MSRDG International Journal of Computer Scientific Technology & Electronics Engineering Volume 1 Issue 1, 1-6, Jan-Feb 2025
2025 Multi-Sense Research Developing Group

Electricity Forecasting Using Pattern Matching Technique

B. Senthilnayaki[1], K. Venkatalakshmi[2]
[1]Department of IT, University College of Engineering, Villupuram,Tamil Nadu, India
[2]Department of ECE, University College of Engineering, Tindivanam,Tamil Nadu, India
[1]senthilnayaki@gmail.com

Received: 05.01.2025 Revised:05.02.2025 Accepted: 15.02.2025 Published:28.02.2025

Abstract - *Introduces a new method for time series forecasting based on a pattern matching technique. The procedure involves two steps. Initially, clustering methods are used to partition and tag the dataset, making it easier to represent structured data. Subsequently, a pattern matching forecasting algorithm is utilized to find similar patterns in past data. Predictions are only made where corresponding patterns exist, using labels rather than actual values for increased accuracy. The method has the goal of optimizing forecasting performance through the use of past pattern identification.*

Keywords - *pattern sequence, prediction, time series, indexing, clustering*

1. Introduction

Data mining is the extraction of valuable knowledge from large data sets. It is automatically scanning huge volumes of data to discover patterns, trends, and relationships beyond mere statistical analysis. Data mining utilizes complex mathematical algorithms to segment data and forecast future occurrences. The major aim of data mining is to convert raw data into an intelligible format for future decision-making. The fundamental characteristics of data mining are: Automatic Pattern Detection: Extracting concealed patterns without human intervention. Outcome Prediction: Forecasts based on past occurrences. Actionable Information: Obtaining useful information to make decisions. Large Dataset Orientation: Processing enormous levels of structured as well as unstructured data.

Data mining models are constructed with algorithms which operate on datasets. These models can be used with the original dataset as well as new datasets, a term referred to as scoring. Predictive data mining creates rules which assist in making inferences about certain conditions from provided data points. Data Sources and Integration, Organizations collect enormous volumes of data in diverse forms, such as: Operational Data: Sales, costs, inventory, payroll, and accounting records. Non-Operational Data: Industry sales predictions and macroeconomic markers. Metadata: Information on information, like database schema and dictionary definitions. Through the study of patterns and relationships in these data sets, companies can transform raw data into meaningful insights. For instance, supermarket sales information can be studied to learn about consumer buying behavior and assess the success of promotion campaigns. Components of Data Mining, Data mining has three major components: Data Integration: Integrating data from multiple sources into one view, usually in a data warehouse. Data Mining Process: Obtaining useful patterns from the integrated data set. Information Organization & Presentation: Organizing insights in a manner to help understand and make decisions.

Forecasting and Time Series Analysis, Forecasting is forecasting future values using past data. It is divided into three categories: Short-term forecasting: Forecasts within a short time horizon (e.g., days or weeks). Medium-term forecasting: Estimates for one or two years. Long-term forecasting: Estimates for several years ahead. Temporal databases hold relational data with time-related information, including: Sequence Databases: Purchase histories of customers, web clickstreams, and biological sequences. Time-Series Databases: Stock prices, inventory levels, and weather readings.

Data mining methods assist in analyzing these datasets to find trends, periodic patterns, and cyclical behavior. For instance, mining bank transaction data can maximize manpower scheduling according to traffic patterns among customers. Likewise, stock market data analysis can identify investment opportunities. Clustering in Data Mining, Clustering is an unsupervised classification method that clumps similar data points together into different classes. Clustering aims to achieve the highest intra-group similarity with inter-group distinction. One popular clustering algorithm is the k-means algorithm, which divides data into k groups. Determining the ideal number of clusters, however, calls for using certain indices. For example, clustering can divide consumers into income, driving record, and buying behaviors categories, which can facilitate targeted marketing. The ability of clustering enhances with increasing data volume, making it a valuable component in contemporary data mining practices.

Pattern Matching Forecasting (PMF), A new data mining method, Pattern Matching Forecasting (PMF), employs the nearest neighbor approach to forecast future values from historical patterns with similarities. PMF is especially useful for time-series forecasting, where there are several labels identifying predictive accuracy. PMF is a general-purpose forecasting algorithm that guarantees stability by observing inherent patterns in a dataset. PMF finds its use in finance, economics, production planning, and telecommunications.

2. Related work

A hybrid method for electricity price forecasting based on Wavelet Transform (WT) and Multiple Linear Regression (MLR). The method is of specific interest for the purpose of risk management in deregulated electricity markets with high price volatilities. Hybrid Model: The method combines WT and MLR by employing various models for various price segments. Implementation: The model was applied to Australia's National Electricity Market (NEM) for day-ahead prediction. Comparison: The new method performed better than an analytical model, a basic MLR model, and an Artificial Neural Network (ANN) model. Wavelet Selection: The Daubechies wavelet of order two was the most effective in enhancing accuracy [1]. A neuro-fuzzy methodology for electricity demand forecasting and compares it with the artificial neural networks (ANNs) and Box–Jenkins ARIMA model. Hybrid Intelligence: The research compares the performance of the emerging fuzzy neural network (EFuNN), which combines fuzzy logic with neural networks. Comparison with Other Models: The EFuNN model performed better than conventional ARIMA, ANN (trained on CGA and BP algorithms), and the Victorian Power Exchange (VPX) predictions. Dataset and Evaluation: The models were trained on 10 months of 30-minute interval electricity demand data from Victoria, Australia. Performance: The neuro-fuzzy system showed the highest accuracy in comparison with the methods tested. Relevance to Energy Markets: Adaptive Forecasting: Application of an adaptive system such as EFuNN indicates that neuro-fuzzy models are apt for dealing with non-linearity and dynamic shifts in energy demand. Comparison with Conventional Models: The higher performance compared to ARIMA and ANN suggests that hybrid models are capable of providing improved predictions for short-term load forecasting. Real-World Application: The research is applicable to power grid operators, enabling generation and distribution optimization based on more accurate demand predictions [2]. A detailed review and analysis of electricity price forecasting methods in deregulated markets. Extensive Methodological Review: The article classifies price forecasting methods into: Stochastic Time Series Models – Statistical methods such as ARIMA and GARCH. Causal Models – Techniques that employ fundamental market variables. Artificial Intelligence (AI)-Based Models – Methods such as neural networks, fuzzy logic, and hybrid methods.Quantitative Analysis: The research compares forecasting models on the basis of: Prediction Time Horizon, Input/Output Variables, Preprocessing Techniques, Model Architecture, Data Points Used for Evaluation. Comparison Tables: The article summarizes results in table format for ease of reference and comparison. Classification of Price-Influencing Factors: Classifies major factors utilized by researchers and their effect on forecasting accuracy. Application to Different Electricity Markets: Examines the performance of different models across various electricity markets globally. Significance in Energy Forecasting: Benchmark for Researchers: Based on this review, researchers can choose the most appropriate model for a given set of market conditions. Practical Insights for Market Participants: Policymakers and traders can utilize this information to enhance risk management and bidding. Model Selection Criteria: Offers a systematic evaluation framework to evaluate forecasting methods for a given application [3]. Proposes a neuro-fuzzy network-based day-ahead electricity price forecasting, using sophisticated machine learning methods to address the nonlinear, time-variant, and complex nature of electricity prices. Hybrid Forecasting Method: Mutual Information (MI)-based Feature Selection: A new MI method is employed to choose the most appropriate input variables. Cascaded Neuro-Evolutionary Algorithm (CNEA): Each module of the forecasting model is composed of a neural network (NN) trained with an evolutionary algorithm (EA). The forecasters are organized in a cascaded format to iteratively refine the predictions. Iterative Parameter Tuning: An optimization process adjusts both the MI-based feature selection and CNEA parameters. Performance Evaluation: Experimenting with real electricity market data from PJM (USA) and Spanish electricity markets. Compared with recent price forecasting models, which have better accuracy and robustness. Significance in Electricity Markets:Improved Accuracy: Cascaded structure provides more reliable predictions than conventional models. Robustness to Market Variability: Good at coping with non-stationary price signals in competitive energy markets. Practical Application: Applicable to market players requiring accurate price forecasts for risk management and bidding purposes [4]. This series of research articles presents several short-term electricity price forecasting methods employing neural networks, wavelet transforms, ARIMA, and combined models. Short-Term Forecasting with Neural Networks [5]. Trained with a three-layer feedforward neural network (FNN) employing the Levenberg-Marquardt algorithm to forecast next-week electricity prices. Employed in mainland Spain and California electricity markets. Critical for market players to improve bidding strategies and profit maximization. Wavelet Transform & ARIMA for Day-Ahead Forecasting [6] Wavelet transform is utilized to divide historical data into successive time series. ARIMA is then utilized on these segments for forecasting. Efficient for dealing with time-series trends and seasonality. Transformation Function Model (Nogales) Designed to forecast previous electricity prices using transformation functions. Utilizes median values as a measure for improved forecasting accuracy.

Hybrid ANN-Fuzzy Logic Model[7] Uses Artificial Neural Networks (ANNs) and Fuzzy Logic together. Feedforward ANN with three layers (input, hidden, output). Hidden layer nodes carry out fuzzification, enhancing prediction precision for nonlinear, imprecise price trends. Neural Networks (FNN, ANN): Good for pattern recognition but can be prone to overfitting. Wavelet-ARIMA: Good at picking up trends and seasonality but less responsive to abrupt market shifts. Transformation Function Model: Statistical model, helpful in dealing with missing data but not necessarily predictive of future trends. Hybrid ANN-Fuzzy Model: Ideal for uncertain, complicated market conditions since it integrates the learning ability (ANN) and reasoning (fuzzy logic).

This collection of studies offers different statistical, machine learning, and hybrid methods for electricity price forecasting. The following is an overview of major contributions and their relative merits. Major Forecasting Models & Methods:

Weighted Nearest Neighbor (WNN) Method [8] Compares a test tuple with training tuples to enhance prediction accuracy. Applied to next-day electricity price forecasting. Strength: Suitable for pattern recognition and dealing with sparse data.

Bayesian Model [9] Depending on probability distribution to forecast price movements. Prediction is done on the basis of class occurrence probability.Strength: Better than deterministic models in coping with uncertainty.

Data Mining Framework (Support Vector Machine - SVM) [10] Employs SVM to identify price spikes (above-average values). Enhances prediction accuracy by learning from historical price behavior. Strength: Applicable for nonlinear data patterns and outliers.

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) Model [11] Statistical model capturing electricity price volatility clustering. Ideal for next-day prediction in extremely volatile markets. Strength: Good for markets with lots of frequent movements.

Mixed Model (Model 24 + Model 48)[12] Combination of two models in a hybrid approach: Model 24: Forecasts 24-hour-ahead prices. Model 48: Forecasts 48-hour-ahead prices. Strength: Enhances short-run forecast accuracy. Relevance in Energy Markets: Machine Learning Models (SVM, WNN) are suitable for nonlinear and unstable price trends. Bayesian & GARCH Models best describe uncertain and variable price patterns. Hybrid Models (Mixed, ANN-Fuzzy, Wavelet-ARIMA) enhance accuracy by combining several approaches.

3. System architecture

Phase 1: Clustering (Preprocessing Step), K-Means Clustering is applied to cluster similar data points. The user needs to specify the number of clusters (k), which can be difficult. To find the best k, validity indices are utilized, making sure that the clustering structure is significant.

Phase 2: Forecasting, The data is then utilized for electricity price forecasting. By grouping similar price patterns first, forecasting models can make more accurate predictions based on the historical behavior of each cluster. Importance of Clustering in Forecasting, Lessens Complexity of Data: Clustering makes it possible to slice the electricity market data, which simplifies the ability of the forecasting model to identify patterns. Enhances Accuracy of Prediction: Rather than training on the entire data, the forecasting model learns from more homogenous subgroups, which eliminates errors.Handles Market Variability: Individual price regimes (e.g., peak hours versus off-peak hours) may be modeled in isolation to enhance robustness.

Selecting an Appropriate Number of Clusters (k): Poor selection will result in overfitting (excessive number of clusters) or loss of information (low number of clusters). Validity metrics such as Silhouette Score, Davies-Bouldin Index, and Dunn Index assist in establishing the optimum k-value as shown Figure 1.

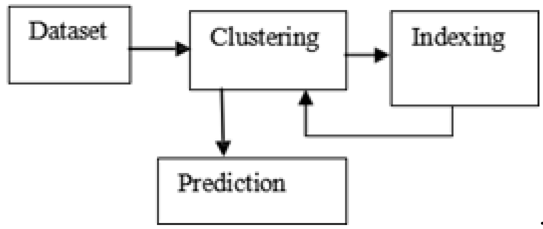

Fig. 1 System Architecture

The PMF algorithm is specifically for time series forecasting, combining the use of clustering methods and recursive forecasting. Most Important Elements of the PMF Method: Clustering Phase, K-Means Clustering is used to cluster similar patterns from the dataset. The algorithm numbers and labels the clusters from each sample, excluding other features. Indexing Phase (Identifying the Ideal Number of Clusters). Uses Validity Indices to identify the ideal clustering setup: Silhouette Index – Calculates how well each sample fits into a cluster. Dunn Index – Maximizes inter-cluster distance and minimizes intra-cluster variance. Davies-Bouldin Index – Assesses compactness and separation of the clusters. Pattern Matching Forecasting (PMF)- Sequence Matching: Searches past data for matching sequences of labels. Calculates average values of matched sequences for forecasting. Recursive Forecasting (Closed-Loop Method): Each forecasted sample is used to re-enter the dataset. Permits multi-step (long-term) forecasting more than one day ahead. Window Size Selection: Cross-validation selects the best window size. The window with the lowest prediction error is selected.

4. Discussion

Dataset Selection: The household electricity power dataset was employed for verifying the forecasting methodology. This is a widely employed dataset in studies of forecasting on account of having rich historical patterns and published baselines. Performance Evaluation Using Clustering Validity Indices. The following clustering assessment measures were utilized to confirm K-Means clustering performance: Silhouette Index (Figure 2) Tells us the extent to which data points conform to their specified cluster. Computed by averaging intra-cluster nearest inter-cluster difference. Higher values represent improved clustering quality. Davies-Bouldin Index (Figure 3) Picks out clusters with high inter-cluster and low intra-cluster similarity. Lower DB index represents improved clustering, as the clusters are well-separated. Dunn Index (Figure 4) Computes the ratio of the smallest inter-cluster distance to the largest intra-cluster distance. Higher Dunn index represents better-separated and compact clusters. Comparison with Existing Work: The forecasting results of the proposed method were compared with existing forecasting models. Through clustering-based PMF (Pattern Matching Forecasting), the approach showed enhanced accuracy in predicting household electricity consumption.

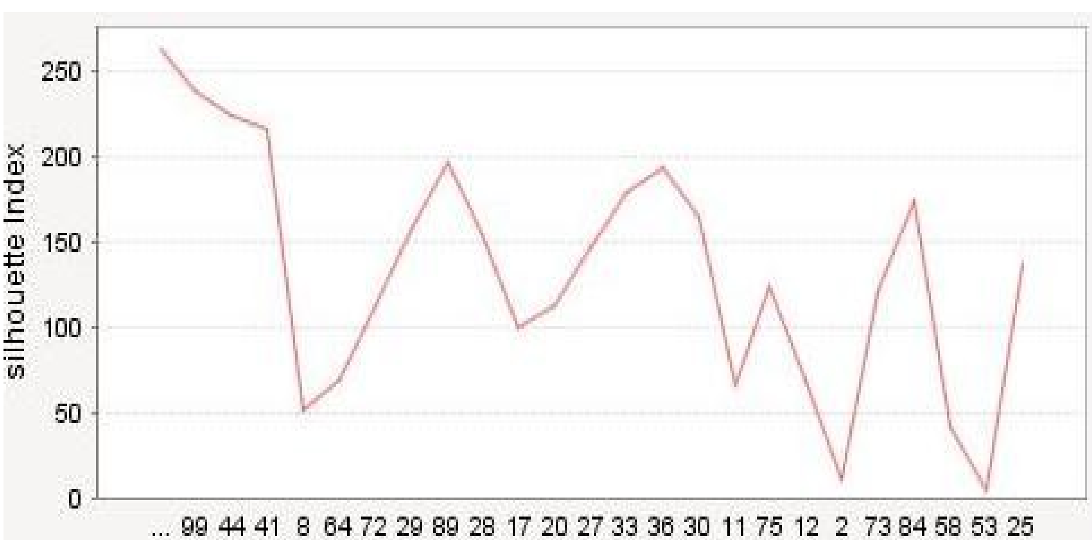

Fig. 2 Silhouette Index

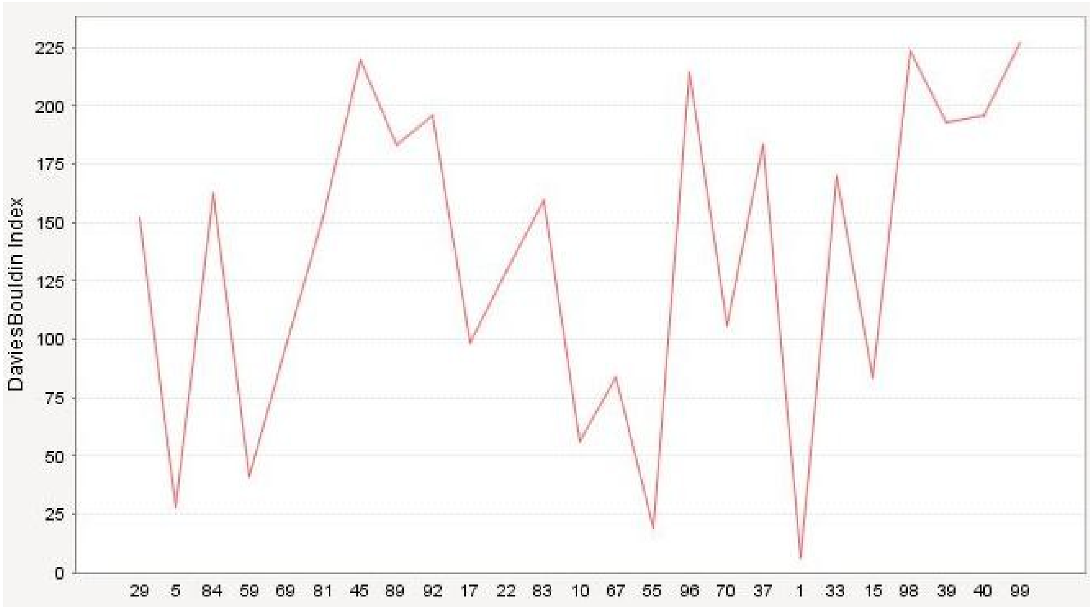

Fig. 3 Davies-Bouldin Index

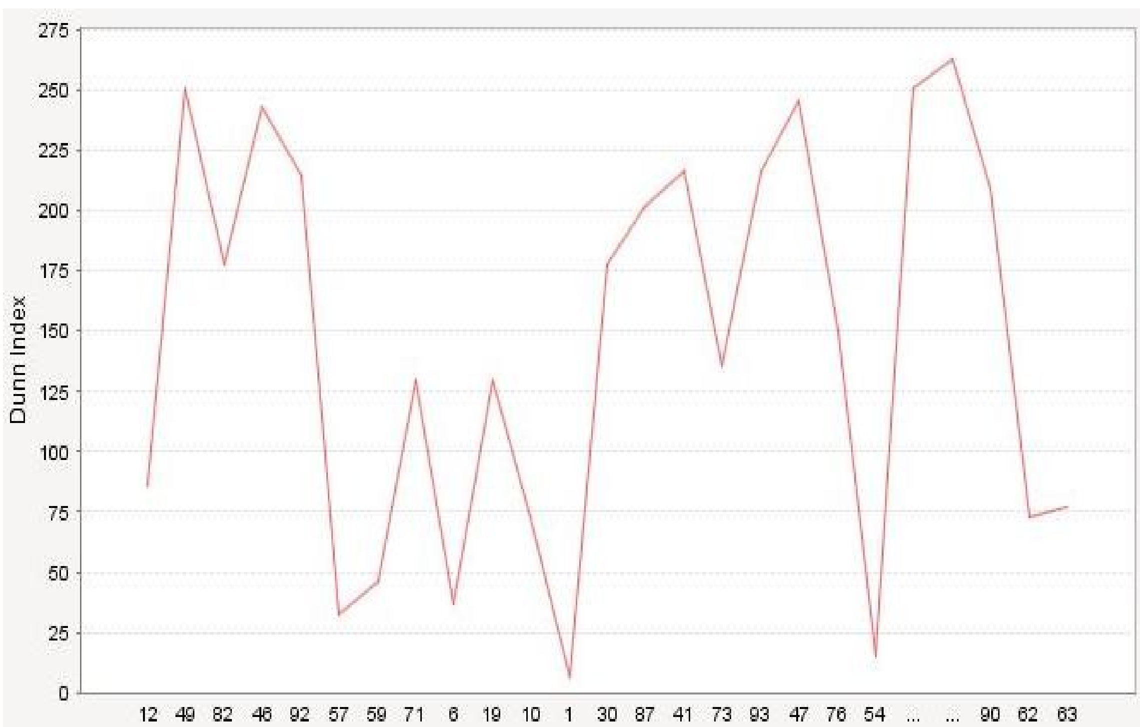

Fig. 4 Dunn Index

5. Conclusion

This paper presents a new time series forecasting algorithm to predict actual-world time series data, with a novel clustering-based method. The approach is particularly targeted at electricity price forecasting, and the accuracy is tested against current approaches. Key Contributions of the Proposed Forecasting Method, Labeling-Based Clustering for Prediction. Rather than employing raw time series values per se, the algorithm initially clusters the data and labels. The labels alone are employed for forecasting, and the actual values are included at the last step. Pattern Matching Forecasting (PMF) for Increased Accuracy. Looks for past label sequences with the same current pattern. Applies recursive feedback (closed-loop) to enhance long-term prediction. Validated Performance - Accuracy of forecasting was evaluated on electricity price data. Results show improved performance over classical models. This clustering + forecasting approach using a combination of both captures patterns and interdependencies in electricity price patterns effectively. The label-based method lessens noise and enhances generalization.

References

[1] S.K. Aggarwal, L.M. Saini, and A. Kumar, “Price Forecasting Using Wavelet Transform and ISE Based Mixed Model in Australian Electricity Market,” Int’l J.Energy Sector Management, vol. 2, no. 4, pp. 521-546, 2008.

[2] S.K. Aggarwal, L.M. Saini, and A. Kumar, “Electricity Price Forecasting in Deregulated Markets: A Review and Evaluation,” Int’l J. Electrical Power and Energy Systems, vol. 31, no. 1, pp. 13-22,2009.

[3] A. Abraham and B. Nath, “A Neuro-Fuzzy Approach for Forecasting Electicity Demand in Victoria,” Applied Soft Computing J., vol. 1, no. 2, pp. 127-138, 2001.

[4] F.J. Nogales and A.J. Conejo, “Electricity Price Forecasting through Transfer Function Models,” J. Operational Research Soc., vol. 57, pp. 350-356, 2006.

[5] R. Xu and D.C. Wunsch II, “Survey of Clustering Algorithms,” IEEE Trans. Neural Networks, vol. 16, no. 3, pp. 645-678, May 2005.

[6] A. Troncoso, J.C. Riquelme, J.M. Riquelme, J.L. Martıinez, and A. Goimez, “Electricity Market Price Forecasting Based on Weighted Nearest Neighbours Techniques,” IEEE Trans. Power Systems,vol. 22, no. 3, pp. 1294-1301, Aug. 2007.

[7] N.M. Pindoria, S.N. Singh, and S.K. Singh, “An Adaptative Wavelet Neural Network-Based Energy Price Forecasting in Electricity Markets,” IEEE Trans. Power Systems, vol. 23, no. 3, pp. 1423-1432, Aug. 2008.

[8] F. Martinez_ Alvarez, A. Troncoso, J.C. Riquelme, and J.S. AguilarRuiz,“LBF: A Labeled- Based Forecasting Algorithm and Its Application to Electricity Price Time Series,” Proc. Eighth IEEE Int’l Conf. Data Mining, pp. 453-461, 2008.

[9] G. Li, C.C. Liu, C. Mattson, and J. Lawarreie, “Day-Ahead Electricity Price Forecasting in a Grid Environment,” IEEE Trans. Power Systems, vol. 22, no. 1pp. 266-274, Feb. 2007.

[10] C. Garcia-Martos, J. Rodriguez, and M.J. Sanchez, “Mixed Models for Short-Run Forecasting of Electricity Prices: Application for the Spanish Market,” IEEE Trans. Power Systems, vol. 22, no. 2, pp. 544- 552, May 2007.

[11] R.C.Garcia, J.Contreras, M.van Akkeren, and J.B. Garcia, “AGARCH Forecasting Model to Predict Day-Ahead Electricity Prices,” IEEE Trans. Power Systems, vol. 20, no. 2, pp. 867-874,May 2005.

[12] A.J. Conejo, M.A. Plazas, R. Espinola, and B. Molina, “Day-AheadElectricity Price Forecasting Using the Wavelet Transform andARIMA Models,” IEEE Trans. Power Systems, vol. 20, no. 2,pp. 1035- 1042, May 2005.