n+1) &= P(n+1) + V(n) \end{aligned} \quad (2)$$

where $l_j, p_j \in \{0,1,......\}$ are defined as lag steps.

$$W_P(n) = [w_{1P}(n) \quad ... \quad ... \quad w_{1L_W}(n)]^T \quad (3)$$

where $W(n+1)$ is an $L_W \times 1$ state vector (neural network Weights in n+1[th] iteration with dimension; L_W)and $f[.]$ is the MLP input-output relationship. The input variable $L(n),...L(n-n1)$ and $P(n),...,P(n-n2)$ (e. g. load and price at step time n ,…,n-n2) for MCP prediction at time n+1 are available in real time. The proper values of l_i and p_i should be selected based on "sensitivity method" as discussed in [10]. Consider the basic state-space framework as in Equation 1. Given the noisy observation $P(n+1)$, a recursive estimation for $W(n+1)$ can be expressed in the following form;

$$\hat{W}(n+1/n+1) = \hat{W}(n+1/n) + K(n+1)[P(n+1) - \hat{P}(n+1/n)] \quad (4)$$

To drive the EKF formula considering input uncertainty, a first order Taylor series expansion of $f[.]$ around the estimated weights $\hat{W}(n+1/n)$ is performed;

$$\begin{aligned} P(n+1) = {} & f[\hat{W}(n), L(n-l_j), P(n-p_j)] \\ & + H(n)[W(n+1) - \hat{W}(n+1/n)] + \text{HOT} + q(\text{n}+1) \end{aligned} \quad (5)$$

In the above, $H(n)$ is the partial derivative of $f[.]$ with respect to $W(n)$ at the estimated weights. i.e., a Jacobian matrix with dimension $L_P \times L_W$:

$$H(n) = \left(\frac{\partial f[W(n+1), L(n-l_j), P(n-p_j)]}{\partial W(n)}\right)^T \Bigg|_{W(n+1)=\hat{W}(n+1/n)} \quad (6)$$

and the higher order terms (HOT) are neglected. Then the estimated output $\hat{P}(n+1/n)$ is given by

$$\hat{P}(n+1/n) = f[\hat{W}_P(n), L(n-l_j), P(n-p_j)] \quad (7)$$

with measurement residual calculated as below:

$$\begin{aligned} &\tilde{P}(n+1/n) = P(n+1) - \hat{P}(n+1/n) \cong \\ &H(n+1)[W(n+1) - \hat{W}(n+1/n)] + V(n) \end{aligned} \quad (8)$$

$$\hat{Z}(n+1/n) = \tilde{P}(n+1/n)$$

The innovation covariance $\sum_Z (n+1/n) = \Sigma_P(n+1/n) + R(n)$ is the covariance of the measurement residual $\delta_z(n+1/n) = Z_p(n+1) - \hat{Z}_p(n+1/n)$.

$$\Sigma_P(n+1/n) = H(n+1)\Sigma_w(n+1/n)H(n+1)^T + Q(n) \quad (9)$$

$$\Sigma_Z(n+1/n) = H(n+1)\Sigma_w(n+1/n)H(n+1)^T + R(n) + Q(n) \quad (10)$$

$$K(n+1) = \Sigma_w(n+1/n)H(n+1)^T \Sigma_Z(n+1/n)^{-1} \quad (11)$$

$$\Sigma_W(n+1/n+1) = [I - K(n+1)H(n+1)]\Sigma_W(n+1/n) \quad (12)$$

$$\Sigma_W(n+1/n) = \Sigma_W(n/n) + G(n)\Sigma_U(n/n)G(n)^T \quad (13)$$
$$\Sigma_u(n/n) = \mu\Sigma_w(n/n)$$

μ set to a small constant (for example $\mu \approx 1e-4$ for data length 1000) [28].

$$\Sigma_w(0/0) = \psi$$
$$E[V(n)V(n)^T] = R(n)$$
$$E[q(n)q(n)^T] = Q(n)$$

where $K(n+1)$ is Kalman gain and $\Sigma_W(n+1/n)$ is weight covariance matrix. In the above, the term $\Sigma_w(0/0) = \psi$ is initial covariance matrix of weight, which can select by arbitrary value.

IV. UKF PRICE FORECASTING

A. UKF Parameter Estimation

The EKF more than 30 years of experience in the estimation community has shown that is computationally expensive and difficult to implement. The UKF address the approximation issues of the EKF. The state distribution is again represented by a GRV, but is now specified using a minimal set of carefully chosen sample points [31]. No explicit calculations of Jacobians or Hessians are necessary to implement this algorithm. Furthermore, the overall number of computations is the same order as the EKF. This sample points completely capture the true mean and covariance of the GRV, and when propagated through the true uncertain dynamic non-linear system, captures the posterior mean and covariance accurately to the 3rd order (Taylor series expansion) for any non-linearity. Consider propagating a random vector variable $W(n)$ (neural network Weights in nth iteration) through a nonlinear function f[.], as (2). Assume $W(n)$ has mean $E[W(n)] = \overline{W}(n)$ and covariance $E[(W(n) - \hat{W}(n/n))(W(n) - \hat{W}(n/n))^T] = \sum_W(n/n)$. To calculate the statistics of $P(n)$, we form a matrix $\omega(n)$ of $2L_W + 1$ sigma vectors $\omega_i(n)$ (with corresponding weights W_i), according the following:

$$W_0^m = \lambda/(L_w + \lambda)$$
$$W_0^C = \lambda/(L_w + \lambda) + (1 - \alpha^2 + \beta) \quad (14)$$
$$W_i^m = W_i^c = 0.5/(L_w + \lambda) \qquad i = 1,\ldots\ldots,2L_W$$
$$\lambda = \alpha^2(L_w + \kappa) - L_w$$

$$\omega_0(n) = \overline{W}(n)$$
$$\omega_i(n) = \overline{W}(n) + \sqrt{(L_W + \lambda)\sum_W(n/n)_i} \qquad i = 1,\ldots..,L_w$$
$$\omega_i(n) = \overline{W}(n) - \sqrt{(L_W + \lambda)\sum_W(n/n)}_{i+L_W} \qquad i = L_w + 1,\ldots..,2L_w \quad (15)$$
$$\omega(n) = [\omega_0(n) \;\ldots\ldots\; \omega_{2L_W}(n)]$$

where $\Sigma_W(n/n)_i$ means ith column of the matrix $\Sigma_W(n/n)$. λ is a scaling parameter. α determines the spread of sigma points around $\overline{W}$, and is usually set to a small positive value. κ is a secondary scaling parameter which is usually set to 0, and β is used to corporate prior knowledge of the distribution of $W(n)$. These sigma vectors are propagated through the linear dynamic system,

$$\omega_i(n+1) = \omega_i(n) \qquad i = 0,\ldots..,2L_W \quad (16)$$

$$\hat{W}(n+1/n) = \sum_{i=0}^{2L_W} W_i^m \omega_i(n+1) \quad (17)$$

And covariance matrix of residue $\widetilde{W}(n+1/n) = W(n+1) - \hat{W}(n+1/n)$ is

$$\Sigma_w(n+1/n) = \sum_{i=0}^{2L_W} W_i^c[\omega_i(n+1) - \hat{W}(n+1/n)] \times [\omega_i(n+1) - \hat{W}(n+1/n)]^T + G(n)\Sigma_u(n/n)G(n)^T \quad (18)$$
$$\Sigma_u(n/n) = \mu\Sigma_w(n/n)$$

By propagating vector variable $\omega_i(n+1)$ from nonlinear function f[.] we have:

$$\rho_i(n+1) = f[\omega_i(n+1), L(n-l_j), P(n-p_j)] \quad (19)$$
$$i = 0,\ldots,2L_W$$

And the mean and covariance for $P(n)$ are approximated using a weighted sample mean and covariance of the posterior sigma points,

$$\hat{P}(n+1/n) = \sum_{i=0}^{2L_W} W_i^m \rho_i(n+1) \quad (20)$$

Using recursive estimation for $W(n+1)$ expressed in (4) the following form for measurement update equations obtained:

$$\Sigma_P(n+1) = \sum_{i=0}^{2Lw} W_i^c[\rho_i(n+1) - \hat{P}(n+1/n)][\rho_i(n+1) - \hat{P}(n+1/n)]^T + R(n) \quad (21)$$

$$\Sigma_Z(n+1) = \Sigma_P(n+1) + Q(n) \quad (22)$$
$$\hat{Z}(n+1/n) = \hat{P}(n+1/n)$$

$$\Sigma_{WZ}(n+1) = \sum_{i=0}^{2Lw} W_i^c[\omega_i(n+1) - \hat{W}(n+1/n)][\rho_i(n+1) - \hat{Z}(n+1/n)]^T \quad (23)$$

$$K(n+1) = \Sigma_{WZ}(n+1)\Sigma_Z(n+1)^{-1} \quad (24)$$

$$\Sigma_w(n+1/n+1) = \Sigma_W(n+1/n) - K(n+1)\Sigma_p(n+1)K(n+1)^T \quad (25)$$
$$\Sigma_w(0/0) = \psi$$

A. The Dual Unscented Kalman Filter (DUKF)

The dual estimation problem consists of estimating $P(n+1)$ from later information of price and load up to sample time n, and the model parameter $W(n+1)$ from noisy data $Z(n+1)$ simultaneously. For this propose, the following augmented matrix is suggested:

$$\begin{bmatrix} W(n+1) \\ P(n+1) \end{bmatrix} = \begin{bmatrix} W(n) \\ f[W(n+1), L(n-l_j), P(n-p_j)] \end{bmatrix} + \begin{bmatrix} G(n) & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} U(n) \\ q(n) \end{bmatrix} \quad (26)$$
$$Z(n+1) = [0 \quad I]\begin{bmatrix} W(n+1) \\ P(n+1) \end{bmatrix} + V(n)$$

In the above augmented method the state $P(n+1)$ and parameter vector $W(n+1)$ can be estimated simultaneously. So at every time-step, the current estimate of the weights is used in the state estimation procedure and the later state estimati

on is used for to improve the parameter estimation.
The number of input variables that drive the neural network subspaces has been selected by calculating the partial and standard autocorrelation functions suggested in [19].

B. The Decoupled Unscented Kalman Filter (DeUKF)

Using UKF as a learning method for MCP and load forecasting in view of the high dimensionality of weights involved, causing excessive computational requirements. Decoupled extended Kalman filter (DEKF) was developed in [10] and same methodology was used for UKF. In this methodology, the interdependency of the weights across neurons was ignored. So the weight covariance matrix is assumed to be block diagonal. In results section the DUKF method will be presented with significant reduction in computation and much faster than that UKF.

V. RESULTS

The data of New England electricity market in January, Februray and May 2004 were used as three study cases [24]. Different types of error can be used for validation of results. We have used the mean absolute error (MAE) defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - \hat{x}_i\right| = \frac{1}{N}\sum_{i=1}^{N}\left|e_i\right| \qquad (27)$$

Another type of error is the mean absolute percentage of error (MAPE) defined as:

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{\left|x_i - \hat{x}_i\right|}{x_i} \qquad (28)$$

The standard deviation (std) of error is a useful parameter and it is defined as

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[e_i - \bar{e}_i]^2} \qquad (29)$$

where $e_i = x_i - \hat{x}_i$ and $\bar{e}_i$ is the average of e_i values.
The std is a well-known criteria used widely in documents. The root mean square error (rmse) which is composed of variance and bias is more suitable for validation.

$$rmse = \sqrt{\sigma^2 + bias^2} \qquad (30)$$

A. Case Study I

In the first case study, the hourly price and load data corresponding to January 2004 was used. 1008 sample points (corresponding to January 1, Feb 11) were used. The 800 points of data are used as training data which selected randomly, and the final 208 points are used as the test set for validate the method. The price-price, price-load, load-price and load-load correlation coefficients for input determination were investigated (see [19] for detail).
The desired forecasting horizons for both price and load in this case study were considered seven hours. Based on the correlation results is showed in TABLE. I (suggested in [19]), the proper values for parameters were selected (for $l_j, j = 0$ and for $p_j, j = 0,1,2,3$ and 24, so the input number is equal to 6). In fact, the parameters were selected such that the correlation values of the corresponding quantities be larger than 0.5.
A standard 4-1 MLP with tansig for first layer and logsig for second layer activation function was used in case study I. The neural network thus had 29 total number of weights and was trained from Jan 1 to Feb 3, and then predicted from Feb 4 to 11. Several neural network structures were studied to select the network that has the best performance. The results showed that MLP is more suitable. It is clear that, the layer numbers and neurons in each layer affect the results of MLP. Therefore, same 4-1 MLP with similar activation functions was used for both EKF and DUKF. Figure 2 shows the predicted and real values of price for one step prediction. Also, Fig. 3 shows the predicted and real values of load for one step prediction.

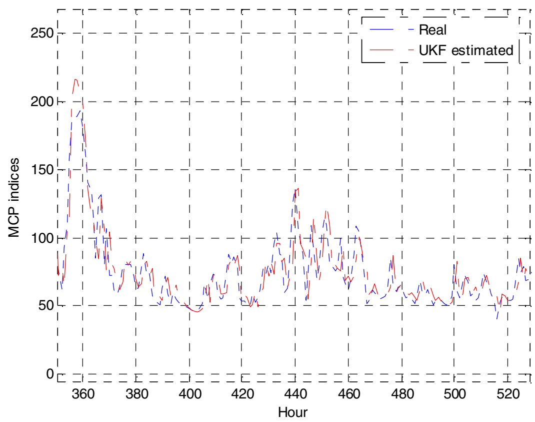

Fig. 2. Proposed method simulation results (DUKF), price forecasting (finer scale, one step prediction, case study I)

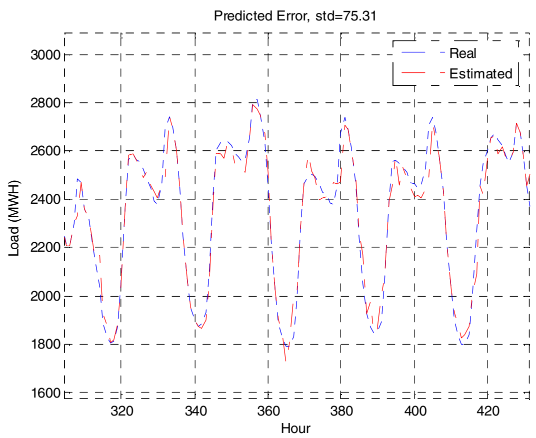

Fig. 3. Proposed method simulation results (DUKF), load forecasting (One step prediction, case study I)

TABLE I
THE CORRELATION COEFFICIENTS BETWEEN PRICE AND LOAD SAMPLES

	One step	*Two step*	*Three step*	*Four step*	*Five step*	*Twenty Four step*
Price(n+24), Load(n)-Load(n-7)	0.635	0.4596	0.430	0.306	0.199	0.09
Price(n+24), Price(n)-Price(n-7)	0.846	0.677	0.584	0.513	0.430	0.612

Fig. 4 shows comparison errors for one step MCP forecasting for three methods at test data (e.g. DUKF and EKF).

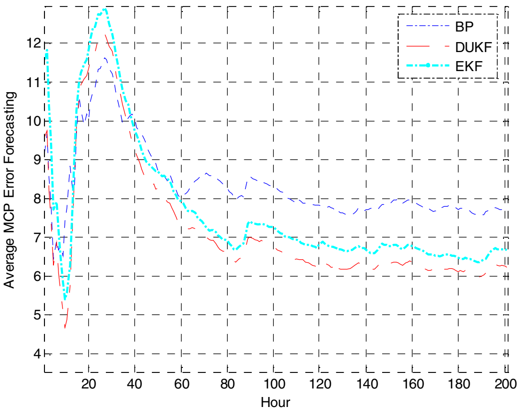

Fig. 4. DUKF, EKF and BP simulation error results, price forecasting (One step prediction, case study I)

The results STD, MAE, MAPE and RMSE values for one to seven steps predictions of price and load are shown in Table II. It can be seen that the smallest MAPE is 5.33% for one step price prediction and 3.01% for one step load prediction. This error increases slightly for higher step predictions. The MAPE values for seven steps price and load predictions are 13.03 and 6.82, respectively.

TABLE II
COMPARING ERROR VALUES BETWEEN PROPOSED and EKF METHODS, (case study I, One step ahead prediction)

Method	Price forecasting				Load forecasting			
	STD	RMSE	MAE	MAPE (%)	STD (Mwh)	RMSE (Mwh)	MAE (Mwh)	MAPE (%)
DUKF	14.9	14.9	5.8	1.9	43.5	44.2	34.3	1.0
EKF	16.5	16.6	7.6	2.5	55.4	59.1	58.6	1.8

B. Case study II

In the second case study, the hourly price and load data corresponding to February 1, 2004 to March 5, 2004 was used. 1008 sample points were used. The overall forecasting performance for these two methods and BP training method are summarized in Table III. The first 600 points of data are used as training data which selected randomly, and the final 240 points are used as the test set for validate the method. The proper values for parameters were selected from correlation coefficients test (for $l_j, j=0$ and for $p_j, j=0,1,2,3$ and $24, 24*7$, so the input number is equal to 7). Comparing the input numbers for this case study and case study I, it can be seen that the correlation coefficients values are higher than later case study (for detail see [19]).

A standard 4-1 MLP with tansig for first layer and logsig for second layer activation function was used in case study, too. This structure was used for four methodologies, similarly. Figure 5 shows comparison errors for seven step MCP forecasting for these four methods at test data (e.g. DUKF, UKF, EKF and BP).

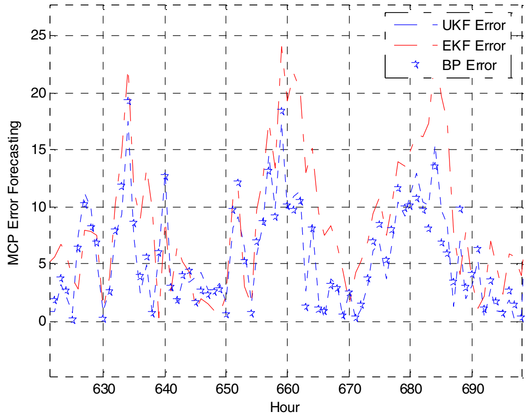

Fig. 5. DUKF and EKF simulation error results, price forecasting (Seven forward step prediction, case study II)

The resultant STD, MAE, MAPE and RMSE values for one, seven and next day predictions of price and load are shown in Table III, IV and V.

TABLE III
COMPARING ERROR VALUES BETWEEN PROPOSED, EKF AND BP METHODS, (case study II, One step forward prediction)

Method	Price forecasting				Load forecasting			
	STD	RMSE	MAE	MAPE (%)	STD (Mwh)	RMSE (Mwh)	MAE (Mwh)	MAPE (%)
DUKF	6.32	6.33	4.57	4.70	38.23	37.9	30.1	0.93
UKF	6.45	6.63	5.1	5.3	39.4	38.2	30.0	0.91
EKF	6.55	6.67	5.25	5.40	45.56	46.23	42.56	1.5
BP	6.43	6.45	4.89	5.03	37.86	38.1	29.5	0.92

TABLE IV
COMPARING ERROR VALUES BETWEEN PROPOSED, EKF AND BP METHODS, (case study II, Seven step forward prediction)

Method	Price forecasting				Load forecasting			
	STD	RMSE	MAE	MAPE (%)	STD (Mwh)	RMSE (Mwh)	MAE (Mwh)	MAPE (%)
DUKF	7.56	7.42	5.01	4.98	140.4	152.1	121.3	4.0
UKF	8.46	8.52	6.22	4.81	143.0	153.8	122.0	4.2
EKF	8.70	8.75	6.47	5.01	146.4	153.5	118.9	4.3
BP	7.43	7.23	4.89	5.03	216.5	266.9	211.0	7.5

TABLE V
COMPARING ERROR VALUES BETWEEN PROPOSED, EKF AND BP METHODS, (case study II, Next Day prediction)

Method	Price forecasting				Load forecasting			
	STD	RMSE	MAE	MAPE (%)	STD (Mwh)	RMSE (Mwh)	MAE (Mwh)	MAPE (%)
DUKF	9.82	10.24	7.90	6.11	186.6	207.1	162.8	5.7
UKF	9.89	10.30	7.93	6.14	188.6	210.1	164.8	5.8
EKF	9.7	10.62	8.28	6.4	195.2	209.8	164.6	5.8
BP	11.04	11.42	9.01	6.97	265.7	336.5	278.7	9.9

In Table VI, the price and load forecasting results obtained by the proposed method (DUKF) are compared to the results obtained by six other techniques for this case study. The above advantages are the average of 100 iterations. These are: MLP using BP, MLP using genetic algorithm, fuzzy-neural approach [7], generalized regression neural network (GRNN), CO-CO approach [19], EKF method [10] and the proposed method (UKF and DUKF). The same training and testing data sets were used for all methods. The resultant STD, MSE, MAE and MAPE for each case considering seven step are shown in this table.

It can be observed that the smallest error was about 4.81% (MAPE) and 7.42 (RMSE) for cases seven step price prediction using proposed method. And smallest error was about 4.01% (MAPE) and 140.4 (STD) for cases seven step load prediction.

The new method has shown to be able to improve the results in about 5-10% at confident interval 95% than the other algorithms. In the other word forecasting results show that this approach is more accurate than conventional techniques which introduced before.

TABLE VI
COMPARING ERROR VALUES BETWEEN DIFFERENT METHODS PERDICTION INTERVAL,
(FEB 25, 2004 to MARCH 5, 2004, ONE STEP PREDICTION)

	Method	***Specification***	*Price forecasting*				*Load forecasting*			
			STD	***RMSE***	***RMSE)***	***MAPE (%)***	***STD (Mwh)***	***RMSE (Mwh)***	***MAE (Mwh)***	***MAPE (%)***
1	***MLP-BP***	trianing steps = 1600	7.43	7.23	4.89	5.03	216.5	266.9	211.0	7.5
2	***MLP-GA***	eneration - number = 2000	7.33	6.92	4.81	4.96	208.23	261.9	201.0	7.1
3	***Fuzzy-Neural [6]***	Fuzzy Logic for Load Prediction GRNN for Price Prediction spread - number = 0.1	6.46	6.63	4.09	7.45	212.23	255.9	220.1	8.1
4	***GRNN***	spread - number = 0.1	7.53	7.92	5.21	5.16	215.23	222.9	211.0	7.2
5	***CO-CO [19]***	2 GRNN *subspaces for Load* & *Price* Spread - number = 0.1	6.35	6.38	4.6	5.60	208.23	261.9	201.0	7.1
6	***EKF [10]***	*trianing* steps = 1600	8.70	8.75	6.47	5.01	146.4	153.5	118.9	4.3
7	***UKF***	*trianing* steps = 1600 $\alpha = 0.005, \beta = 2, \kappa = 0$	8.46	8.52	6.22	4.81	143.0	153.8	122.0	4.2
8	***DUKF***	*trianing* steps = 1600 $\alpha = 0.005, \beta = 2, \kappa = 0$	7.56	7.42	5.01	4.98	140.4	152.1	121.3	4.0

VI. DISCUSSION AND CONCLUTION

Using EKF as a learning method for MCP prediction may suffers a number of serious limitations. At first the existent of Jacobian matrices is a critical problem, especially for some activation functions contain discontinuities (for example hard-limit), so Jacobian matrices might be abruptly. The second, linear transformations can introduce large errors in true posterior mean and covariance; therefore it causes its estimation to diverge. Finally calculating Jacobian matrices can be very difficult and introduces numerous errors. In this paper a new methodology for MCP and load forecasting was introduced to determine neural networks weighting coefficients based on unscented filtering. This approach is able to efficiently forecast the both MCP indices and load in comparison with the other methods. Our DUKF forecasting results are better and reliable than the prediction obtained by using the EKF method, and are comparable with BP methods. From the above simulation results and many other simulations (not reported here), the proposed DUKF method is shown to be more computationally efficient (especially in long term prediction) in price and load prediction than other methods.

REFERENCES

[1] S. Signal and L. Wu, "Training feedforward networks with the extended Kalman algorithm," in Proc. Int. Conf. ASSP, 1989, pp. 1187-1190.
[2] N. J. Francisco, C. Javier, and et al, "Forecasting next-day electricity prices by time series models." *IEEE Transactions on Power Systems*. vol. 17, no 2, pp. 342-348, May 2002.
[3] D. C. Sansom, and T. K. Saha, "Neural Networks for forecasting electricity pool price in a deregulated electricity supply industry." *AUPEC/EECON 99*, Darwin, Australia, 1999.
[4] D. H. Sonhuai, and Zhijian. "Design on spot price forecasting and mutation alarm system In electricity market." in *Proc. Power Systems and Communications Infrastructures for the Future*, Beijing, 2002.
[5] D. C. Park, M.A., El-Sharkawi, R. J. Marks II, L. E. Atlas and M.J. Damborg, "Electric Load Forecasting Using an Artificial Neural Network." *IEEE Transaction on Power Systems*, vol. 6, no. 2, May 1991, pp. 442-449.
[6] T. M. Peng, N. F. Hubele, and G. G. Karady, "Advancement in the application of Neural Networks for short-term load forecasting ," *IEEE Transactions on Power Systems*, vol. 7, pp. 250-257, 1992.

[7] L. Vanaja, C. Fung and T. Gedeon, “A fuzzy-neural approach to electricity load and spotprice forecasting in a deregulated electricity market,” School of Information Technology, Murdoch University, South Street, AUSTARIA
[8] J. J. Guo, and B. L. Peter, “Improving market clearing price prediction by using a committee machine of neural networks”, *IEEE Transactions on Power Systems.* vol. 19, no. 4, pp. 1867-1876, Nov 2004.
[9] P. Rodriguez , G. J. Anders, "Energy price forecasting in the ontario competitive power system market," *IEEE Trans. on Power Systems, vol. 19, no. 1, pp. 366-374, Feb 2004.*
[10] L. Zhang and B. Luh, ‘‘Neural network-based market clearing price prediction and confidence interval estimation with an improved extended Kalman filter method,’’ *IEEE Transactions on Power Systems.* vol. 20, no. 1, pp. 59-66, Feb 2005.
[11] W. A. Wright, ‘Baysian approach to neural-network with input uncertainty,” *IEEE Trans. Neural Networks,* vol. 10, no. 6, pp. 1261-1270, Nov, 1999.
[12] G. E. P. Box and G. M. Jenkins, *Time Series Analysis Forecasting and Control*, 2nd ed. San Francisco, CA :Holden-Day, 1976.
[13] K. Peng, and C. Wen, “An algorithm to determine Neural Network hidden layer size and weight coefficients,” i*n Proc. IEEE Symposition on Intelligent Control*, 2000.
[14] S. Haykin, *Neural Network: A Comprehensive Foundation,* prentice Hall, New Jersey, 1999.
[15] Y. Y. Hong, “Locational marginal price forecasting in deregulated electric markets using a recurrent Neural Network”, *In Proc. IEEE Power Engineering Society Winter Meeting 2001, vol. 2, pp.* 539 – 544.
[16] D. F. Specht, ”A general regression neural network.” *IEEE Transactions on NeuralNnetworks*, vol. 2, pp. 568-576, 1991.
[17] D. Whiteley, “Applying genetic algorithm to Neural Networks learning,” in *Proc. Conference of the socity of Artificial Intyelligence and Simulation of Behavior,* Sussex, England, 1986, pp. 320-331.
[18] M. R. Akbarzadeh, and Ms. Trabi., “Genetic searching in finding perceptron optimum weights and structure,” *in Proc. 11th Iranian Conference of Electrical Engineering,* 2003, vol. 1, pp. 234-230.
[19] A. Karsaz, H. R. Mashhadi, “Cooperative co-evolutionary approach to electricity load and price forecasting in deregulated electricity markets”*, in International Power India Conf. IEEE 2006.*
[20] X. Yao, “A review of evolutionary artificial Neural Networks,” *International J. Intelligent Systems, v*ol. 8, pp. 549-567, 1989.
[21] D. Beasly, R. Martin. “A sequential niche technique for multimodel function operation,” in conf. *Evolutionary Computation, 1993, vol.* 1. pp. 101-125.
[22] A. Berlanga, and A. Sanchis, “A general learning co-evolution method to generalize autonomous robot navigation behavior,” i*n Proc. IEEE* Evolutionary Computation 2000, vol. 1, pp. 769 – 776.
[23] J. Gonzalez, J. Ortega, H. Pomares,F. Fernandez, and A Diaz, “Multiobjective evolutionary opyimizition of size, shape and position parameters of radial basis function networks for function approximation,” *IEEE Transaction on Neural Network 2003,* vol. 14, no. 6, pp. 1478-1495.
[24] [http://www.iso-ne.com/smd/operations_reports/
[25] *S* . Haykin. *Aduptive Filter Theory*. Prentice-Hall, Inc. *3* edition, 1996.

BIOGRAPHIES

Ali Karsaz received the B.S. degree in Electrical Engineering from the Amirkabir University of Technology, Tehran, Iran, in 1999. He Obtained his M.Sc. degree in Electrical Engineering in 2004 at Ferdowsi University of Mashhad. He is currently as a Ph.D. student in Control Engineering at the Ferdowsi University of Mashhad, Mashhad, Iran. His interest area is Artificial Neural Networks, Stochastic and nonlinear Modeling and Estimation, System Identification, Time Series Analysis and prediction, Inertial Navigation Systems.

Hamid Khaloozadeh, received the B.S. degree in control engineering from Sharif University of technology (Tehran, Iran), in 1990, the M.S. degree in control engineering from K.N.Toosi university of technology Tehran, Iran), in 1993, and the Ph.D. degree in control engineering from Tarbiat Moddaress university (Tehran, Iran), in 1998. Since 1998 to 2004 he was a faculty member at Ferdowsi University of Mashhad. He is currently an assistant Professor teaches in the Department of Electrical Engineering in K.N.Toosi University of Technology (Tehran, Iran). His interest area is Digital Control, Nonlinear Modeling, System Identification, Optimal Control, Adaptive Control, Stochastic Estimation and Control, Neural Networks and Time Series Analysis.