Electric Power Systems Research 117 (2014) 115–122

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

Energy price prediction multi-step ahead using hybrid model in the Brazilian market

José C. Reston Filho [a], Carolina de M. Affonso [b,*], Roberto C.L. de Oliveira [b]

[a] IDAAM, Manaus, Amazonas, Brazil
[b] Institute of Technology, Federal University of Pará, Pará, Brazil

ARTICLE INFO

Article history:
Received 18 October 2013
Received in revised form 30 May 2014
Accepted 10 August 2014
Available online 28 August 2014

Keywords:
ARIMA
NN
Electricity market
Energy price forecasting
Explanatory variable selection
Multi-step ahead

ABSTRACT

This paper proposes a new hybrid approach for short-term energy price prediction. This approach combines auto-regressive integrated moving average (ARIMA) and neural network (NN) models in a cascaded structure and uses explanatory variables. A two step procedure is applied. In the first step, the selected explanatory variables are predicted. In the second one, the energy prices are forecasted by using the explanatory variables prediction. Further, the proposed model considers a multi-step ahead price prediction (12 weeks-ahead) and is applied to Brazilian market, which adopts a cost-based centralized dispatch with unique characteristics of price behavior. The results show good ability to predict spikes and satisfactory accuracy according to error measures and tail loss test when compared with traditional techniques. Thus, the model can be an attractive tool to mitigate risks in purchasing power.

1. Introduction

Electricity price forecasting is an important issue to all market participants in order to decide bidding strategies and to establish bilateral contracts, maximizing their profits and minimizing their risks. Energy price typically exhibits seasonality, high volatility and spikes. Also, energy price is influenced by many factors such as power demand, weather, and fuel price. These issues make the energy price prediction a complex task, and several forecasting methods have been reported in the literature in recent years [1].

Electricity price prediction models have been classified mainly in three groups: game theory models, fundamental models and time series models [1,2]. The first group is based on game theory, which analyzes the strategic behavior of the agents and its impact on electricity prices, considering as a key point equilibrium models (like Nash equilibrium) [3]. The second group is based on fundamental or simulation models, which simulates the exact physical model of the power system. This approach express electricity prices based on marginal generation costs considering transmission congestion, losses, and other ancillary service requests in power markets [4]. Finally, the third group is based on time series models, which includes regression-based models such as auto-regressive integrated moving average (ARIMA) and generalized auto-regressive conditional heteroscedastic (GARCH), neural network (NN), fuzzy logic and others [5]. These models are mainly focused on data analysis, without examining the underlying physical processes in detail. They use historical prices series and sometimes, others explanatory factors such as temperature, time of day, and load demand. This paper focuses on time series models.

Recently, hybrid forecast methods with promising results have also been proposed in the literature. The benefit of hybrid models is to combine strengths of different techniques providing a robust modeling framework. For instance, Ref. [6] proposed a hybrid method based on wavelet transform and ARIMA models to forecast day-ahead energy prices in the Spain market. The authors conclude that the use of the wavelet transform as a preprocessor of forecasting data improves the predicting behavior of ARIMA models. Ref. [7] proposed a least-squares support vector machine (LSSVM) combined with autoregressive moving average with exogenous variables (ARMAX) model to predict hourly price for an entire month in PJM (Pennsylvania–New Jersey–Maryland) market. The LSSVM method is used as the major forecasting module to predict the initial electricity prices. Next, the ARMAX method is used to improve the forecasting results. The proposed model improved forecasting accuracy compared to a forecasting model using a single LSSVM. However, the method still has low accuracy to forecast peak prices. In [8] a hybrid model is proposed combining time-series methods (ARMAX + GARCH) and an adaptive wavelet neural

* Corresponding author at: Augusto Correa, 01, 66075-110 Belém, Pará, Brazil. Tel.: +55 91 32417944.
E-mail addresses: jcreston@gmail.com (J.C.R. Filho), carolina@ufpa.br (C.d.M. Affonso), limao@ufpa.br (R.C.L. de Oliveira).

http://dx.doi.org/10.1016/j.epsr.2014.08.006
0378-7796/

network (AWNN) model for day-ahead price forecast in PJM market. In this method power load is used as explanatory variable. The ARMAX model is used to catch the linear relationship between price return series and explanatory variable load series. The GARCH model is used to show the heteroscedastic character of residuals, and AWNN is used to present the nonlinear, non-stationary impact of load series on electricity prices. Ref. [9] proposed a combination of wavelet transform, NN and an evolutionary algorithm for day-ahead price prediction in PJM market. The obtained results were compared with other traditional techniques and presented smaller errors.

Unlike load forecasting methods which have prediction errors below 3%, price forecasting techniques are still in their early stages of maturity and the reported errors generally range from 5% to 36% [10], and vary according to the technique used and the market analyzed. A review of different electricity price forecasting methods can be found in [1], and most methods are applied to PJM, New England and Spain markets. It is important to mention that few studies were developed in energy price prediction area applied to the Brazilian market [11,12].

In this context, this paper proposes a new hybrid approach for short-term energy price prediction using explanatory variables. The proposed method is a two-step procedure. In the first step, the selected explanatory variables are predicted. In the second step, the electricity price is forecasted by using the explanatory variables prediction. The results obtained with the proposed approach are compared with traditional techniques like ARIMA, GARCH, Exponential Smooth and NN. The prediction time horizon is 12 weeks-ahead.

The main contributions of this paper are the forecast model itself and its application to the Brazilian market, which adopts a cost-based model with unique features of energy price behavior. Some important aspects of the proposed price forecasting method are:

- combines ARIMA and NN models and does not assumes that the relationship between the linear and nonlinear components are additive (may be multiplicative for example), which may degrade predictor performance;
- uses explanatory variables selection technique;
- is less complex compared to other estimation methods;
- adopts a prediction time horizon of n-steps ahead;
- is applied to the Brazilian market;

The proposed hybrid model was applied to the Brazilian market and the results were compared to other techniques available in the literature applied to other markets. The results show comparable or even better error measures than other estimation methods.

This paper is organized as follows. Section 2 presents the main features and peculiarities of the Brazilian electricity market. The explanatory variable selection method is discussed in Section 3. Section 4 presents the proposed hybrid approach and important aspects of data-preparation. The results are presented and discussed in Section 5. Finally, the main conclusions are summarized in Section 6.

2. Brazilian electricity market

The Brazilian energy market operates with two trading environment, one regulated and the other free [13]. On the free market, buyers and sellers are free to establish bilateral contracts and negotiate prices, quantities, delivery dates, and conditions. On the regulated market, a pool of purchasing agents (distributors) buys power from selling agents (generators, independent power producers or self-producers) in public auctions under set prices. The differences between the amounts of energy contracted and those effectively consumed or produced by the agents are accounted in the short-term market (spot market) based on the spot price called PLD (settlement price for the differences).

The PLD is calculated weekly and is based on the system marginal cost of operation, obtained as a result of an optimization process to dispatch generators. The PLD is limited by a minimum and maximum value established by the Brazilian Electricity Regulatory Agency (ANEEL), and is evaluated to each submarket associated with the country regions: North, Northeast, Center-west/Southeast and South.

The optimization model adopted considers the immediate benefit of using water in the reservoirs (immediate cost) and the future benefits of its storage (future cost), reducing the use of fuel in thermal units [14]. As a consequence, the spot price obtained based on marginal cost of operation is strongly dependent on the water level and inflow energy in reservoir plants, and does not respond directly to load variations as in most energy markets.

Then, Brazil uses a cost-based market instead of a bid-based market, and adopts a tight pool model with a centralized and least cost dispatch organized by National System Operator (ONS). This scheme is adopted due to the country peculiarities, which has an installed capacity of 121 GW where 65.96% corresponds to hydro generation. The hydro system is composed of several reservoirs capable of multi-year regulation located at the same river with different owners.

3. Explanatory variables selection

In prediction models, explanatory variable is one that may explain or may cause differences in a response variable. The term "explanatory variable" is preferred by some authors over "independent variable" since the quantities treated as "independent variables" may not be statistically independent [15–17]. In this case, the dependent variable is referred to as "response variable".

In this study, the response variable is the PLD (Brazilian spot price). Also, the demand has been the most commonly examined explanatory variable in price forecasting studies [18]. However, according to market particularities, energy prices can also be explained by other variables [16]. The potential explanatory variables used in this paper are selected based on the Brazilian market behavior, and according to the Brazilian National System Operator the most important variables are related to hydrological conditions, power load and fuel prices of thermal units. These variables are:

- Stored energy in reservoirs (% MLT);
- Inflow energy in reservoirs (% MLT);
- Total hydro generation (MWmed);
- Total thermal generation (MWmed);
- System power load (MWmed);

where MLT is the long-term average (historical average of 79 years).

After the identification of the potential explanatory variables, an explanatory variable selection method is applied to find the optimal set of input variables required to describe the behavior of the energy price, which should contain minimum degree of redundancy [17]. Two procedures were applied: significance and importance analysis. The significance was analyzed with the global F-test implemented by means of analysis of variance (ANOVA) [19]. The aim is to test how two or more input variables act together to affect the output variable, and determine whether they improve prediction of the desired output. The p-value associated to the F-Statistic in the general regression model is used to select the significant variables. Variables with a p-value lower than a fixed threshold will be selected ($p < 0.05$). The predictor importance was computed with a leave-one-out method, based on the residual sum of squares (SSE) by removing one predictor at a time from the final full model [20].

The results of significance and importance to each submarket are presented in Table 1. The significance analysis shows that hydro generation should be eliminated from North and Northeast submarkets. Also, importance analysis indicates that inflow energy should be removed from Center-west/Southeast and South submarket. Then, the selected set of explanatory variables to each submarket is:

- North: Stored energy, Inflow energy and Load;
- Northeast: Stored energy, Inflow energy, Thermal Generation and Load;
- Center-west/Southeast: Stored energy, Hydro generation, Thermal generation and Load;
- South: Stored energy, Hydro generation, Thermal generation and Load.

4. Proposed Methodology

Consider an energy price time series y_t which is function of n explanatory variables u_t^i, $i = 1,\ldots,n$. The explanatory variable u_t^i is composed of a linear relation structure and a nonlinear component such as:

$$u_t^i = f(L_t, N_t) \tag{1}$$

where L_t denotes the linear component and N_t denotes the nonlinear component. First, we let ARIMA model to forecast the linear components of the explanatory variables series 12-step ahead:

$$\hat{z}_{t+12}^i = L^{ARIMA}(u_t^i) \tag{2}$$

Then, a NN is used to perform the nonlinear modeling of the explanatory variables.

$$\hat{u}_{t+12}^i = N^{NN1}(\hat{z}_{t+12}^i) \tag{3}$$

Later, another NN is used to forecast the energy price using the prediction of all explanatory variables.

$$\hat{y}_{t+12} = N^{NN2}[\hat{u}_{t+12}^1, \hat{u}_{t+12}^2, \ldots, \hat{u}_{t+12}^n] \tag{4}$$

This paper uses the data mining software IBM SPSS Modeler, originally named SPSS Clementine [21]. This software provides all techniques used in this paper to develop and test the proposed hybrid method, such as: ANOVA, PCA, ARIMA, NN and others. The proposed hybrid method is presented in Fig. 1 and can be summarized as follows:

1. Create a large database comprising historical data of PLD and potential explanatory variables that affect the short-term energy price.
2. Apply a explanatory variable selection technique to identify the optimal set of explanatory variables (u_t^i) to each submarket.
3. Forecast each explanatory variable ($\hat{u}_{t+12}^i$) 12-weeks-ahead. This process is described in the box "Forecast each explanatory variable u^i "of the flowchart.

   First, an ARIMA model is used to analyze the linear relations of the series. A single ARIMA model is built based on the entire series. After a satisfactory ARIMA model is found, this model can be used for prediction purposes. The price forecast 12-week ahead is reached via recursion, by feeding input variables with the forecaster outputs. Then, when the price of one week ahead is forecasted, it is used as input for the price prediction of the next week and this cycle is repeated until the price of the next 12 weeks are predicted.

   The ARIMA provides an initial price forecast for the first NN (NN1), which is used to catch the nonlinear relations of the explanatory variables series. In order to guarantee generalization capacity, it is important to reduce the dimension of the input vectors and choose the better learning set before training the NN. This is carried out during the data preparation process applying Principal Component Analysis (PCA) and a balancing procedure, which will be discussed next. The NN has one output node used to forecast the price of the next week. Iterative forecasting technique is used to generate price forecasts of the next 12 weeks. In this case, the forecast values are iteratively used as inputs for the next forecasts.
4. Forecast the PLD 12 weeks ahead ($\hat{y}_{t+12}$ using a second NN (NN2) preceded by the data preparation process again. The NN inputs are the explanatory variables predicted 12 weeks ahead. Again, this NN has one output node and predicts the 12 weeks ahead using iterative forecasting technique.

4.1. ARIMA

ARIMA models are classical time series prediction method, applied to non-stationary linear time series [22]. As the application of these models is very common, it is described here briefly. The ARIMA(p,d,q) model is defined as:

$$\phi_P(q^{-p})\Delta^d y(k) = \theta_0 + \theta_q(q^{-q})V(k) \tag{5}$$

where p and q are the order of the parameters ϕ and θ respectively, θ_0 is the model constant and the operator q^{-p} delays the sample of p steps, y is the predicted variable and v is the model error. Δ^d is the differencing operator given by:

$$\Delta = 1 - q^{-1} \tag{6}$$

The identification of ARIMA model is usually based on the analyses of Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF), and includes data transformation (d differencing) to make the time series stationary. Once the model is specified, its parameters should be estimated, which usually involves the use of a least-square process. After this, a diagnosis check is used to validate the model, examining the residuals. If the residual term is a white noise process, then the model can be used for forecasting purposes. Otherwise, the process should be repeated until an adequate model is found. In this paper, the ARIMA model identification (chose appropriate p, d, q values) and parameter estimation are automatically executed by the software using the Ljung–Box statistics [22].

4.2. Neural network

Neural networks have been successfully applied to a variety of complex problems due to its ability to learn non-linear relationships between input and output patterns, which would be difficult to model with conventional methods [23]. One of the most commonly used NN architecture is the multi-layer perceptron with the back-propagation algorithm, which was used in this paper [24]. It consists of an input layer, hidden layers and an output layer as shown Fig. 2. The back-propagation is a supervised learning algorithm used to adjust the weights and node biases of the neural network in order to minimize the error between the actual output and the desired output of the network [25].

The data set is usually divided into a training set and a test set. The training set is used to construct the NN model and map the relationship between input patterns and output patterns. The training set is used to measure the predictive ability of the NN. Once the neural network is properly trained, it can interpolate/extrapolate patterns using a limited amount of input data.

In this study, many tests were conducted varying the number of hidden layers and neurons to find the neural network architecture that produces best generalization accuracy to each submarket,

Table 1
Significance and importance analyses to each submarket.

	North region		Northeast region		CW/SE region		South region	
	p-Value	Imp	*p*-Value	Imp	*p*-Value	Imp	*p*-Value	Imp
Stored energy	0.001	0.62	0.000	0.18	0.000	0.53	0.000	0.2
Inflow energy	0.050	0.23	0.000	0.09	0.002	–	0.024	–
Hydro generation	0.263	–	0.116	–	0.000	0.05	0.000	0.09
Thermal Generation	–	–	0.000	0.61	0.000	0.3	0.000	0.25
Load	0.023	0.15	0.000	0.12	0.000	0.12	0.000	0.46

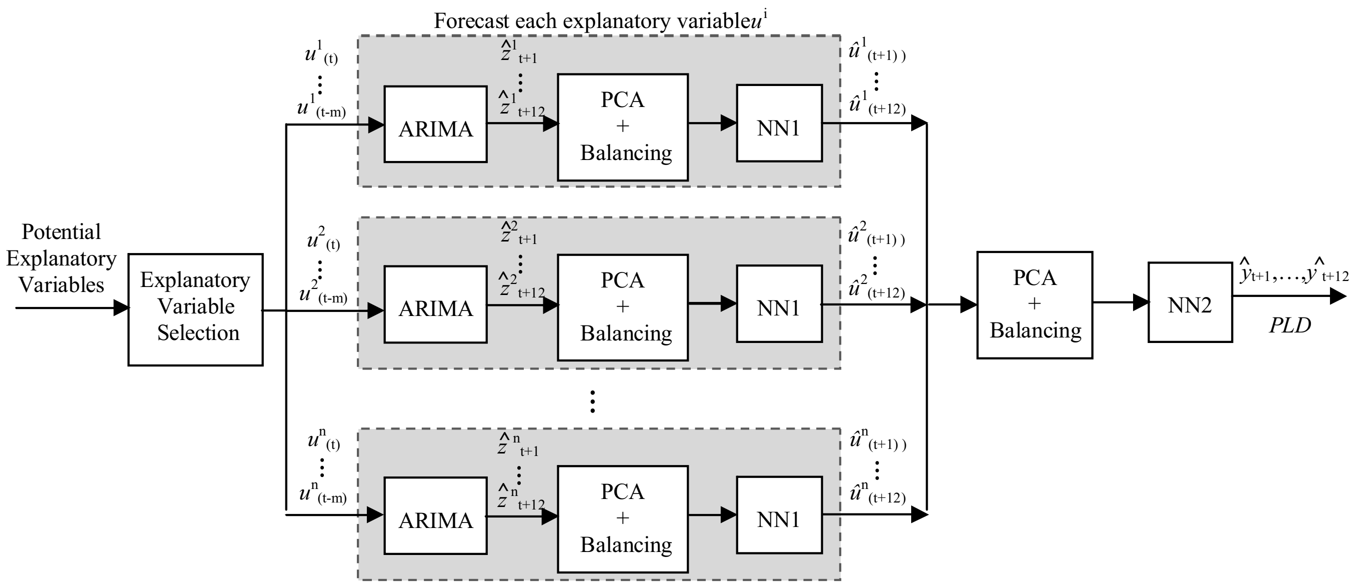

Fig. 1. Flowchart of the proposed hybrid model.

explanatory variable and PLD. In all cases, the NN architecture adopted is composed of one input layer, three hidden layers and one output layer, with sigmoid function in all layers. The training set contains 80% of the data and the test set contains 20% of the data.

4.3. Database

The database used in this paper contains the electricity prices data taken from Brazilian Electrical Energy Commercialization Chamber website [26] presented on a weekly basis, and the explanatory variables data taken from National System Operator website [27], presented on a daily basis. The data was first standardized to a weekly basis and a large database was constructed using data from 2002 to 2009 to each submarket: North, Northeast,

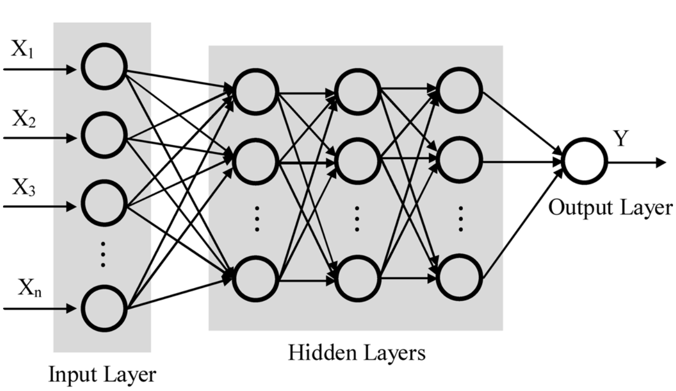

Fig. 2. Neural network model.

South and Center-west/Southeast. Table 2 provides the summary statistic to all variables and Fig. 3 shows the time series data to the Center-west/Southeast region.

4.4. Data preparation

Sometimes, the dimension of the input vector is large but the components of the vectors are highly correlated (redundant). In this situation it is useful to reduce the dimension of the input vectors discarding repetitive data, and choose only those which contain maximum information regarding different patterns or variations of the whole set of input data. This procedure is called Principal Component Analysis and is applied in this paper during the data preparation process [28,29].

Also, many problems involve the analysis of rare patterns of occurrences. As an example, Fig. 4 shows the histogram of the PLD series to Center-west/Southeast region. It is possible to see that some pattern occur more often than others. Also, most of the time the price remains at low values, under R$100.00. The energy price rarely reaches high values, above R$300.00. However,

Table 2
Summary statistics of the variables from Center-west/Southeast region.

Variable	Minimum	Mean	Maximum	Standard deviation
Stored energy	20.71	65.30	87.64	16.7
Inflow energy	47.86	103.87	182.00	24.38
Hydro generation	8854.14	17,635.02	23,378.14	3253.84
Thermal Generation	192.86	1112.57	3258.86	615.93
Load	19,295.57	28,048.88	34,668.00	3148.61
PLD	4.0	74.65	684.00	117.53

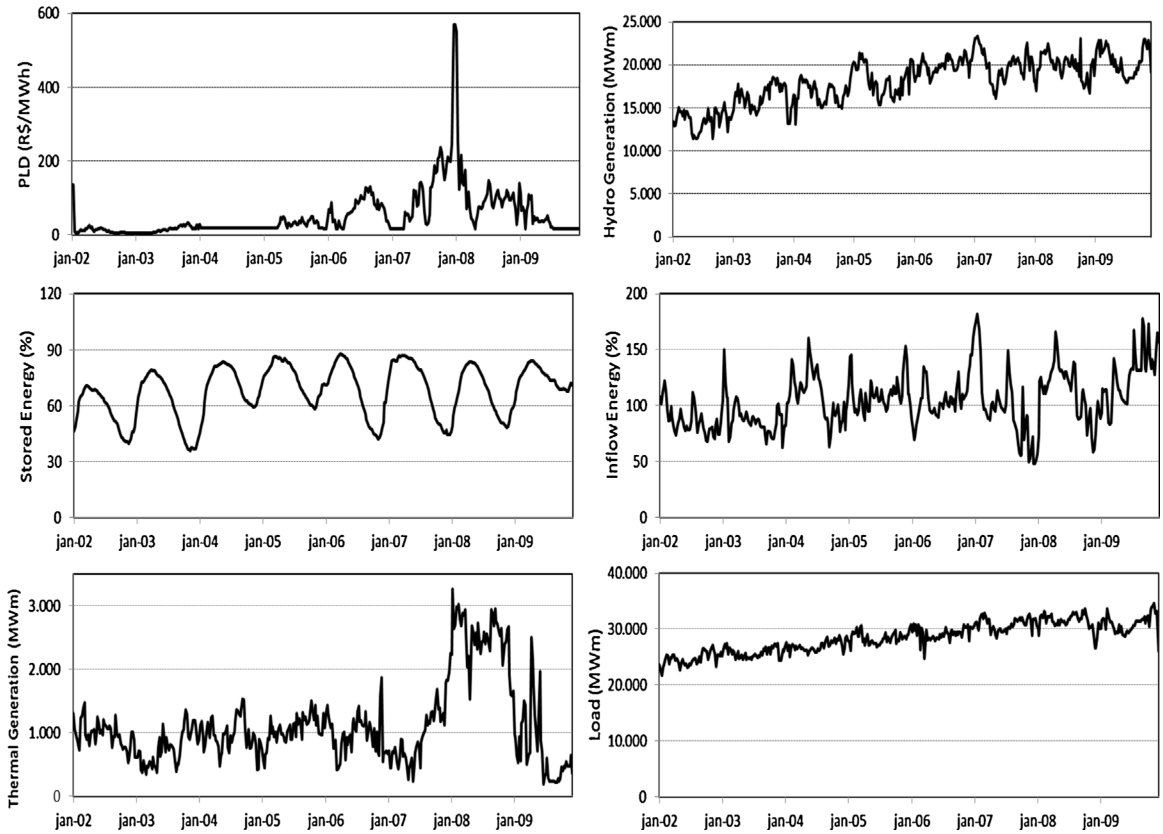

Fig. 3. Time series data to Center-west/Southeast region.

neural networks are sensitive to imbalanced data sets since it causes difficulties in the learning process and can deteriorate the model performance [30]. Then, data balancing was applied in this paper during the data preparation process.

5. Simulation results

The methodology proposed in this paper is applied to the Brazilian electricity market. Some criteria commonly used to evaluate price forecasting accuracy are employed in this paper: root mean square error (RMSE), mean absolute error (MAE) and mean

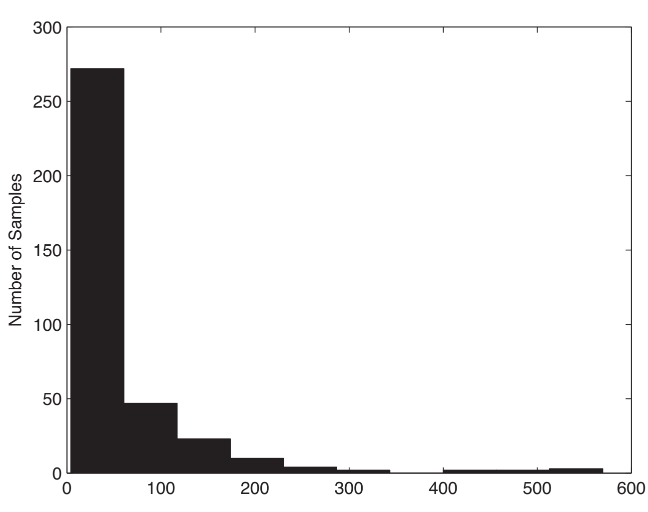

Fig. 4. Histogram of the PLD series to Center-west/Southeast region.

absolute percentage error (MAPE) [31]. These quantities are calculated by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i^{true} - P_i^{forecast})^2} \quad (7)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|P_i^{true} - P_i^{forecast}| \quad (8)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|P_i^{true} - P_i^{forecast}|}{P_{AVE}} \times 100\% \quad (9)$$

$$P_{AVE} = \frac{1}{N}\sum_{i=1}^{N}P_i^{true} \quad (10)$$

where N is the number of samples, P_i^{true} is the actual price and $P_i^{forecast}$ is forecasted price.

Fig. 5 presents energy price error measures obtained with the proposed hybrid model 12-weeks ahead and others traditional methods: ARIMA, Exponential Smooth, GARCH and NN. As the results show, the proposed hybrid model has lower forecasting errors (RMSE, MAE and MAPE) compared to the other methods. This behavior is also observed in all regions. The only exception is the South region, where the hybrid model presented similar MAPE to GARCH model. Also, the proposed hybrid model presented higher linear correlation (LC) and lower standard deviation (SD) to all regions as shown in Table 3.

The results obtained with the proposed hybrid model were compared with some forecast methods available in the literature. In [7], the accuracy obtained in terms of MAE was 2.763 $/MWh

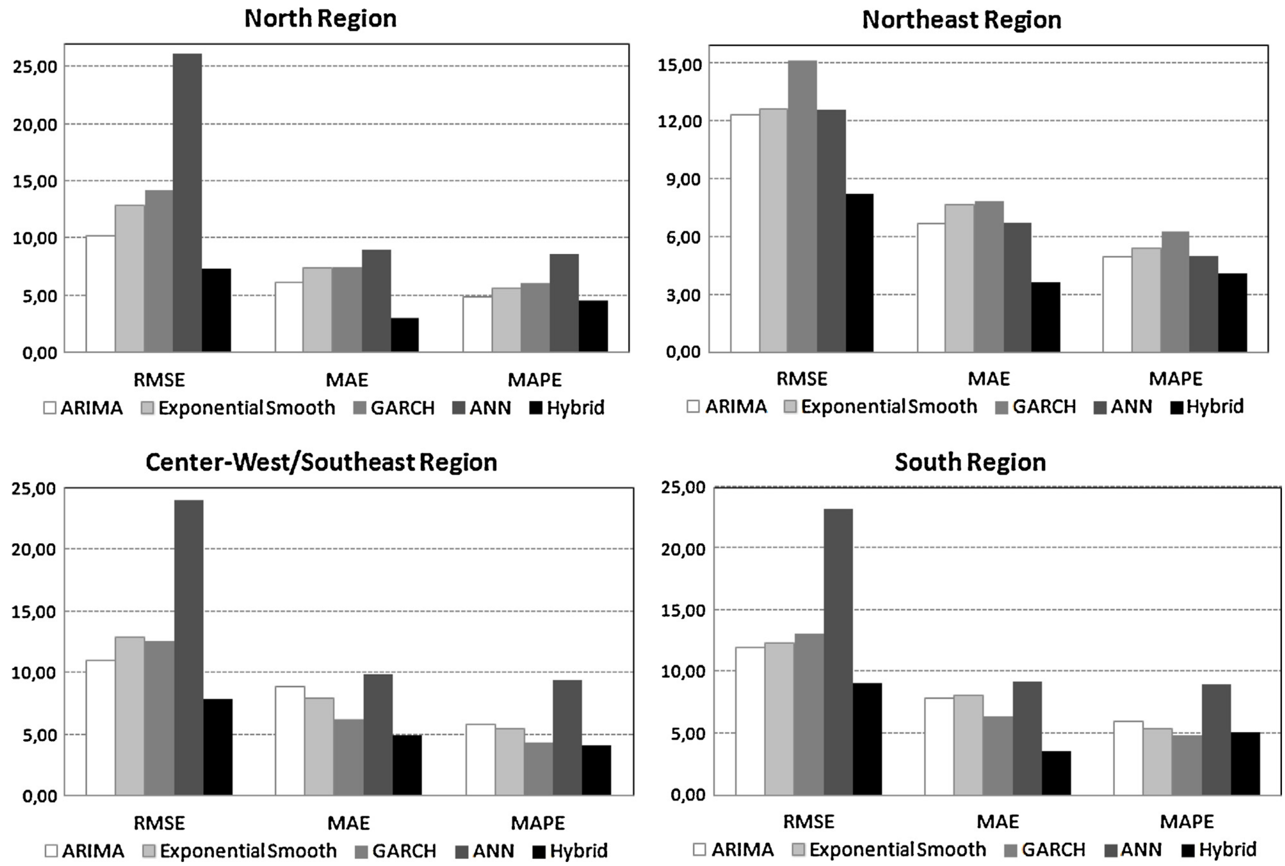

Fig. 5. RMSE, MAE and MAPE to short-term energy price prediction.

Table 3
Linear correlation (LC) and standard deviation (SD) to short-term energy price prediction.

	North		Northeast		Center-west/Southeast		South	
	LC	SD	LC	SD	LC	SD	LC	SD
ARIMA	0.96	33.63	0.90	34.35	0.96	32.08	0.92	25.36
ExpSmooth	0.95	37.00	0.89	37.18	0.96	33.94	0.92	28.50
GARCH	0.75	22.25	0.93	23.67	0.88	18.63	0.70	19.80
NN	0.77	42.50	0.85	22.00	0.81	34.17	0.69	32.05
Hybrid	0.99	14.72	0.97	19.75	0.99	11.30	0.98	14.75

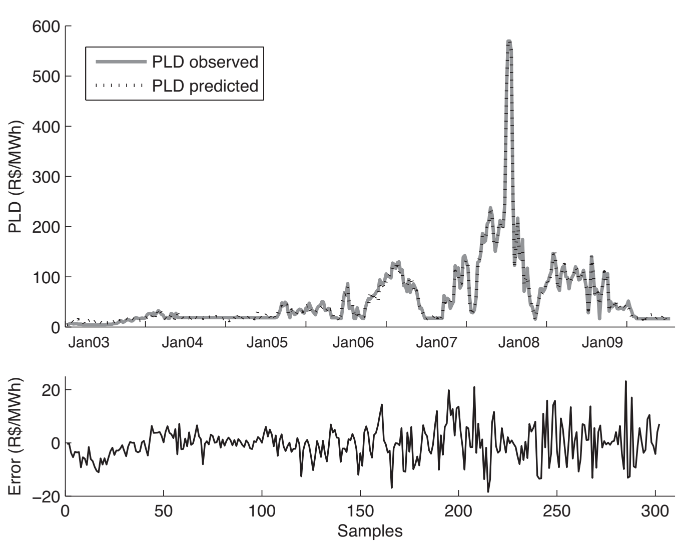

Fig. 6. Energy price observed and predicted and absolute error obtained with the hybrid model to Center-west/Southeast region.

and 7.0989 $/MWh to PJM market. The MAE obtained with our proposed method applied to the four Brazilian submarkets were: 3.04 R$/MWh, 3.64 R$/MWh, 4.95 R$/MWh, and 3.58 R$/MWh. In [6] the weekly error MAPE obtained with the proposed method was from 4.8% $/MWh to 11.3% $/MWh to the Spain market. In [8] the weekly error MAPE obtained with the proposed method was from 4.64% $/MWh to 11.28% $/MWh to the PJM market. The weekly error MAPE obtained in [9] was from 3.41% $/MWh to 5.15% $/MWh to the PJM market. The MAPE obtained with our proposed method applied to the four Brazilian submarkets varies from 4.06% R$/MWh to 5.11% R$/MWh. Then, the maximum MAE and MAPE obtained in our study are comparatively lower than those

Table 4
Results of Kupiec test to 12-weeks ahead.

Region	LR_{POF} statistics				
	ARIMA	GARCH	NN	ExpSmooth	Hybrid
North	15.48	18.39	5.82	7.85	0.298
Northeast	12.74	24.69	4.78	10.19	1.108
Center-west/Southeast	18.39	28.05	3.83	7.85	1.108
South	12.74	15.48	2.97	18.39	0.298

Table 5
MAPE criteria for superior time horizon forecast.

Region	12-weeks ahead			24-weeks ahead			36-weeks ahead		
	ARIMA	NN	Hybrid	ARIMA	NN	Hybrid	ARIMA	NN	Hybrid
North	5879	8201	4481	17,839	20,000	9353	20,431	23,115	16,387
Northeast	11,851	10,759	4122	14,215	19,813	9148	23,930	27,717	15,885
Center-west/Southeast	10,666	11,838	4066	17,363	17,597	9253	28,237	22,694	15,274
South	10,600	11,770	5111	19,032	20,591	11,922	23,400	24,235	17,838

obtained in Refs. [6–9]. Hence, we can conclude that the proposed hybrid model provides reliable price forecasts for 12-weeks ahead in the Brazilian market.

Fig. 6 shows the short-term energy price observed and predicted with the proposed hybrid model to Center-west/Southeast region. The absolute error is also presented in this figure. It is possible to see that the forecast results obtained with the proposed model are quite close to the actual PLD values, including when price spike occurs. Then, the results obtained with the proposed model are sufficiently good.

Since the RMSE, MAE and MAPE criteria cannot capture tail losses, this paper also evaluated the performance of the proposed method using the unconditional coverage tail-loss test developed by Kupiec [32,33]. Kupiec's test, also known as the POF-test (proportion of failures), measures whether the number of exceptions is consistent with the confidence level. The likelihood ratio LR_{POF} statistic is evaluated by:

$$LR_{POF} = -2\ \ln \frac{(1-p)^{T-n} p^n}{[1-n/T]^{T-n}(n/T)^n} \tag{11}$$

where p is the probability of failure, T is the number of observations, n number of exceptions. If the value of the LR_{POF} statistic exceeds the critical value of the χ^2 distribution (chi-squared), the model is deemed as inaccurate.

In this paper the confidence level adopted is 99%, and the chi-squared critical value with one degree of freedom is 6.635. The results are presented in Table 4, and according to the Kupiec test (LR_{POF}) the proposed hybrid technique is the best model. The worst models are GARCH and NN to all region, ARIMA to North and South region, and Exponential Smooth to Center-west/Southeast region.

The behavior of the hybrid proposed model to superior time horizon was also considered in this paper. Table 5 shows MAPE error for 12, 24 and 36 weeks ahead for the proposed hybrid model, ARIMA and NN model for sake of comparison. The results show that the hybrid model presents smaller error to all regions. Also, forecast performance deteriorates as the prediction time horizon increases. It is important to mention that better results were obtained considering a prediction time horizon inferior than 12-weeks ahead. However, the time horizon of 12-weeks ahead is more appropriate to risk management practices in the Brazilian market.

6. Conclusions

In this paper, a hybrid approach combining ARIMA and NN models is proposed for short-term energy price prediction using explanatory variables. First, the model predicts the explanatory variables that affect the energy price. Then, the energy prices are forecasted by using the explanatory variables prediction. The model considers multi-step ahead price prediction (12 weeks-ahead) and is applied to the Brazilian electricity market.

The results obtained with the proposed methodology are compared with traditional techniques like ARIMA, GARCH, Exponential Smooth and NN, and the proposed method outperforms these techniques. Also, the results obtained with the proposed hybrid model applied to the Brazilian market presented accuracy level sufficiently good compared with other prediction methods reported in the literature, which are applied to other markets. Since the error criteria cannot capture tail losses, this paper also evaluated the performance of the proposed method using the unconditional coverage tail-loss test developed by Kupiec. The results show that the proposed hybrid technique is the best model.

Results considering a superior time horizon were also analyzed, and the proposed technique outperforms standards techniques such as NN and ARIMA for 24 and 36 weeks ahead. Then, this technique can be an important tool to help market participants to reduce risk in purchasing and buying power in the Brazilian market.

It is important to mention that the methodology proposed in this paper was developed based on the characteristics of the Brazilian electricity market, and the price forecasting results were obtained only to the Brazilian market. Then, at this moment, it is not possible to assess the degree of generality of the algorithm. As future works the authors intend to investigate the performance of the proposed method in other electricity markets to have a more general approach.

Acknowledgement

This paper was supported by Fapeam.

References

[1] S.K. Aggarwal, L.M. Saini, A. Kumar, Electricity price forecasting in deregulated markets: a review and evaluation, Int. J. Electr. Power Energy Syst. 31 (2009) 13–22.

[2] A.M. Gonzalez, A.M. San Roque, J. Garcia-Gonzalez, Modeling and forecasting electricity prices with input/output hidden Markov models, IEEE Trans. Power Syst. 20 (2005) 13–24.

[3] X. Guan, D. Pepyne, Y.C. Ho, Gaming and price spikes in electric power markets, IEEE Trans. Power Syst. 16 (2001) 402–408.

[4] Y. Fu, Z. Li, Different models and properties on LMP calculations, in: IEEE Power Eng. Soc. General Meeting Proc., Montreal, 2006.

[5] D. Singhal, K.S. Swarup, Electricity price forecasting using artificial neural networks, Int. J. Electr. Power Energy Syst. 33 (2011) 550–555.

[6] A.J. Conejo, M.A. Plazas, R. Espínola, A.B. Molina, Day-ahead electricity price forecasting using the wavelet transform and ARIMA models, IEEE Trans. Power Syst. 20 (2005) 1035–1042.

[7] X. Yan, N.A. Chowdhury, Mid-term electricity market clearing price forecasting: a hybrid LSSVM and ARMAX approach, Electr. Power Energy Syst. 53 (2013) 20–26.

[8] L. Wu, M. Shahidehpour, A hybrid model for day-ahead price forecasting, IEEE Trans. Power Syst. 25 (2010) 1519–1530.

[9] N. Amjady, F. Keynia, Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method, Int. J. Electr. Power Energy Syst. 30 (2008) 533–546.

[10] H. Zareipour, A. Janjani, H. Leung, A. Motamedi, A. Schellenberg, Classification of future electricity market prices, IEEE Trans. Power Syst. 26 (2011) 165–173.

[11] A.R. Queiroz, F.A. Oliveira, J.W.M. Lima, P.P. Balestrassi, Simulating electricity spot prices in Brazil using neural network and design of experiments, in: IEEE PowerTech Conf. Proc., Lausanne, 2007, pp. 2029–2034.

[12] J.C.R. Filho, C.M. Affonso, R.C.L. Oliveira, Pricing analysis in the Brazilian energy market: a decision tree approach, in: IEEE PowerTech Conf. Proc., Bucharest, 2009, pp. 1–6.

[13] E. Melo, E.M.A. Neves, L.H.A. Pazzini, The Brazilian electricity model: an overview of the current structure and market design, in: 8th European Energy Market International Conference Proc., Croatia, 2011, pp. 1–6.

[14] S. Granville, G.C. Oliveira, L.M. Thomé, N. Campodónico, M.L. Latorre, M. Pereira, L.A. Barroso, Stochastic optimization of transmission constrained and large scale hydrothermal systems in a competitive framework, in: IEEE General Meeting Conf. Proc., vol. 2, Toronto, 2003, pp. 1101–1106.

[15] B.S. Everitt, Cambridge Dictionary of Statistics, 3rd ed., Cambridge University Press, 2006.

[16] I.E. Arciniegas Rueda, A. Marathe, Important variables in explaining real-time peak price in the independent power market of Ontario, Util. Policy 13 (2005) 27–39.
[17] R. May, G. Dandy, H. Maier, in: K. Suzuki (Ed.), Review of Input Variable Selection Methods, Artificial Neural Networks – Methodological Advances and Biomedical Applications, vol. 219, InTech, 2001, p. 234.
[18] P. Mandal, T. Senjyu, T. Funabashi, Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market, Energy Convers. Manag. 47 (2006) 2128–2142.
[19] I. Lind, L. Ljung, Regressor and structure selection in narx models using a structured ANOVA approach, Automatica 44 (2008) 383–395.
[20] S. Sweet, K. Grace-Martin, Data Analysis with SPSS: A First Course in Applied Statistics, 4th ed., Pearson Publishing as Allyn & Bacon, Boston, MA, 2012.
[21] SPSS, Data Mining, Statistical Analysis Software Predictive Analytics, Decision Support Systems, http://www.spss.com (accessed 06.12.13).
[22] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis, Forecasting and Control, Holden-Day, New Jersey, 1976.
[23] S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, New York, 1994.
[24] R. Rojas, Neural Networks – A Systematic Introduction, Springer-Verlag, Berlin/New York, 1996.
[25] L.V. Fausett, Fundamentals of Neural Networks: Architectures, Algorithms, and Applications, Prentice-Hall, Englewood Cliffs, NJ, 1994, pp. 461.
[26] Website of Brazilian Electrical Energy Commercialization Chamber. http://www.ccee.org.br/
[27] Website of Brazilian National System Operator. http://www.ons.org.br/home/
[28] R. Nisbet, J. Elder, G. Miner, Handbook of Statistical Analysis & Data Mining Applications, Academic Press, Elsevier, 2009.
[29] H. Abdi, L.J. Williams, Principal component analysis, Wiley Interdiscip. Rev.: Comput. Stat. 2 (2010) 433–459.
[30] D.L. Olson, Data set balancing, in: Data Mining and Knowledge Management, Lecture Notes in Computer Science, vol. 3327, 2005, pp. 71–80.
[31] P. Mandal, A. Ul Haque, J. Meng, A.K. Srivastava, R. Martinez, A novel hybrid approach using wavelet, firefly algorithm, and fuzzy ARTMAP for day-ahead electricity price forecasting, IEEE Trans. Power Syst. 28 (2013) 1041–1051.
[32] P.H. Kupiec, Techniques for verifying the accuracy of risk measurement models, J. Deriv. 3 (2) (1995) 73–84.
[33] A. Cifter, Forecasting electricity price volatility with the Markov-switching GARCH model: evidence from the Nordic electric power market, Electr. Power Syst. Res. 102 (2013) 61–67.