(PAPER FOR PRESENTATION ONLY IN THE WORKSHOP OF APPLIED INFRASTRUCTURE - BERLIN, GERMANY, OCTOBER 11, 2003)
(A SIMILAR VERSION OF THIS PAPER IS SUBMITTED TO THE "IEEE POWER SYSTEMS JOURNAL")

A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices

Reinaldo C. Garcia
German Institute of Economic Research, DIW (Berlin), Germany

Javier Contreras
Universidad de Castilla – La Mancha, Spain

Marko van Akkeren
PMI Group, Walnut Creek, CA, USA

and

João Batista C. Garcia
Dexia Bank, Brussels, Belgium

September 27, 2003

Keywords: Electricity markets, forecasting, volatility, time series analysis, GARCH models.

JEL-Classification: A, C, D, E, G, H, K, L, M, P, Z

Contact Details: Reinaldo C. Garcia

Full affiliation: Senior Researcher, Department of Energy, Transportation and Environment, German Institute of Economic Research, DIW
Address: Königin-Luise-Straße 5, 14195 Berlin, Germany
Tel: + 49 30 897 89 666
Fax: + 49 30 897 89 113
E-mail: rgarcia@diw.de

Contact Details: Javier Contrears
Full affiliation: Associate Professor, E.T.S. de Ingenieros Industriales, Universidad de Castilla – La Mancha
Address: E.T.S. de Ingenieros Industriales, Universidad de Castilla – La Mancha, 13071 Ciudad Real, Spain.
E-mail: Javier.Contreras@uclm.es

Contact Details: Marco Van Akkeren
Full affiliation: Forecasting Department, PMI Group (walnut Creek, CA, USA)
E-mail: marko.vanakkeren@pmigroup.com

Contact Details: João Batista C. Garcia
Full affiliation: Manager, Derivatives Group, Dexia Bank, Brussels, Belgium
E-mail: joaobatista.crispinianogarcia@dexia.com

A GARCH Forecasting Model to Predict Day-Ahead Electricity Prices

Abstract: Price forecasting is becoming increasingly relevant to producers and consumers in the new competitive electric power markets. Both for spot markets and long-term contracts, price forecasts are necessary to develop bidding strategies or negotiation skills in order to maximize benefit. This paper provides a method to predict next-day electricity prices based on the GARCH methodology being already used to analyse time series models in general. A detailed explanation of the aforementioned GARCH models and results from mainland Spain and Californian markets are presented.

I Introduction

Price forecasting has become a very valuable tool in the currently upheaval of deregulation in electricity markets. The companies that trade in electricity markets make extensive use of price prediction techniques either to bid or to hedge against volatility. When bidding in a pool system, the market participants are requested to express their bids in terms of prices and quantities. Since the bids are accepted in order of increasing price until the total demand is met, a company that is able to forecast the pool price can adjust its own price/production schedule depending on hourly pool prices and its own production costs [1].

Another market instrument to trade in the market is the bilateral contract system. In this setting, a buyer and a seller agree on a certain amount to be transferred through the network at a certain fixed price. This price is agreed by both sides beforehand and it is also based on price predictions. The reason is that most of the deregulated electricity markets use a mixed bag of pool and bilateral contracts. If this is the case, companies have to optimize their production schedules such that they can hedge pool price volatility via bilateral contracts. Thus, a good knowledge of future pool prices helps to valuate more accurately bilateral contracts.

In recent years, several methods have been applied to predict prices in electric markets. For example, Transfer function models [2] and ARIMA models have been tested in the Spanish [3] and the Norwegian markets [4]. In addition, Artificial Neural Networks (ANN) have been applied to the England-Wales pool [5] and the Australian market [6]. Other techniques, such as Fourier Transform [7] and stochastic modeling [8] have addressed the same problem.

As mentioned earlier, one key aspect of pool prices is their volatility, at least during certain periods. Also note that not only price volatility is important per se, but also crucial to calculate average annual prices and to derive from them bilateral contract prices.

Spot price volatility has been recently studied in several publications. Benini et al. [9] have analyzed several markets, such as Spain, California, England and Wales and the PJM system. Mount [10] has claimed that a uniform auction worsens this problem as compared to a discriminatory auction in the England-Wales system. The Californian market has also served as a benchmark to apply Value-at-Risk models [11] or stochastic linear regression models [12].

General Autoregressive Conditional Heteroskedasticity (GARCH) models [13][14] consider that the price series is not invariant (i.e. the error term: real value minus forecasted value does not have 0 mean and constant variance), as it happens in an ARIMA process. The error term is now assumed to be serially correlated and can be modeled by an Autoregressive (AR) process. Thus, a GARCH process can measure the implied volatility of a series due to price spikes. For example, California experienced huge price spikes during the summer of 2000 that led to the closure of the market [11][12] until new rules were developed.

This paper focuses on day-ahead forecasts of daily electricity markets with high volatility periods using a GARCH methodology. Our GARCH models provide the 24 forecasts of the clearing prices for the next day based on historical data [15][16]. To illustrate the models, forecasts of prices in mainland Spain [17] and California [18] using GARCH processes are presented and discussed.

The paper is organized as follows. In Section II a general GARCH methodology and the two models obtained for the Spanish and Californian day-ahead markets are shown. Section III presents numerical results of the simulations and Section IV states several conclusions.

II Garch Methodology

The information recovery process in time-series analysis uses historical observations to derive estimates of current and future values of the dependent variable. Among the most popular estimation techniques are the maximum-likelihood (ML) approach which requires the availability of information on the entire probability distribution, generalized methods of moments (GMM) which reduces the informational requirements to specific moments of the data, and nonparametric procedures. Nonlinear neural network (NNM) models represents a more modern estimation technique which has gained popularity in recent years. Given the Gaussian distribution for the time series is satisfied, we adhere to the Box-Jenkins modelling approach of parsimony, i.e. using the fewest model parameters as supported by the data, to estimate an ARMA process with conditional-heteroscedastic (GARCH) error components.

The ARMA(p,q) process includes components of both autoregressive and moving average terms and is defined as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \quad (1)$$

or applying the backshift operator, L,

$$(1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p)\, y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q)\, \varepsilon_t \quad (2)$$

and may be equivalently be expressed as:

$$\psi(L)\, y_t = \omega(L)\, \varepsilon_t \quad (3)$$

To simplify our analysis, we assume y_t is covariance-stationary, meaning that the moments of our process are not dependent on t. In traditional ARMA estimation, the basic assumption on the error terms include zero mean and constant variance, or specifically (i) $E(\varepsilon_t)=0$, (ii) $E(\varepsilon^2_t)=\sigma^2$, and (iii) $E(\varepsilon_t\varepsilon_s)=0$ for s=t. In particular, assumptions (ii) and (iii) do not necessarily need to hold. Our research has found that the generalized-heteroscedastic error specification is strongly supported by the data, and moreover significantly improves both goodness of fit and out-of-sample predictive ability of our model for estimating energy price movements.

To accommodate the possibility of serial correlation in volatility, the autoregressive conditional heteroscedasticity-ARCH(m) class of processes is introduced by Engle (1982) that takes the form:

$$\sigma^2_t = c + \alpha_1 x^2_{t-1} + \alpha_2 x^2_{t-2} + \ldots + \alpha_m x^2_{t-m} = c + \alpha(L)\, x^2_t \quad (4)$$

Since σ^2_t cannot be negative, it follows that c>0 and $\alpha_1 >= 0$ for i=1,2,…,m. In order for (4) to reflect the degree of autocorrelation displayed by the data, it may be necessary for m to become relatively large. The GARCH(r,m) model proposed by Bollerslev (1986) reduces the dimensionality by adding autoregressive terms in:

$$\sigma^2_t = c + \alpha_1 x^2_{t-1} + \alpha_2 x^2_{t-2} + \ldots + \alpha_m x^2_{t-m} + \beta_1 \sigma^2_{t-1} + \beta_2 \sigma^2_{t-2} + \ldots + \beta_m \sigma^2_{t-r} \quad (5)$$

$$= c + \alpha(L)x^2_t + \beta(L)\sigma^2_t \quad (6)$$

The GARCH(1,1) represents the most popular model expressed as:

$$\sigma^2_t = c + \alpha_1 x^2_t + \beta_1 \sigma^2_{t-1}$$
$$= c + \alpha_1(x^2_t - \sigma^2_{t-1}) + (\alpha_1 + \beta_1)\sigma^2_{t-1} \quad (7)$$

because it incorporates mean reversion and the dynamics of σ^2 can be explained through past volatility shocks α_1. The general scheme applied to obtain the proposed GARCH model is as follows:

Step 0. A class of models is formulated assuming certain hypotheses.
Step 1. A model is identified for the observed data.
Step 2. The model parameters are estimated.
Step 3. If the hypotheses of the model are validated go to Step 4, otherwise go to Step 1 to refine the model.
Step 4. The model can be used to forecast.
In the following subsections, each step of the above scheme is detailed.

Step 0

In this step, a general GARCH formulation is selected to model the price data. This selection is carried out by careful inspection of the main characteristics of the hourly price series. In most of the competitive electricity markets this series presents: high frequency, non-constant mean and variance, and multiple seasonality (corresponding to daily and weekly periodicity, respectively), among others. These factors are among the main ones applied when selected the GARCH model.

Step 1

A trial model, as seen in (1), must be identified for the price data. In a first trial, the observation of the autocorrelation and partial autocorrelation plots of the price data can help to make this selection. In successive trials, the same observation of the residuals obtained in Step 3 (observed values minus predicted values) can refine the structure of the functions in the model.

Step 2

After the functions of the model have been specified, the parameters of these functions must be estimated. Good estimators of the parameters can be found by maximizing the likelihood with respect to the parameters. The Eviews System is used to estimate the parameters of the model in the previous step.

Step 3

In this step, a diagnosis check is used to validate the model assumptions of the GARCH model. Among the test to validate the assumptions of the GARCH model chosen is a careful inspection of the autocorrelation and partial autocorrelation plots of the residuals.

Step 4

In step 4, the model from Step 2 can be used to predict future values of prices.

As a result of these five steps, the final models for the Spanish and Californian electricity markets for the year 2000 are shown next in (8) and (9), respectively:

- For Spain, it was applied the following GARCH(2,1) Model:

$$p_t = c + (\phi_1 L + \phi_2 L^2 + \phi_2 L^3 + \phi_{24} L^{24} + \phi_{48} L^{48} + \phi_{120} L^{120} + \phi_{144} L^{144} + \phi_{168} L^{168}) + \varepsilon_t \quad (8)$$

- For California, it was applied the following GARCH(2,1) Model:

$$p_t = c + (\phi_1 L + \phi_2 L^2 + \phi_2 L2^3 + \phi_{24} L^{24} + \phi_{47} L^{47} + \phi_{48} L^{48} + \phi_{120} L^{120} + \phi_{144} L^{144} + \phi_{167} L^{167} + \phi_{168} L^{168} + \phi_{169} L^{169} + \phi_{192} L^{192}) + \varepsilon_t \quad (9)$$

The results obtained applying the two models described above are presented in the next section.

III Numerical Results

The Garch models in (8) and (9) have been applied to predict the electricity prices of mainland Spain and California, respectively. The data set used to obtain the proposed GARCH model for the Spanish market consists of hourly electricity prices from September 01, 1999 to December 31, 2000. Similarly, the data set used to model the California market consists of data from January 1, 2000 to December 31, 2000.

For the Spanish electricity market, three weeks have been selected to forecast and validate the performance of the Garch model. The first one corresponds to the last week of May 2000 (from May 25th to 31st.) The second one corresponds to the last week of August 2000 (from August 25st to 31th), which is typically a low demand week. The third corresponds to the third week of November 2000 (from November 9th to 15th), which is typically a high demand week. For the California electricity market, the week of April 3rd. to 9th, 2000 has been chosen. This week is prior in time to the beginning of the dramatic price volatility period that took place afterwards.

Numerical results with the GARCH models are presented. Figs.1-4 show the forecasted prices resulting from the GARCH models for each of the four weeks studied; three for the Spanish electricity market, and one for the Californian market, together with the actual prices. Fig. 1 corresponds to the selected week in May for the Spanish market. The seven daily mean errors for this May week of the Spanish market appear in Table I. A good performance of the prediction method can be observed. The daily mean errors are around 4%, where the lowest mean error is 2.60% and the highest one, 7.60%.

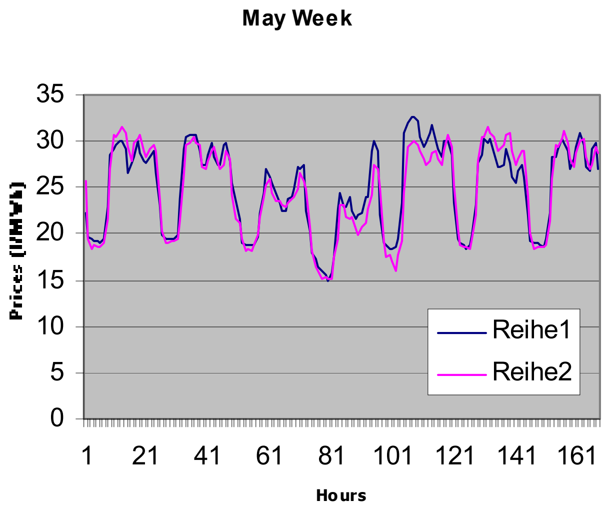

Fig 1. Forecast of May week in the Spanish market. Prices in €/MWh.

TABLE I
DAILY MEAN ERRORS OF MAY WEEK IN THE SPANISH MARKET

Days	1	2	3	4	5	6	7
Mean	4.62%	2.90%	3.11%	5.95%	7.60%	5.22%	3.40 %

Fig. 2 corresponds to the selected week in August for the Spanish market (from August 25st to 31th.) The seven daily mean errors appear in Table II. The daily mean errors for the selected week in August are around 7%, where the lowest mean error is about 4.8% and the highest one, 10.4%. Fig. 3 corresponds to the selected week in November for the Spanish market (from November 9th to 15th in 2000.) The seven daily mean errors appear in Table III. The daily mean errors for this selected week in November are around 6%, where the lowest value is 5.5% and the highest one, 10.4%.

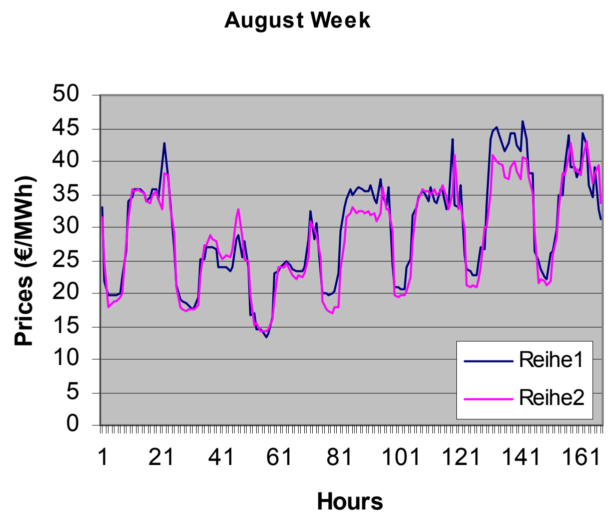

Fig. 2. Forecast of August week in the Spanish market. Prices in €/MWh.

TABLE II
DAILY MEAN ERRORS OF AUGUST WEEK IN THE SPANISH MARKET

Days	1	2	3	4	5	6	7
Mean	4.81%	6.66%	5.36%	10.4%	7.27%	10.1%	7.17%

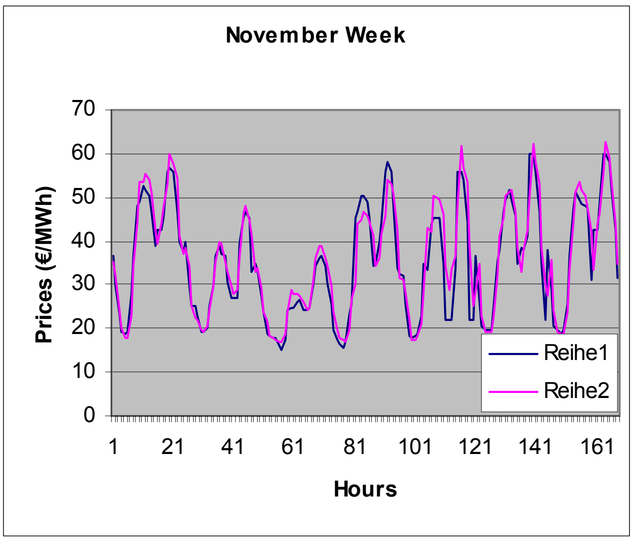

Fig. 3. Forecast of November week in the Spanish market. Prices in €/MWh.

TABLE III
DAILY MEAN ERRORS OF NOVEMBER WEEK IN THE SPANISH MARKET

Days	1	2	3	4	5	6	7
Mean	10.2%	5.81%	6.68%	6.19%	6.66%	5.47%	6.01%

Fig. 4 corresponds to the selected week in April (3rd. to 9th. of April, 2000) for the

California market. The seven daily mean errors for this week appear in Table IV. The daily mean errors are around 4%, where the lowest value is 3.0% and the highest one, 4.6%

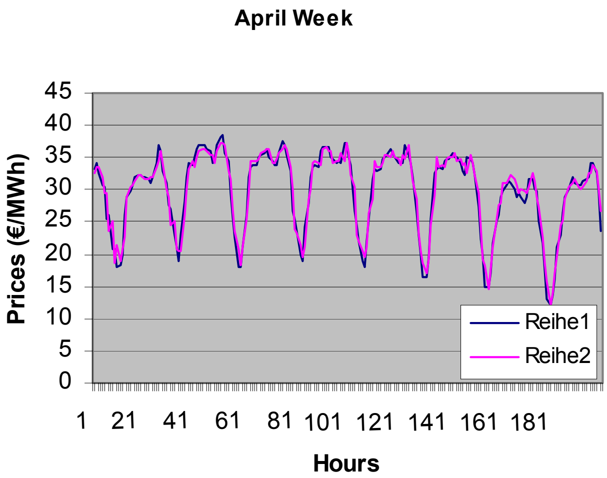

Fig. 4. Forecast of April week in the Californian market. Prices in \$/MWh.

TABLE IV
DAILY MEAN ERRORS OF APRIL WEEK IN THE CALIFORNIAN MARKET

Days	1	2	3	4	5	6	7
Mean	4.23%	3.03%	3.81%	3.32%	3.38%	4.64%	4.09%

In general, it could be said that the results obtained by the model are quite reasonable, as the errors obtained are no larger than 10%. Nevertheless, to verify the prediction accuracy of the GARCH model, different statistical measures are utilized.

For the four weeks under study, the average prediction error of the 24 hours is computed for each day. Then, the average of the daily mean errors is calculated: Mean Week Error (MWE.) Finally, the Forecast Mean Square Error (FMSE) for the 168 hours of each week is derived.

Tables V and VI present the numerical results as follows. The second column of both tables shows the percentage Mean Week Error (MWE), and the third one presents the square root of the Forecast Mean Square Error (FMSE):

$$\sqrt{FMSE} = \sqrt{\sum_{i=1}^{168} (p_t - \hat{p}_t)^2}$$

where p_t and $\hat{p}_t$ are the actual and forecasted prices, respectively. Note that prices and $\sqrt{FMSE}$ are measured in €/MWh and \$/MWh in the Spanish and Californian markets, respectively.

In addition to the four weeks under study, and for the sake of completion, Table V shows the statistical measures for the last week of the first ten months of the year 2000 in Spain, and November, in which the third week is selected. Table V also shows the results for the weeks of April 3rd to 9th, the week of August 21st to 27th, and the week of November 13th to 19th, 2000, for the California market. Note that, after April 2000, this market experienced high spikes that provoked its collapse at the end of 2000.

TABLE V
STATISTICAL MEASURES

Spanish Market	MWE (%)	$\sqrt{FMSE}$
January	7.77	55.68
February	6.78	35.62
March	8.99	54.00
April	9.07	31.36
May	4.68	20.35
June	8.37	49.86
July	7.61	39.57
August	7.33	36.12
September	7.90	52.86
October	10.76	70.53
November	6.72	56.61
California Market	MWE (%)	$\sqrt{FMSE}$
April	3.52	35.87
August	8.69	203.94
November	3.15	156.92

All the study cases have been run on a PC with one processor Pentium IV with 128 Gb of RAM at 1000 MHz. Running time, including estimation and forecasting, has been under two minutes for each one of the cases, applying the Eviews software.

IV Conclusions

This paper has proposed two GARCH models to predict hourly electricity prices in the electricity markets of Spain and California, respectively. Average errors in the Spanish market have been around 7% (depending on the studied week), and around

4% in the Californian market. These errors are very reasonable ones, taking into account the complex nature of price time series and the results previously reported in the technical literature. The differences in the models for the Spanish and California markets may reflect different bidding structures.

In the future, improvements to the models, such as special treatment for weekend data (calendar effect), the inclusion of exogenous variables (demand, water storage, etc.) and the issue of volatility in the price will be addressed.

REFERENCES

S. de la Torre, A. J. Conejo and J. Contreras, "Simulating Oligopolistic Pool-Based Electricity Markets: A Multiperiod Approach," accepted for publication in *IEEE Trans. Power Systems*. Paper TPWRS-00466-2002.

F. J. Nogales, J. Contreras, A. J. Conejo and R. Espínola, "Forecasting Next-Day Electricity Prices by Time Series Models," *IEEE Trans. Power Systems*, vol. 17, no. 2, pp. 342-348, May 2002.

J. Contreras, R. Espínola, F. J. Nogales and A. J. Conejo, "ARIMA Models to Predict Next-Day Electricity Prices," *IEEE Trans. Power Systems*, vol. 18, no. 3, pp. 1014-1020, August 2003.

O. B. Fosso, A. Gjelsvik, A. Haugstad, M. Birger and I. Wangensteen, "Generation Scheduling in a Deregulated System. The Norwegian Case," *IEEE Trans. Power Systems*, vol. 14, no. 1, pp. 75-81, February 1999.

B. Ramsay and A. J. Wang, "An Electricity Spot-Price Estimator with Particular Reference to Weekends and Public Holidays," *Proc. of the Universities Power Engineering Conference, UPEC'97*, Manchester, UK, September 1997.

B. R. Szkuta, L. A. Sanabria and T. S. Dillon, "Electricity Price Short-Term Forecasting using Artificial Neural Networks," *IEEE Trans. Power Systems*, vol. 14, no. 3, pp. 851-857, August 1999.

J. D. Nicolaisen, C. W. Richter, Jr. and G. B. Sheblé, "Signal Analysis for Competitive Electric Generation Companies," *Proc. of the Conference on Electric Utility Deregulation and Restructuring and Power Technologies 2000*, City University, London, U.K., 4-7 April, 2000.

J. Valenzuela and M. Mazumdar, "On the Computation of the Probability Distribution of the Spot Market Price in a Deregulated Electricity Market," in *Proc. of the 21st Power Industry Computer Applications International Conference, PICA'2001*, Sydney, Australia May 2001.

M. Benini, M. Marracci, P. Pelacchi and A. Venturini, "Day-Ahead Market Price Volatility Analysis in Deregulated Electricity Markets," in *Proc. 2002 IEEE Power Engineering Society Summer Meeting, Jul. 21-25 2002, Chicago, IL, United States*, vol. 3, pp. 1354-1359.

T. Mount, "Market Power and Price Volatility in Restructured Markets for Electricity," *Decision Support Systems*, vol. 30, pp. 311-325, 2001.

R. W. Dahlgren, C.-C. Liu and J. Lawarree, "Volatility in the California Power Market: Source, Methodology and Recommendations," *IEE Proc.-Gener. Transm. Distrib.*, vol. 148, no. 2, pp. 189-193, March 2001.

A. Fossa, "The California Electricity Market Analysis during the Summer 2000: Application of Stochastic Techniques," Ph.D. dissertation, Dipartimento de Ingegneria Elettrica, Università degli Studi de Pavia, 2001.

J. D. Hamilton, *Time Series Analysis*. Princeton University Press, 1994.

W. Enders, *Applied Econometric Time Series*. New York: Wiley, 1995.

Compañía Operadora del Mercado Español de Electricidad (Spanish Market Operator), http://www.omel.es.

California Electricity Market Data, University of California Energy Institute, http://www.ucei.berkeley.edu/datamine/datamine.htm.

J. J. González and P. Basagoiti, "Spanish Power Exchange and Information System Design Concepts, and Operating Experience," in *Proc. of the 21st Power Industry Computer Applications International Conference, PICA'99*, Santa Clara, USA, May 1999.

F. Albuyeh and Z. Alaywan, "Implementation of the California Independent System Operator," in *Proc. of the 21st Power Industry Computer Applications International Conference, PICA'99*, Santa Clara, USA, May 1999.