

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Short-Term Energy Price Prediction Multi-Step-Ahead in the Brazilian Market Using Data Mining

José C. Reston Filho, Carolina de M. Affonso and Roberto Célio L. de Oliveira

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/48472>

1. Introduction

The electricity price tend to be very volatile due to weather conditions, fuel price, economic growth and many others factors [1]. As a consequence, electricity markets participants face high risks in bilateral contracts and short-term market. With regard to short-term market, generators sell energy at variable pool prices while their fuel cost are fixed. Also, distributors supply energy to most of their costumers at an annual fixed tariff, but they have to purchase electricity at a variable pool price. Then, a reliable tool to forecast electricity price is absolutely crucial for risk management in energy markets.

Many papers have proposed hybrid models to energy price prediction. The benefit of the hybrid model is to combine strengths of the techniques providing a robust model capable of capturing the nonlinear nature of the complex time series, producing more accurate forecasts. Reference [2] provides a hybrid methodology that combines both ARIMA and Artificial Neural Network (ANN) models for predicting short-term electricity prices. In [3], a novel technique to forecast day-ahead electricity prices is presented based on Self-Organizing Map neural network (SOM) and Support Vector Machine (SVM) models. Reference [4] proposes a novel price forecasting method based on wavelet transform combined with ARIMA and GARCH model.

The major data mining functions that are developed in research communities include summarization, association, prediction and clustering. This work deals with the energy price prediction problem multi-step-ahead in the Brazilian market. The ARIMA model is used to predict the variables that affect the short-term energy price (exogenous input), instead of predicting the energy price directly as in [2]. The results obtained with the

methodology proposed are compared with the traditional ARIMA techniques. The historical data are from January 2006 to December 2009.

Some papers have already proposed the use of exogenous input to predict the energy price [5,6]. However, no work has been reported so far with energy price prediction models for the Brazilian market. Regarding the Brazilian market, most of the papers deal with risk analysis, optimal contract portfolio, and load prediction [7,8,9,10]. The main contribution of this chapter lies in the application of energy price forecasting methodologies applied to the Brazilian market, which adopts the tight pool model with unique characteristics of energy price behavior.

Another contribution of this study is to consider price spikes in the data base, and treat them equally as the normal prices. A price that is much higher than the normal price is usually considered as price spike. Most energy price forecast methods remove price spikes as noise and deal only with the normal prices, or build two different prediction models separately for both normal prices and spikes [5,11,12].

The next sections of this chapter are organized as follows. Section 2 describes the main features and peculiarities of the Brazilian electricity market. The proposed methodology and important aspects of data-preparation are addressed in Section 3. Section 4 presents the results and Section 5 presents the main conclusions.

2. The Brazilian System

The Brazilian System system has an installed capacity of 91GW where 82% corresponds to hydro generation, 15.2% to thermal generation, 2.19% to nuclear power and only 0.64% corresponds to biomass and wind generation [13].

The hydro system is characterized by large reservoirs with multi-year regulation capacity, arranged in complex cascades over several river basins.

Brazil still has an undeveloped hydro potential of 145,000 MW. Then, it is expected that the system remains predominantly hydro in the future.

The country is fully interconnected by a 80,000 km meshed grid, with voltages levels from 230 kV to 765 kV ac, plus two 600 kV dc links connecting the binational Itaipu hydro power plant to the main grid.

The National System Operator (ONS) is responsible to operate, supervise and control power generation and transmission grid in the Brazilian system. The Electrical Energy Commercialization Chamber (CCEE) is the body responsible for energy market transactions, such bilateral and short-term market contracts.

The Brazilian National Interconnected System (SIN) has four geoelectric submarkets organized by regions: North, Northeast, South and Center-west/Southeast. These markets can import/export energy from/to each other.

2.1. Tight Pool model

In general, there are two types of power pool arrangements: loose pool and tight pool. In the loose pool model there is no common dispatch center and each company in the group has its own dispatch center. The generation dispatch is carried out through auctions where generators and demand agents bid for price and quantities. Then, agents are paid at the same price, the market-clearing price, defined by the equilibrium between supply and demand.

In a tight pool model the generation dispatch is centralized by an Independent System Operator (ISO) in order to maximize the energy production by the system as a whole. The Brazilian system adopts the tight pool model with the dispatch centralized by the ONS due to the predominance of the hydro generation. This model is adopted to make efficient use of hydroelectric reservoir. The water is stored during the “wet” years (favorable inflow energy) in order to increase energy production in the dry years, reducing the generation from thermal power plants.

In the Brazilian model, hydro plants are dispatched with basis on their expected opportunity costs (“water values”) calculated by a multi-stage stochastic optimization model (Stochastic Dual Dynamic Programming - SDDP) considering several inflow scenarios with uncertainties [14].

The SDDP model minimizes the marginal price of the system operation considering the immediate benefit of using water in the reservoirs (immediate cost) and the future benefits of its storage (future cost), measured in terms of the economy expected by the use of fuel in thermal units [15]. Then, the spot price used in the short-term market is not calculated from the equilibrium between demand and supply, but from the Lagrange multipliers of the stochastic dispatch model instead.

2.2. Short-term spot market

Wholesale energy markets have a structure organized by a long term market (forward or bilateral contracts) and a spot market. In long term contracts sellers and buyers of energy negotiate freely the terms of volume, price and duration. A spot market represents the short term 24-hour look-ahead market condition in which prices and generation dispatch are defined.

Since generators and loads does not bid prices in the Brazilian market, the market settlement is an accounting procedure controlled by CCEE, given by the difference between the energy produced and the energy volumes registered in financial forward contracts.

Positive or negative differences are settled in the spot market through the spot price, which is named PLD (Settlement Price for the Differences). Figure 1 illustrates this commercialization process. The PLD price is determined by a stochastic optimization model and is limited by a minimum and maximum price. It is computed on a weekly basis for each

load level (low, medium and high) in each submarket (North, Northeast, Center-west/Southeast and South).

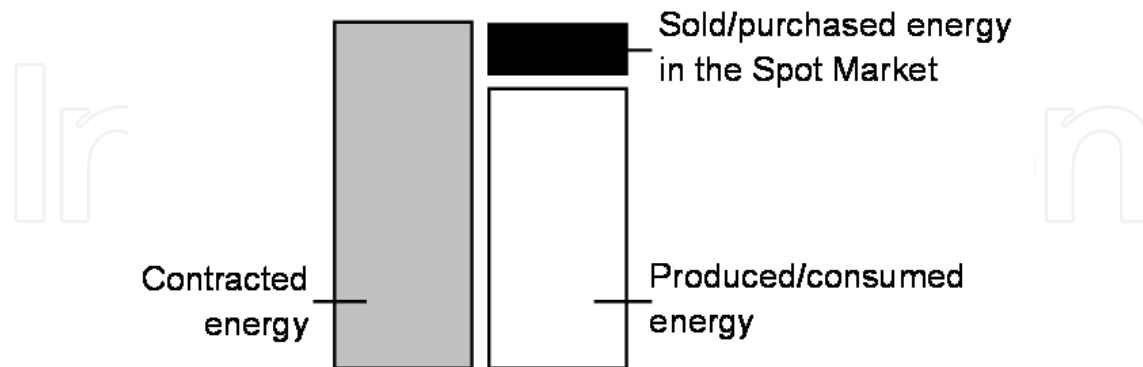


Figure 1. Commercialization process in the Spot Market [17].

2.3. Exogenous input

In general, forecasting loads and prices in the wholesale markets are mutually intertwined activities since the main variable that drives the price is the power demand [16]. For this reason, the demand has been the most commonly examined explanatory variable in price forecasting studies. However, the Brazilian market is a tight pool model with no price bids from producers and consumers. The Brazilian short-term energy price (PLD) is obtained from optimization models. Then the demand does not respond to energy price variations.

On the other side, the Brazilian short-term energy price is strongly dependent on the water level and inflow energy in reservoirs of the hydropower plants, since hydroelectric generation is predominant in Brazil. The result is that Brazilian short-term energy price is very volatile and dependent on the system's hydrological conditions.

Table 1 shows the exogenous input used in this study to forecast the short-term electricity price to the Brazilian market. These inputs are selected based on the methodology used by the Brazilian Independent System Operator. According to then, the most important variables involved in the computation of the PLD are variables related to hydrological conditions, system power load and fuel prices of thermal units [17].

Exogenous Input	Definition	Unit
HyGen	Total hydro generation	MWmed
TherGen	Total thermal generation	MWmed
Load	System power load	MWmed
StEn	Stored energy in reservoirs	% MLT*
InEn	Inflow energy in reservoirs	% MLT*

Table 1. Exogenous Input (*MLT: long-term average - historical average of 79 years)

Figure 2 shows linear correlation graphs of the input attributes from the proposed hybrid model to South region. It is possible to note that the variables do not exhibit a linear relation between them, which justifies its use in the prediction with hybrid model. The same behavior was observed to the other regions.

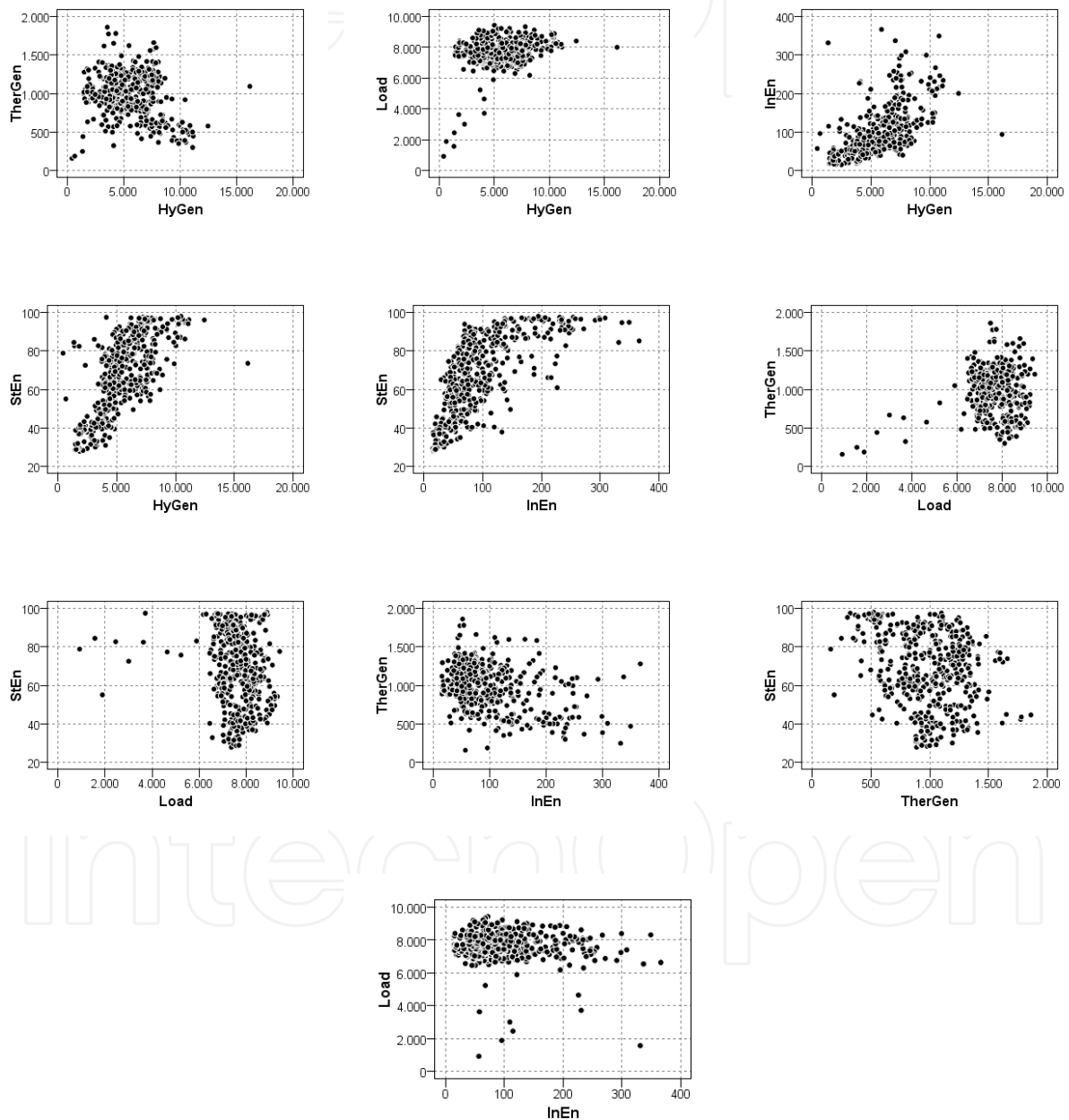


Figure 2. Linear correlation graphs of the input attributes from the proposed hybrid model to South region.

Figure 3 shows the behavior of PLD, storage energy and inflow energy from May 2006 through April 2007 to the Center-west/Southeast region.

This region is characterized by two distinct seasons: dry, which falls between May and November, and wet, which falls between December and April.

During the dry season the inflow energy is lower, and tends to increase during the wet season. Stored energy also presents this behavior with a delayed relationship. Also, PLD tends to be higher during the dry season due to the use of the thermal power plants.

The strong relationship between these variables reinforces the need to use them as exogenous input to forecast the short-term electricity price to the Brazilian market.

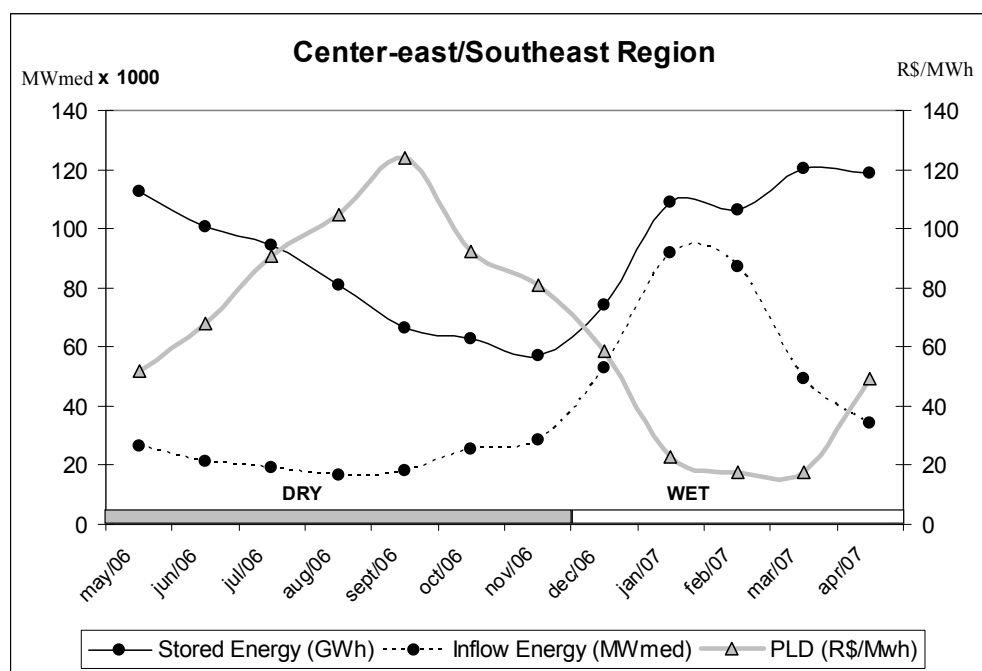


Figure 3. Relationship between PLD, stored energy and inflow energy.

3. Proposed hybrid model

The behavior of the short-term energy price may not be easily captured by stand-alone models since time series data may include a variety of characteristics such as seasonality and heteroskedasticity. A hybrid model having both linear and nonlinear modeling abilities could be a good alternative for predicting energy price data. Figure 4 shows the flowchart of the proposed hybrid model. The main steps of the algorithm are presented below and the details will be discussed next. This study uses the data mining software SPSS Clementine to develop and test the proposed methodology [18].

Step 1. Create a large data base composed by historical data of the short-term energy price and the attributes that affect the short-term energy pricing – exogenous input (U_i): stored energy, inflow energy, hydro generation, thermal generation and power load. These are the exogenous input.

- Step 2.** Forecast the linear relationship of the exogenous input (U_i) 12-steps-ahead (12 weeks ahead) with the ARIMA model.
- Step 3.** Apply PCA and Balacing in data preparation process to reduce the dimension of the input vectors and choose the better learning set before training the first ANN.
- Step 4.** Forecast the non-linear relationship from the exogenous input (U_i) 12-steps-ahead (12 weeks ahead) with the ANN model.
- Step 5.** Apply again the PCA and Balacing in data preparation process to the second ANN.
- Step 6.** Forecast the short-term energy price 12-steps-ahead (12 weeks ahead) with the ANN model.

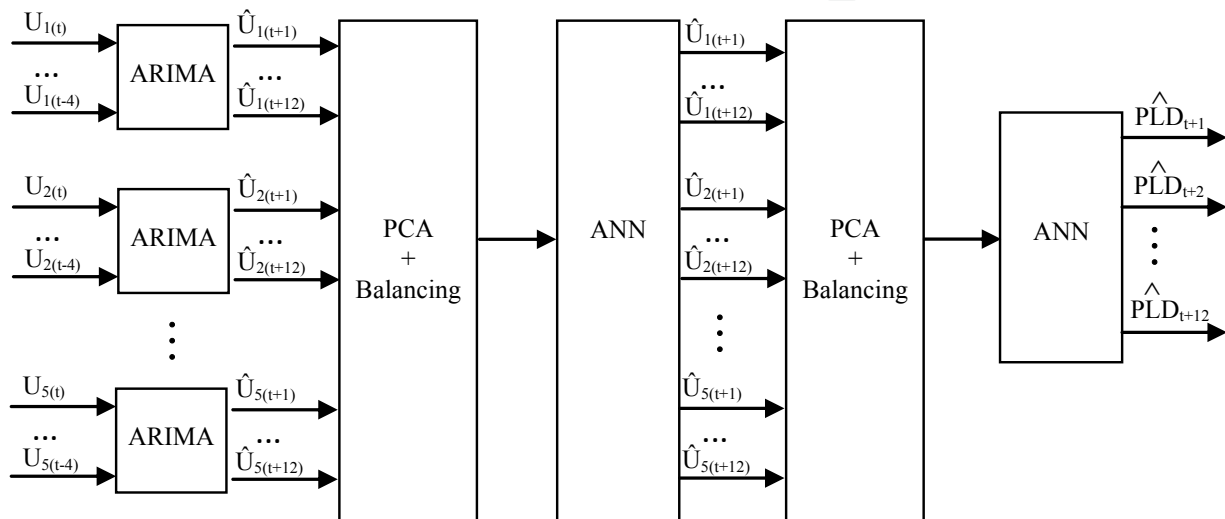


Figure 4. Flowchart of the proposed hybrid model.

3.1. Autoregressive Integrated Moving Average (ARIMA)

The ARIMA (autoregressive integrated moving average) model predicts a value in a response time series as a linear combination of its own past values, past errors, and current and past values of other time series [19].

An ARIMA model, usually referred as ARIMA (p, d, q), can be described by equation (1):

$$\phi P(q-p)\Delta dy(k) = \theta_0 + \theta q(q-q)V(k) \quad (1)$$

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. Δd is the differencing operator given by:

$$D = 1 - q - 1 \quad (2)$$

The ARIMA modeling approach involves the following four steps: model identification, parameter estimation, diagnostic checking and forecast future outcomes based on the known data. Identification of the general form of a model includes appropriate differencing of the series to achieve stationary and normality. Then, the temporal correlation structure of

the transformed data is identified by examining its autocorrelation (ACF) and partial autocorrelation (PACF) functions [20].

The second step is the estimation, which can be done using an iterative procedure to minimize the prediction error, such as least square method [21].

The third step is the diagnostic checking to investigate the adequacy of the model. Tests for white noise residuals (uncorrelated and normally distributed around a zero mean) indicate whether the residual series contains additional information that might be utilized by a more complex model. The last step is forecast future outcomes based on the known data.

3.2. Artificial Neural Networks (ANN)

Neural networks are flexible computing frameworks for modeling a broad range of nonlinear problems [22]. It can be considered a black box that is able to predict an output pattern when it recognizes a given input pattern. ANN makes no prior assumptions concerning the data distribution. A neural network can be trained by the historical data of a time series in order to capture the characteristics of this time series. The model parameters (connection weights and node biases) will be adjusted iteratively by a process of minimizing the forecast errors. The algorithm used in this work is the Backpropagation algorithm and the ANN architecture is the multilayer perceptron (multilayer feed-forward network). This neural network is widely used and consists of an input layer, hidden layers and an output layer of neurons.

3.3. Data preparation

The historical data used to create the database are available on Brazilian Electrical Energy Commercialization Chamber website [17] and in the National System Operator website [13]. The exogenous input data were on a daily basis and the PLD data was on a weekly basis. Thus, all data was first standardized to a weekly basis, consuming a lot of effort. Then, a large database was constructed for each one of the four submarkets (North, Northeast, South and Center-west/Southeast) for the period from April 2001 to December 2009.

The variables used to create the database to each submarket are the PLD as the goal attribute, and the exogenous input as input variables (total hydro and thermal generation, system power load, stored energy and inflow energy in reservoirs). Figure 5 shows the time series data to the South region.

In data mining, the data set is usually cleaned before applying forecasting algorithms. It means that price spikes (outliers) are usually removed as noise to avoid very large forecasting errors introduced by the outliers. In this work, we chose not to eliminate any noise or discrepant samples. The decision was to create an estimation model also capable to map price spikes since they have significant impact on the electricity market. The idea is that the prediction model can be used in a risk analysis tool, where the exact value of the energy price is not as important as its range of variation.

The Principal Component Analysis (PCA) technique is applied to reduce the dimension of the input vectors eliminating data highly correlated (redundant) [23]. In addition, analysis of rare events is performed since some pattern occurs less often than others. As an example, Figure 6 shows the histogram of the PLD series. Most price scenarios are very low and only a few are high. However, models built with neural networks algorithms are very sensitive to imbalanced data sets. Then, balancing data sets is necessary to equilibrate the bias in the learning process [24].

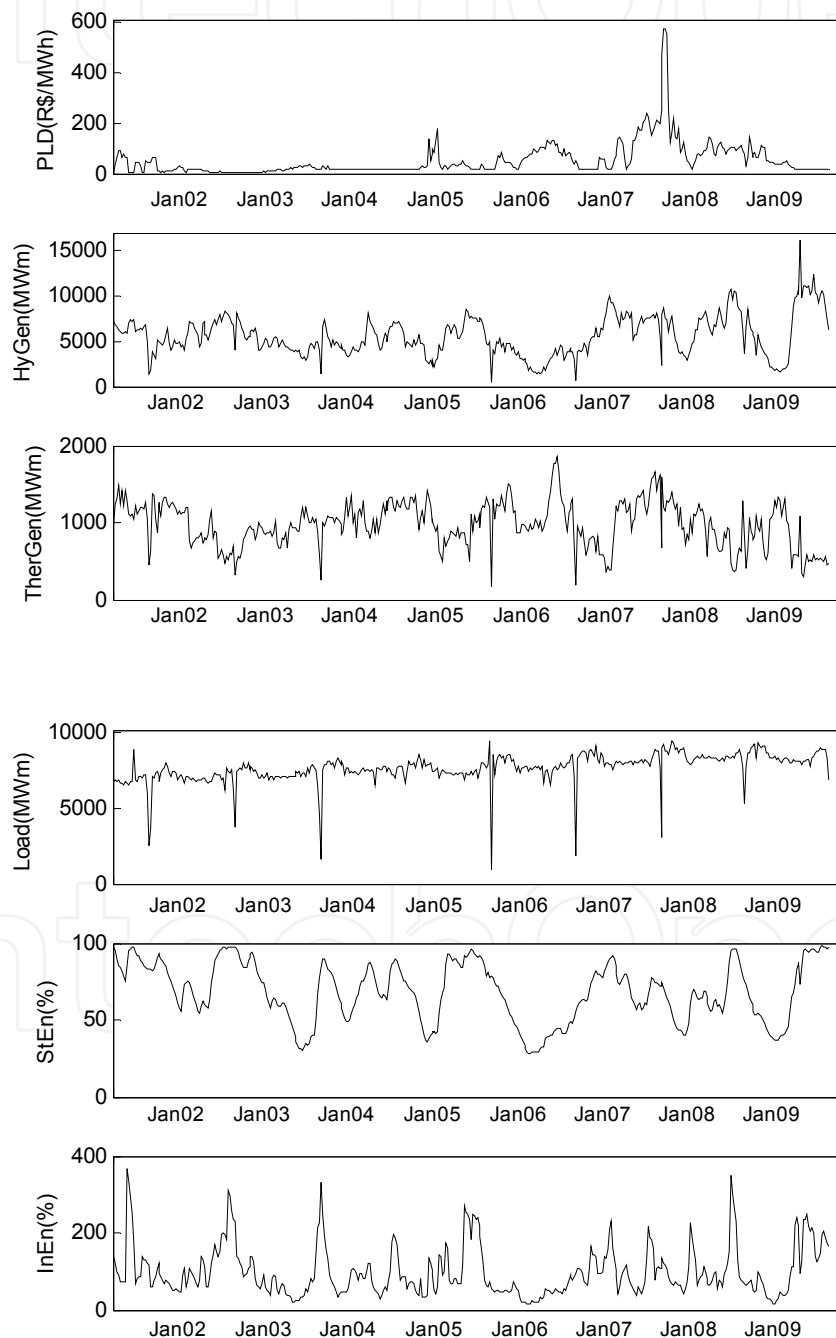


Figure 5. Time series data to South region.

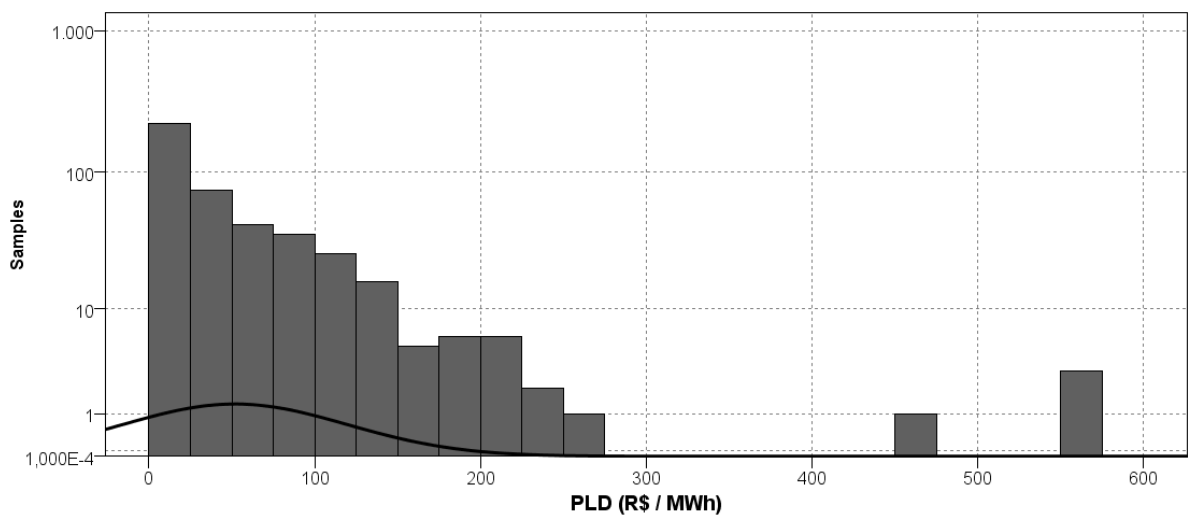


Figure 6. Histogram of the PLD series to South region.

4. Results

The proposed hybrid model was applied to the Brazilian electricity market. Several tests were made to identify the neural network architecture that produces best generalization accuracy for each attribute (short-term energy price PLD, hydro generation, thermal generation, power load, stored energy and inflow energy in reservoirs) varying the number of hidden layers and number of neurons. Best results were obtained with the ANN configuration presented in Table 2, using hyper tangent function in all layers. The same architecture is used to predict the PLD to all regions (North, Northeast, South and Center-west/Southeast). The data set is divided considering the training set with 80% of the data and the test set with the remaining 20% of the data.

Layers	Number of Neurons
Input layer	5
Hidden layer 1	20
Hidden layer 2	15
Hidden layer 3	10
Output layer	1

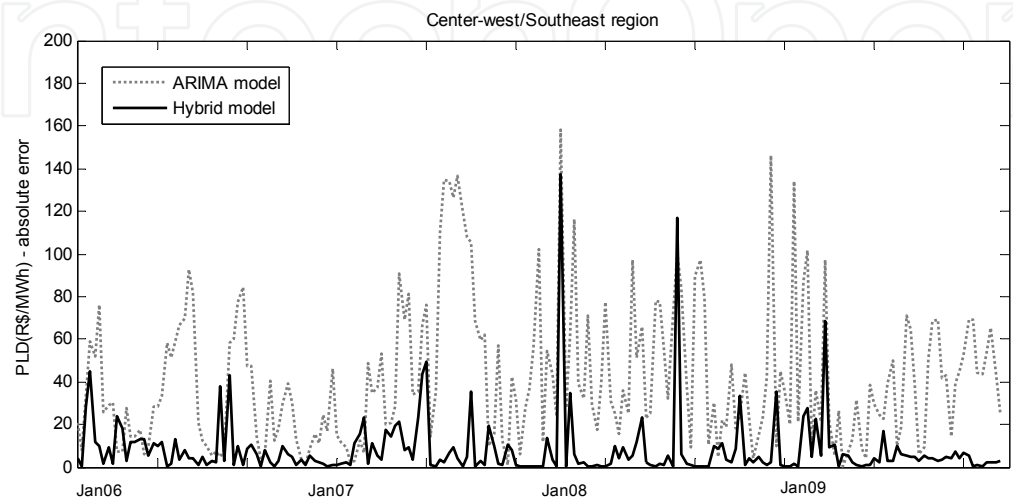
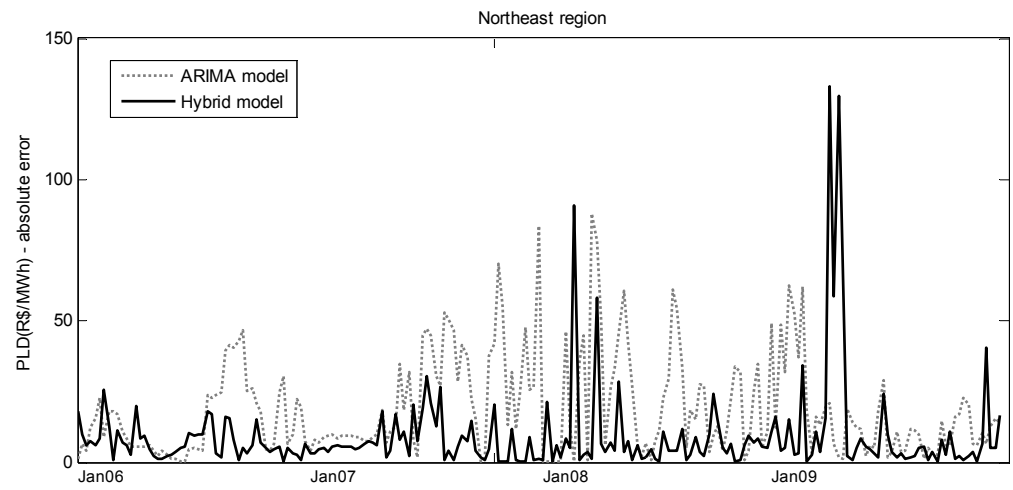
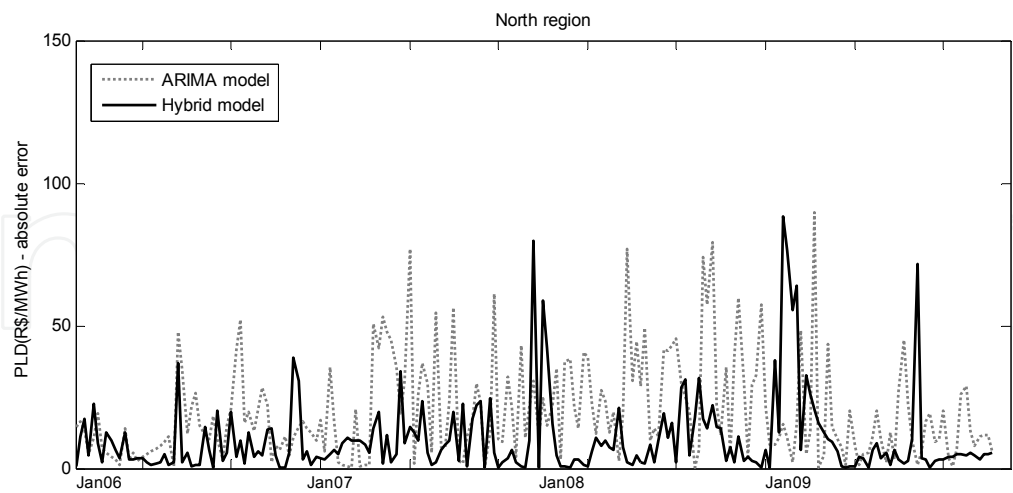
Table 2. Artificial Neural Networks Architecture.

The results are compared with the ARIMA traditional techniques. Some accuracy measures commonly used are employed in this study to analyze the results: the mean square error (MSE), standard deviation and linear correlation. Table 3 gives a statistical comparison of the short-term energy price prediction obtained from the ARIMA and the hybrid model for each region with both training and test set. The hybrid model provides better results for both training and test set, with lower error, lower standard deviation and higher linear correlation.

North region		
	ARIMA	Hybrid
Mean Square Error	18.78	10.589
Standard Deviation	22.37	15.719
Linear Correlation	0.894	0.987
Northeast region		
	ARIMA	Hybrid
Mean Square Error	13.12	10.587
Standard Deviation	23.39	21.231
Linear Correlation	0.915	0.964
Center-west / Southeast region		
	ARIMA	Hybrid
Mean Square Error	11.33	8.292
Standard Deviation	22.65	11.305
Linear Correlation	0.942	0.995
South region		
	ARIMA	Hybrid
Mean Square Error	13.308	11.048
Standard Deviation	26.489	14.754
Linear Correlation	0.935	0.98

Table 3. Performance of the proposed hybrid model.

Figure 7 shows the absolute error obtained for the ARIMA model and the hybrid model to regions North, Northeast, Center-west/Southeast and South. The hybrid model presented superior results. It is important to mention that better results were obtained applying the proposed methodology to predict energy prices with less steps-ahead. However, for practical issues related to the Brazilian market design, which has unique features, the prediction to 12-steps ahead (12 weeks) is more suitable to risk management practices.



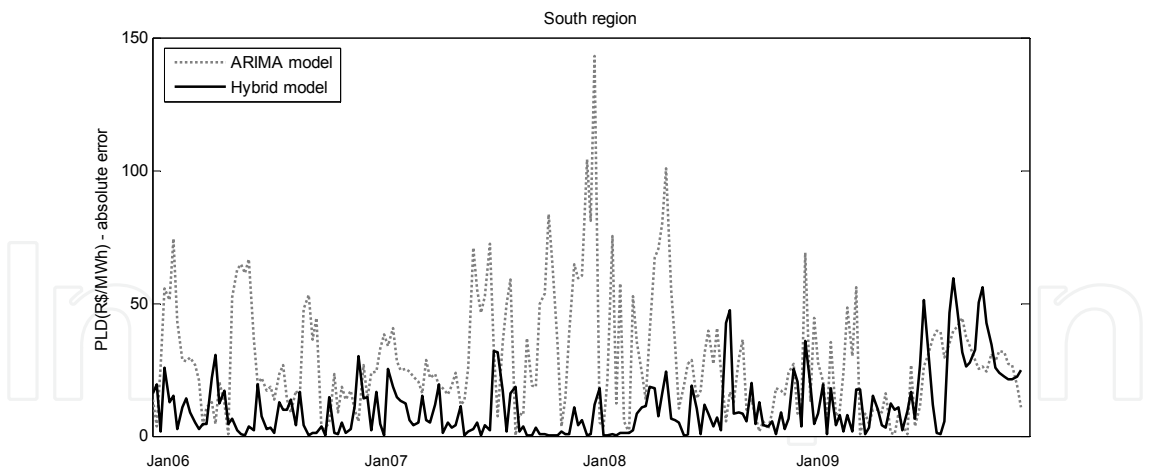


Figure 7. Absolute error obtained for ARIMA and Hybrid model.

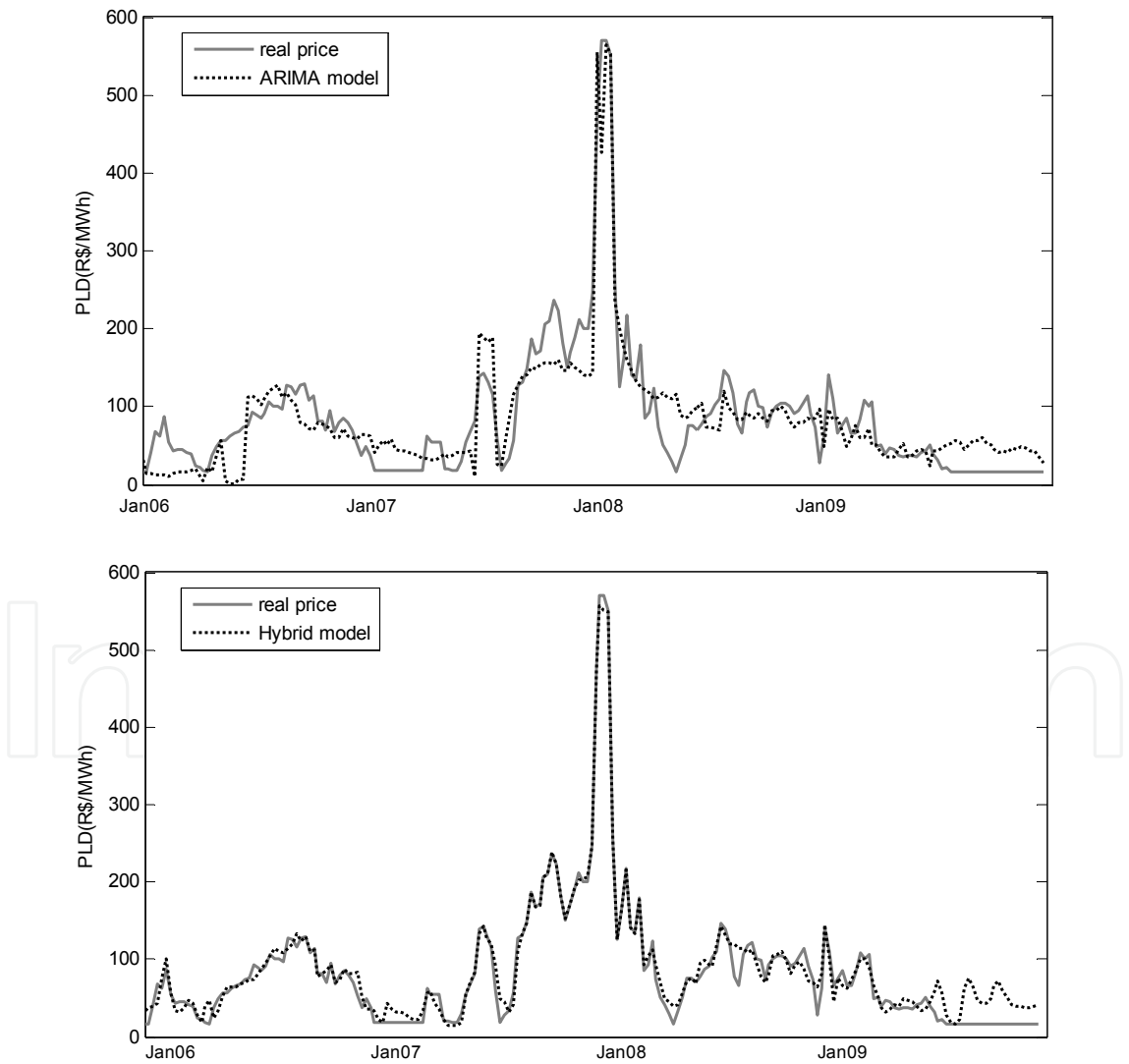


Figure 8. Short-term energy price observed and predicted by the ARIMA and hybrid model to South region.

Figure 8 shows the observed and predicted short-term energy price (PLD) obtained from the ARIMA and hybrid model to South region. The results show that the hybrid model produces better predictions than the ARIMA model. Also, the proposed hybrid model has a strong ability of predicting spikes. Note that this accuracy is obtained with serious insufficient data containing spikes, and spikes are caused by many stochastic events that cannot be entirely considered in the model. Furthermore, the price prediction is being made to 12-steps ahead (12 weeks), which represents a considerable time. For these reasons, we can say that the results are sufficiently good.

5. Conclusions

In this chapter, a hybrid model combining ARIMA and ANN with exogenous input is proposed for short-term energy price prediction in the Brazilian market. The ARIMA model is used to predict the variables that affect the short-term energy price (exogenous input), instead of predicting the energy price directly. This methodology is encouraged by the way energy price is computed in the Brazilian market by National System Operator. The exogenous input are: stored energy, inflow energy, hydro generation, thermal generation and power load. These are the most important attributes involved in the computation of the short-run marginal cost.

After the time series of the exogenous input are predicted, a second ANN is used to forecast the energy price multi-step ahead (12-weeks-ahead). In order to guarantee ANN generalization capacity, a data preparation process is first applied, which includes Principal Component Analysis (PCA) and balancing (analysis of rare patterns of occurrences). Software SPSS Clementine was used to develop and test the proposed methodology. The results obtained with the proposed methodology are compared with the ARIMA traditional techniques.

The results show that the proposed hybrid method performs the short-term electricity price prediction 12- steps ahead with high accuracy. This work provides a valuable contribution for price forecasting in the Brazilian market that can help market participants in their risk management practices.

Author details

José C. Reston Filho
Federal University of Pará (UFPA), Belém, Pará, Brasil

Carolina de M. Affonso
Institute of Technology - ITEC, Federal University of Pará (UFPA), Belém, Pará, Brasil

Roberto Célio L. de Oliveira
Institute of Technology - ITEC, Federal University of Pará (UFPA), Belém, Pará, Brasil

Acknowledgement

This work was supported in part by FAPEAM – AM, Brazil.

6. References

- [1] Shahidehpour M, Alomoush M (2001) *Restructured Electrical Power Systems: Operation, Trading, and Volatility*. New York: Marcel Dekker. 489p.
- [2] Areekul P, Senjyu T, Toyama H, Yona A (2010) A Hybrid ARIMA and Neural Network Model for Short-Term Price Forecasting in Deregulated Market. *IEEE Transactions on Power Systems*. 25:524–530.
- [3] Niua D, Liua Da, Dash Wub D (201) A soft computing system for day-ahead electricity price forecasting. *Applied Soft Computing*. 10: 868–875.
- [4] Tan Z, Zhang J, Wangb J, Xu J (2010) Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. *Applied Energy*. 87: 3606–3610.
- [5] Conejo A J, Plazas M A, Espinola R, Molina A B (2005) Day-ahead electricity price forecasting using wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*. 20: 1035-1042.
- [6] Amjady N, Keynia F (2008) Day ahead price forecasting of electricity markets by a miked data model and hybrid forecast method. *International Journal of Electrical Power Energy Systems*. 9: 533-546.
- [7] Marzano L G B, Melo A C G, Souza R C (2003) An Approach for Portfolio Optimization of Energy Contracts in the Brazilian Electric Sector. *Proc. 2003 IEEE Bologna Powertech Conf. Bologna-Italy*.
- [8] Barroso L A, Rosenblatt J, Guimarães A, Bezerra B, Pereira M V (2006) Auctions of Contracts and Energy Call Options to Ensure Supply Adequacy in the Second Stage of the Brazilian Power Sector Reform. *Proc. 2006 IEEE Power Engineering Society General Meeting Conf. Montreal- Canada*.
- [9] Soares L J, Medeiros M C (2008) Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. *International Journal of Forecasting*. 24: 630–644.
- [10] Leme R C, Turrioni J B, Balestrassi P P, Zambroni de Souza A C, Santos P S (2008) A Study of Electricity Price Volatility for the Brazilian Energy Market. *5th International Conference on European Electricity Market – EEM. Lisboa-Portugal*.
- [11] Amjady N, Keynia F (2011) A new prediction strategy for price spike forecasting of day-head electricity markets. *Applied Soft Computing*. 11: 4246-4256.
- [12] Lu X, Dong Z I, Li X (2005) Electricity market price spike forecast with data mining techniques. *Electric Power Systems Research*. 73:19–29.
- [13] National System Operator (ONS) [Online]. Available: <http://www.ons.com.br>
- [14] Granville S, Oliveira G C, Thomé L M, Campodónico N, Latorre M L, Pereira M, Barroso L A (2003) Stochastic Optimization of Transmission Constrained and Large Scale Hydrothermal Systems in a Competitive Framework. in *Proc. 2003 IEEE General Meeting. Toronto-Canada*.
- [15] Lino P, Barroso L A, Fampa M, Pereira M V, Kelman R (2003) Bid-Based Dispatch of Hydrothermal Systems in Competitive Markets. *Annals of Operations Research*. 120: 81-97.

- [16] Mandal P, Senjyu T, Funabashi T (2006) Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy Conversion and Management*. 47: 2128-2142.
- [17] Brazilian Electrical Energy Commercialization Chamber (CCEE) [Online]. Available: <http://www.ccee.org.br>
- [18] SPSS, Data Mining, Statistical Analysis Software. [Online]. Available: <http://www.spss.com>
- [19] Box G E P, Jenkins G M, Reinsel G C (1976) *Time Series Analysis, Forecasting and Control*. New Jersey: Holden-Day. 784p.
- [20] Mishra A K, Desai V R (2005) Drought forecasting using stochastic models. *Stochastic Environmental Research Risk Assessment*. 19: 326–339.
- [21] Morettin P A, Toloi C M C (2006) *Análise de Séries Temporais*. São Paulo: Edgard Blucher. 538p.
- [22] Haykin S (1994) *Neural Networks: A Comprehensive Foundation*. New York: Macmillan. 842p.
- [23] Nisbet R, Elder J, Miner G (2009) *Handbook of Statistical Analysis & Data Mining Applications*. Oxford: Academic Press (Elsevier).824p.
- [24] Olson D L (2005) Data Set Balancing. *Lecture Notes in Computer Science-Data Mining and Knowledge Management*. 3327: 71-80.