# Electricity Price Prediction Model Based on Simultaneous Perturbation Stochastic Approximation

**Hee-Sang Ko*, Kwang Y. Lee** and Ho-Chan Kim†**

**Abstract** – The paper presents an intelligent time series model to predict uncertain electricity market price in the deregulated industry environment. Since the price of electricity in a deregulated market is very volatile, it is difficult to estimate an accurate market price using historically observed data. The parameter of an intelligent time series model is obtained based on the simultaneous perturbation stochastic approximation (SPSA). The SPSA is flexible to use in high dimensional systems. Since prediction models have their modeling error, an error compensator is developed as compensation. The SPSA based intelligent model is applied to predict the electricity market price in the Pennsylvania–New Jersey–Maryland (PJM) electricity market.

**Keywords:** electricity market, intelligent time series, price prediction, simultaneous perturbation stochastic approximation

## 1. Introduction

Many countries are restructuring their electrical power industry as a means of introducing deregulation and competition by unbundling generation, transmission and distribution functions, and allowing open market access. In the deregulated environment, a market-based system of electricity transactions has been introduced [1] where electricity is traded as a commodity and the balance of supply and demand significantly influences its price. The theory of electricity spot marketpricing states that the hourly spot price can be determined by such factors as fuel and maintenance costs, the availability of generator and network, and costs to compensate for transmission losses [2].

The electrical power industries have seen many changes over the last decade. Regulated or state-owned monopoly markets have been deregulated. This happened first in the United Kingdom and New Zealand, followed several years later by Sweden, Norway, Australia, New England, New York, California, and Pennsylvania–New Jersey–Maryland (PJM) [3, 4]. In Canada, market-based electricity pricing has come to Alberta and is coming to Ontario.

There is no reason why producers and consumers of electrical power cannot meet in a properly designed marketplace to decide on the price of their product. But electrical power is different from most other commodities. It cannot appreciably be stored and system stability requires constant balance between supply and demand. Most users of electricity are, on short time scales, unaware of or indifferent to its price. These two facts drive the extreme price volatility or price spikes of the electrical power market.

In the modeling of spot price of electricity, one common approach is to observe the price for a long period and fit a statistical model based on the observed time series [5]. The other approach is called Ryan and Mazumdar production costing model, used to represent the main variables that affect the spot price of electricity [6]. In that work, the information on the probability distribution of prices is of particular use in managing risk and improving decision-making. Moreover, the estimated price depends on many factors such as the periodicity of demand, temperature, and other meteorological influences, the loading order of generators, etc. Traditional methods based on statistical and probabilistic approaches may not be suitable to represent data generated by human activities such as the prices in the power exchange market. For example, market-based transactions of electricity tend to make prices more volatile in the high demand region than in the low demand region, depreciating the traditional assumption of random process [7].

In this paper, the uncertain market prices are represented by an intelligent time-series model. The parameter of the intelligent time-series model is obtained by the simultaneous perturbation stochastic approximation (SPSA) [8]. Most mathematical algorithms for search and optimization play a large role in finding the best solutions in many problems. They start with an initial guess, and

† Dept. of Electrical Engineering, Cheju National University, Korea. (hckim@cheju.ac.kr)
* Wind Energy Research Center, Korea Institute of Energy Research, Korea. (hee-sang@ieee.org)
** Dept. of Electrical and Computer Engineering, Baylor University, USA. (Kwang_Y_Lee@baylor.edu)
Received 8 June, 2007 ; Accepted 4 December, 2007

this predicted solution is updated on an iteration-by-iteration basis with the aim of improving the performance measure. The powerful merit of the SPSA is that it uses perturbation signals. Such signals actually help in finding the global optimum because a possibility to be trapped in a local minimum can be minimized. This is the main advantage of the SPSA.

Since prediction models always have a modeling error, this paper presents an error compensator to compensate for the modeling error. The proposed model is applied to the Pennsylvania–New Jersey–Maryland (PJM) hourly time-series electricity market data [4].

## 2. An Intelligent Time-Series Model Based on the Newton-Backward Operator

Consider a nonlinear time-invariant discrete-time system, represented by

$$y(k+1) = f(y(k), y(k-1), \cdots, y(k-N)) \tag{1}$$

where $y(k\text{-}i),\ i = 0,1,\cdots,N$ denote the delayed data. It can be shown that the delayed signals are made of increments or differences. The Newton backward difference operator [9] is defined as

$$\begin{aligned} &\Delta^n f(k) = \Delta^{n-1} f(k) - \Delta^{n-1} f(k-1), \quad n \ge 1 \\ &\Delta^0 f(k) = f(k) \end{aligned} \tag{2}$$

Using the difference operator (2), the model (1) can be represented as

$$y(k+1) = f(y(k), \Delta y(k), \cdots, \Delta^N y(k)) \tag{3}$$

Equation (3) is expanded into Taylor series in [9].

$$\begin{aligned} y(k+1) &= f(y(k), \Delta y(k), \cdots, \Delta^N y(k)) \\ &= f(y(k-1), \cdots, \Delta^N y(k-1)) \\ &\quad + \left( \frac{\partial f}{\partial y(k-1)} \right) (y(k) - y(k-1)) + \cdots \\ &\quad + \left( \frac{\partial f}{\partial \Delta y^N (k-1)} \right) (\Delta^N y(k) - \Delta^N y(k-1)) \\ &= y(k) + \sum_{i=1}^{N} a_i \Delta^i y(k) + O(k) \end{aligned} \tag{4}$$

where $a_i = \partial f / \partial \Delta^i y(k-1)$ and $O(k)$ represents the high order terms. By subtracting $y(k)$, (4) can be represented as follows:

$$\Delta y(k+1) = \sum_{i=1}^{N} a_i \Delta^i y(k) + O(k) \tag{5}$$

By neglecting high order terms, the free model is defined as the following:

$$\Delta \hat{y}(k+1) = \sum_{i=1}^{N} a_i \Delta^i y(k) \tag{6}$$

or, dividing both sides with $\Delta$,

$$\hat{y}(k+1) = \sum_{i=1}^{N} a_i \Delta^{i-1} y(k) \tag{7}$$

where $N$ is the order of the proposed intelligent time-series prediction model, and $\hat{y}(k+1)$ denotes the estimate of $y(k+1)$. The remaining problem now is how to determine parameters $a_i$.

## 3. The Description of the Error Compensator

For better accuracy, it is required to consider a wide range of fluctuating conditions. In real application, however, it is impractical to consider all conditions. Therefore, when the prices are fluctuating, errors between the predicted value and real value inevitably exist even though the intelligent time-series model may have been completely trained for a previously given data set. Thus, there exists a modeling error $E(k)$ at time $k$. The predicted value considering the modeling error becomes as follows:

$$y^*(k+1) = \hat{y}(k+1) + \hat{E}(k+1) \tag{8}$$

where $\hat{E}(k+1)$ is the estimate of the modeling error $E(k+1)$, $\hat{y}(k+1)$ is the predictor output in (7), and $y^*(k+1)$ is the corrupted estimate.

The error can be estimated by extrapolating the previous error using the Newton-backward-difference formula (NBDF) [10] as follows:

$$\hat{E}(k+1) = \sum_{r=1}^{l} (-1)^r \binom{r}{l+1} \Delta^r E(k) \tag{9}$$

where $E(k) = y(k) - \hat{y}(k)$ is the modeling error at time $k$, $\Delta^r$ is the backward difference operator defined in (2), $l$ is the extrapolation order, and the binomial-coefficient notation is defined as

$$\binom{r}{l+1} = \frac{r(r-1)\cdots(r-l)}{(l+1)!} \quad (10)$$

## 4. Simultaneous Perturbation Stochastic Approximation

To determine parameters, the SPSA method is applied [8] to the least squares problem, which is defined to minimize the loss function $E(\theta)$:

$$\text{Min } E(\theta) = \sum_{i=1}^{n} (y(k-i+1) - \hat{y}(k-i+1))^2 \quad (11)$$

where $\theta = [a_1 \cdots a_N]^T$ is the parameter vector of the prediction model, and $y$ and $\hat{y}$ indicate the real data and estimated data of the proposed prediction model, respectively.

The stochastic optimization is of great practical importance, which may be stated as the problem of finding a minimum point, $\theta^* \in R^p$, of a real-valued function $L(\theta)$ or the loss function that is observed in the presence of noise. A common desire of optimization problems in many applications is that the algorithm reaches the global minimum rather than becoming stranded at a local minimum value. The SPSA uses only the measurements of objective function. This contrasts with other algorithms requiring direct measurements of the gradient of the objective function, which are often difficult or impossible to obtain. Furthermore, the SPSA is especially efficient in high-dimensional problems in terms of providing a good solution with a relatively small number of measurements of the objective function. The essential feature of the SPSA, which provides its power and relative ease of use in difficult multivariate optimization problems, is the underlying gradient approximation that requires only two measurements of objective function per iteration, regardless of the dimension of the optimization problem. These two measurements are made by simultaneously varying in a proper random fashion all of the variables in the problem. This contrasts with the classical finite-difference method where the variables are varied one at a time. If the number of variables being optimized is $p$, then the finite-difference method takes $2p$ measurements of the objective function at each iteration to form one gradient approximation while the SPSA takes only two measurements.

A fundamental result on relative efficiency is as follows: Under reasonably general conditions, the SPSA achieves the same level of statistical accuracy for a given number of iterations even though SPSA uses $p$ times fewer measurements of the objective function at each iteration. This indicates that the SPSA will converge to the optimal solution within a given level of accuracy with $p$ times fewer measurements of the objective function than the standard gradient method. Furthermore, the SPSA formally accommodates noisy measurements of the objective function. This is an important practical concern in a wide variety of problems involving Monte Carlo simulations, physical experiments, feedback systems, or incomplete knowledge such as fluctuating electricity market price.

The basic unconstrained SPSA algorithm is in the general recursive stochastic approximation (SA) form

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \eta_k \hat{g}_k(\hat{\theta}_k) \quad (12)$$

where $\hat{g}_k(\hat{\theta}_k)$ is the simultaneous perturbation estimate of the gradient $g(\theta) \equiv \partial L(\theta)/\partial\theta$ at the iterate $\hat{\theta}_k$ based on the measurements of the loss function and $\eta_k$ is a nonnegative scalar gain coefficient.

The essential part of (12) is the gradient approximation $\hat{g}_k(\hat{\theta}_k)$. This gradient approximation is formed by perturbing the components of $\hat{\theta}_k$ one at a time and collecting a loss measurement $E(\bullet)$ at each of the perturbations. This requires $2p$ loss measurements for a two-sided finite difference approximation. All elements of $\hat{\theta}_k$ are randomly perturbed together to obtain two loss measurements $E(\bullet)$. For the two-sided simultaneous perturbation gradient approximation, this leads to

$$\hat{g}_k(\hat{\theta}_k) = \frac{E(\hat{\theta}_k + c_k\Delta_k) - E(\hat{\theta}_k - c_k\Delta_k)}{2c_k} \begin{bmatrix} \Delta_{k1}^{-1} \\ \Delta_{k2}^{-1} \\ \vdots \\ \Delta_{kp}^{-1} \end{bmatrix} \quad (13)$$

where the mean-zero $p$-dimensional random perturbation vector, $\Delta_k = [\Delta_{k1}, \Delta_{k2}, \cdots, \Delta_{kp}]^T$, has a user-specified distribution and $c_k$ is a positive scalar. Because the numerator is the same in all $p$ components of $\hat{g}_k(\hat{\theta}_k)$, the number of loss measurements needed to estimate the gradient in SPSA is two, regardless of the dimension $p$.

The procedure how the SPSA iteratively produces a sequence of estimates is summarized as below:

**Step 1: Initialization and coefficient selection**

Pick initial guess $\theta_0$ of $g(\theta) \equiv \left.\frac{\partial L(\theta)}{\partial \theta}\right|_{\theta=\theta_0}$ in (12) and nonnegative coefficients $\eta, c, A, \alpha$, and $\gamma$ in the SPSA gain sequences $\eta_k = \eta/(A+k+1)^{\alpha}$ and $c_k = c/(k+1)^{\gamma}$.

**Step 2: Generation of simultaneous perturbation vector**

Generate a $p$-dimensional random perturbation vector $\Delta_k$ from a zero-mean probability distribution.

**Step 3: Loss function evaluations**

From Steps 1 and 2, obtain two measurements of the loss function based on the simultaneous perturbation around the current $\hat{\theta}_k$: $E(\hat{\theta}_k + c_k\Delta_k)$ and $E(\hat{\theta}_k - c_k\Delta_k)$ in (13) with the $c_k$ and $\Delta_k$.

**Step 4: Gradient approximations**

Generate the simultaneous perturbation approximation to the unknown gradient $\hat{g}_k(\hat{\theta}_k)$ according to (13).

**Step 5: Updating $\hat{\theta}_k$ estimate**

Use the standard stochastic approximation form in (12) to update $\hat{\theta}_k$ to a new value $\hat{\theta}_{k+1}$.

**Step 6: Iteration or Termination**

Return to Step 2 with $k+1$ replacing $k$. Terminate the algorithm if the maximum allowable number of iterations has been reached.

## 5. Numerical Example

The proposed intelligent time-series model is applied to the Pennsylvania–New Jersey–Maryland (PJM) [4] system for demonstration. In this example, the hourly time-series data is used. The time-series data is first normalized by zero-mean unit variance as follows:

$$y' = x - I_{n\times1}\bar{x}, \qquad x_{scaled} = \frac{y'}{I_{n\times1} std(y')} \tag{14}$$

where $x$ is the time-series data, $I$ is the identity matrix, $n$ is the total number of data, $\bar{x}$ is the mean value of $x$, and *std* stands for the standard deviation. A 2000 time-series data is used. For training, 500 time series data is used, and 1500 time series data is used for the validation. Numerical values for the SPSA and the model are $[a_1\ a_2\ a_3\ a_4\ a_5] = [0.95\ \text{-}0.0593\ 0.1087\ \text{-}0.1313\ 0.0539]$, $\alpha$=0.5, $\gamma$=0.1, $c$=0.01

The prediction result shows very good accuracy, and all price spikes are correctly predicated. Fig. 1 and Fig. 2 indicate the results of training and validation, respectively. The order of the intelligent time-series model and the error compensator is chosen by 5 ($N = l = 5$) in (7) and (9). The reason to choose $N = 5$ is to consider the extension of the prediction for weekly forecasting. The mean square error (MSE) according to the iteration is given in Table 1.

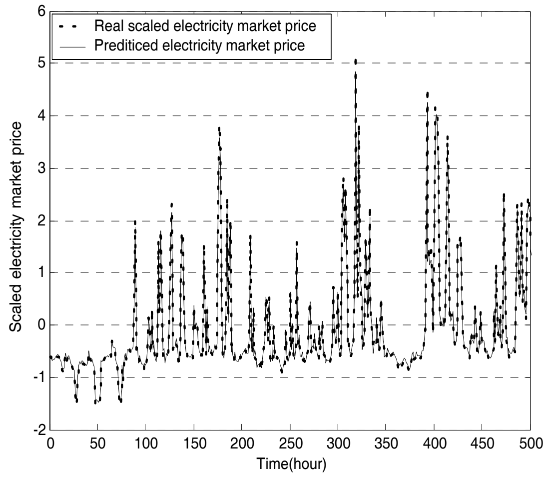


**Fig. 1.** The comparison of the scaled electricity market price between the actual values and the predicted values.

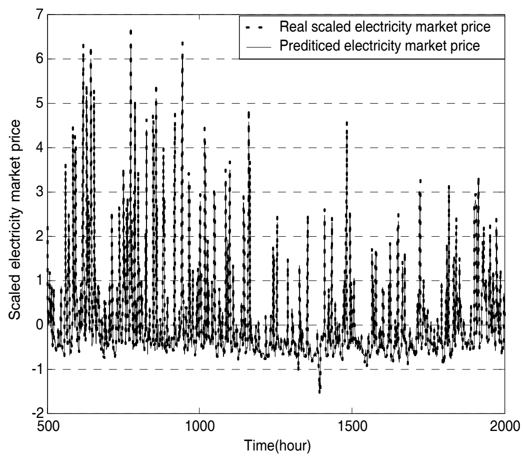


**Fig. 2.** The comparison of the scaled electricity market price between the actual values and the predicted values.

**Table 1.** Mean Square Error

| Iteration | Mean Square Error | | |
|---|---|---|---|
| | 500 (training) | 1500 (validation) | |
| | | Without Error Compensator | With Error Compensator |
| 1000 | 0.0199 | 0.0266 | 0.0239 |
| 3000 | 0.0079 | 0.0101 | 0.0089 |
| 5000 | 0.0032 | 0.0038 | 0.0038 |
| 7000 | 0.0023 | 0.0027 | 0.0025 |
| 10000 | 0.0021 | 0.0024 | 0.00237 |

## 6. Conclusion

This paper presented an intelligent time series model. The parameters of the intelligent time-series model are obtained based on the simultaneous perturbation stochastic approximation. There are three main advantages in the SPSA: free from high-dimensional problems, possible maximization of global optimization, and fast speed for making prediction. The prediction is accurate, and all price spikes are correctly predicted. Therefore, the proposed model can be an effective tool for an electricity supplier or broker to determine a price to offer a contract, which is attractive to a target customer. For future work, the next day electricity market price and long term prediction including monthly and yearly forecast will be conducted.

## References

[1] M. Ilic, F. Galiana, and L. Fink (ed.), *Power Systems Restructuring*, Kluwer, Boston, 1994.

[2] F. C. Schweppe, M. C. Caramanis, R. D. Tabors, and R. E. Bohn, *Spot Pricing of Electricity*, Kluwer Academic Publishers, 1998.

[3] R. Green and D. M. Newbery, "Competition in the British electricity spot market," *J. Political Econ*., vol. 100, no. 5, pp. 929-953, 1992.

[4] PJM Interconnection, LLC, http:// www.pjm.com, 2000.

[5] E. H. Allen and M. D. Ilic, *Price-Based Commitment Decisions in the Electricity Market, London*: Springer-Verlag, 1999.

[6] S. M. Ryan, and M. Mazumdar, "Chronological influence on the variance of electric power production costs," *Operations Research*, vol. 40, pp. 284-292, 1992.

[7] M. Michalik, E. Khan, and W. Mielczarski, "Statistical analysis of electricity demand and spot prices in competitive markets," *in Proceedings of the 4th International Conference on Advances in Power System Control, Operation and Management*, Hong Kong, pp. 457-462, 1997.

[8] J. C. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Wiely, 2003.

[9] H. S. Ko, K. Y. Lee, and H.-C. Kim, "An Intelligent based LQR controller design to power system stabilization," *Electrical Power System Research*, vol. 71, no. 1, pp. 1-9, 2004.

[10] K. Y., Lee and H. S. Ko, "Power system stabilization using free-model based inverse dynamic neuro controller," *in Proceeding of Int. Joint Conf. Neural Network*, Honolulu, USA, no. 3, pp. 2132–2137, 2002.

**Hee-Sang Ko**
He received his B.S. degree in Electrical Engineering from Cheju National University, Jeju, Korea, in 1996, his M.Sc. degree in Electrical Engineering from Pennsylvania State University, University Park, USA, in 2000, and his Ph.D. in Electrical and Computer Engineering from the University of British Columbia, Vancouver, Canada, in 2006. He is a researcher in the wind energy research center at Korea Institute of Energy Research (KIER), Korea. His research interests include wind power generation, power systems voltage and transient stability, data processing for power systems security analysis, electricity market analysis, and system identification.

**Kwnag Y. Lee**
He received his B.S. degree in Electrical Engineering from Seoul National University, Korea, in 1964, M.S. degree in Electrical Engineering from North Dakota State University, Fargo, in 1968, and Ph.D. degree in System Science from Michigan State University, East Lansing, in 1971. He has been with Michigan State, Oregon State, Univ. of Houston, the Pennsylvania State University, and Baylor University where he is currently a Professor and Chair of Electrical and Computer Engineering. His interests include power system control, operation, planning, and intelligent system applications to power systems. Dr. Lee is a Fellow of IEEE, Associate Editor of IEEE Transactions on Neural Networks, and Editor of IEEE Transactions on Energy Conversion. He is also a registered Professional Engineer.

**Ho-Chan Kim**
He received He received B.S., M.S., and Ph.D. degrees in Control & Instrumentation Engineering from Seoul National University in 1987, 1989, and 1994, respectively. He was a research staff from 1994 and 1995 at the Korea Institute of Science and Technology (KIST). Since 1995, he has been with the Department of Electrical Engineering at Cheju National University, where he is currently a professor. His research interests include wind power control, electricity market analysis, and grounding systems.