

# Energy price variation estimation using an ANFIS model

**Julián Moreno, Santiago Medina**

*Universidad Nacional de Colombia. Carrera 80, No. 65-232, Facultad de Minas  
Oficina M8A 311. Medellín, Colombia. Tel.: +574 4255221; Fax: +57 4 4255365.  
jmoreno1@unal.edu.co, smedina@unal.edu.co*

## Abstract

This article presents the implementation of an ANFIS model to estimate price movements in the energy market in Colombia in the short term with regard to hydrological conditions in order to support the decision making process of the involved agents. The proposed model considers three input variables: an ENSO phenomenon forecast, the system's reservoir level and the hydraulic contributions of the rivers to the generation plants. The model was trained and validated with actual monthly data from 1999 until 2007. According to the results, the model captures the price tendency in an accurate way, but it must be complemented with other models that reflect the influence of other factors in the energy price formation process in this country or in others with similar features.

**Key words:** energy market, price formation, hydrological phenomena, ANFIS.

## Estimación de la variación de precios de energía usando un modelo ANFIS

### Resumen

Este artículo presenta la implementación de un modelo ANFIS para estimar los movimientos de precio en el mercado eléctrico Colombiano en el corto plazo respecto a las condiciones hidrológicas con el fin de servir de apoyo a los procesos de toma de decisión de los agentes involucrados. El modelo propuesto considera tres variables de entrada: el pronóstico del fenómeno El Niño, el nivel de embalses del sistema y los aportes de los ríos a las plantas de generación. El modelo fue entrenado y validado con datos reales mensuales desde 1999 hasta 2007. De acuerdo con los resultados obtenidos, el modelo captura la tendencia del precio de manera adecuada, aunque debe ser complementado con otros modelos que reflejen la influencia de otros factores en el proceso de formación del precio de la electricidad en este país o en otros con características similares.

**Palabras clave:** mercado eléctrico, formación de precio, fenómenos hidrológicos, ANFIS.

### Introduction

As in any other market, the goal of the strategies of trading agents in the energy market is to obtain a positive difference between the purchase and sales prices in order to earn enough profits to cover the costs structure and generate greater value for stockholders. In the particular case of Colombia as well as other mar-

kets with similar features and where prices are determined by the market, these agents can take a conservative position and buy all the energy they need through bilateral contracts to fulfill the demand of their users (long term coverage mechanism), or they can take a more aggressive position and speculate, buying energy above or below their demand so they can sell or buy the surplus or deficit in the energy pool (short term

balance mechanism), taking advantage of price fluctuations.

Fluctuations in prices are due mainly to the high volatility of pool prices which are directly affected by climatic conditions (around 70% of electricity in Colombia is produced by hydraulic plants). According to this, a major concern for trading agents is to determine, among other factors [1], how the hydrological conditions affects the energy trading process, and moreover, how all they are related. In this article, the dependence between some of these variables and the energy price variation is studied based on historic data and some statistical proofs like the Pearson correlation coefficient. However, the aim of this work is not to obtain a forecast of such price based on such variables; instead a model that captures its tendency is presented. It is important to highlight that only such variation expectation is studied, but not how it influences trading decisions.

The rest of this document is organized as follows: In the next section the considered variables and their impact in price formation are described. Then, the structure and calibration of the proposed model are discussed and later the validation of such model with a summary of the obtained results is described. The final section presents the conclusions and gives a short outlook.

### **Description of the considered variables**

In countries like Colombia, where most of the generation plants are hydraulic (the rest are mainly thermal using as primary fuel natural gas or coal), weather is an important issue in the energy price formation because it affects the resource availability and therefore the offer. To incorporate this issue into the decision process of the agents several variables which reflect weather conditions may be considered. Such variables may have direct or indirect effects in the energy prices, in the former case because of the resource-offer relation, and in the last case because they may produce speculations within the market and therefore altering price volatility. Based on interviews to experts from several companies of the electric sector, the following variables were identified:

- "El Niño" / South Oscillation phenomenon - ENSO
- Reservoir level (mean level of dams)
- Contributions of the rivers

In annual scale the most important phenomenon for the Colombian hydrology is the ENSO, while in a minor scale there are other phenomena like the Madden-Julian oscillation, the Caribbean and Northeast Pacific hurricanes, the East tropical waves, and the Chocó flow [2]. Regarding these phenomena we consider that reservoir level and the contributions of the rivers concentrate their general effects in the different regions of Colombia, presenting some seasonal features.

In the rest of this section a statistical analysis for each of these variables is presented. The goal of this analysis is not establishing causality relations between them and the prices (even if they can be deduced). Instead, we try to justify its inclusion into the model from a statistical point of view.

#### **ENSO phenomenon**

This phenomenon is a natural event and results of the interaction between the ocean and the atmosphere in the region of the equatorial Pacific. An abnormal pluviometric deficit or drought due to this phenomenon is known as a "Niño", while a surplus is known as a "Niña".

There are several weather indexes which record the evolution of the ENSO in several influence regions (we are interested on 5°N-5°S, 120°-170°W) like the Sea Superficial Temperature - SST 3-4, the South Oscillation Index -SOI, the Outgoing Long wave Radiation in the equatorial zone - OLR, the Multivariate ENSO Index - MEI, etc. Among them, the SST 3-4 (the one we chose) and the MEI are the most important index to forecast the ENSO and they are published by the National Oceanic Atmospheric Administration of the United States - NOAA [3]. A value of the SST 3-4 anomaly above 0.5 is a signal of the occurrence of "El Niño", while a value below -0.5 is a signal of "La Niña".

The Pearson correlation coefficient between SST anomaly and energy price is 0.65. This indicates that high levels of the SST index are associated to rises in this price which can be explained

because the warming produced by the phenomenon causes a fall in the hydrological resources and therefore a rise in the generation costs.

In the model proposed in this article we do not try to make a forecast of the ENSO based on such index. Instead we propose using the NOAA's forecasting in order to foresee the price evolution. This decision, beyond the accuracy of the forecasting models that NOAA uses, is due to the speculation that is created among agents because NOAA is one of the more referenced official sources.

### Reservoir level

This level can be seen as the reserve of available water for the electricity generation and it is defined as the mean of all the dam levels. The evolution of such reserves follows a regular pattern due to winter-summer seasons and the annual scale phenomena. There is a continuous growth during winter, generally in the months of March-April and September-November, and a falling tendency in the remaining months. The reservoir level presents a significant correlation (the Pearson correlation coefficient is  $-0.47$ ) with the energy pool price. This negative relation means that when this level goes down, the water reserves are low, and therefore it is necessary to increase the use of other plants (like thermals), which have higher operation costs.

Additionally, the correlation of this variable with the ENSO is significant. The correlation with the MEI and the SST 3-4 are  $-0.61$  and  $-0.44$  respectively, which means that reservoir level goes down when the index go up (during drought) and vice versa.

### Contributions of the rivers

The flow of the rivers is an important variable because they are the natural supply source for dams and water edge plants. In order to measure this variable, the relation between the current hydrological contribution level and its historic mean may be used. This relation determines if the hydrological contributions of one period are higher or lower than its mean and it allows foreseeing the future water availability for the system in the short term. A way to measure that relation is:

$$RC_t = \frac{CR_t - CHM}{CHM} * 100\% \quad (1)$$

where CR is the contribution of the rivers and CHM, contribution's historic mean.

If  $RC_t$  is negative that means that the contribution of the rivers of the period  $t$  is below the mean of such month (a situation of contribution deficit) and it may produce generation problems if the dam levels are low and therefore it may produce a rise in the prices. Otherwise, if  $RC_t$  is positive that means that the contribution of the rivers in the period  $t$  is above the mean (a situation of contribution surplus) and it may produce a fall in the prices if the dam levels are above normal and "El Niño" is not foresaw in the short term.

This variable presents a significant correlation (the Pearson correlation coefficient is  $-0.52$ ) with the energy pool price.

### Price variation

This variable corresponds to the medium term price variation of the energy pool price. To measure such variable several methods like a log-differenced function may be used. However we propose a more empirical method that relates the medium term past mean prices giving more importance to the more recent ones with the medium term future mean prices. The proposal (whose formulation has a more interpretation purpose than mathematical) consists in using the simple pondered mean of the previous three months compared with the arithmetic mean of the next three, as shown in (2).

$$PV_t = \frac{P2 - P1}{P1} \quad (2)$$

where  $P1 = \frac{1 * p_{t-3} + 2 * p_{t-2} + 3 * p_{t-1}}{6}$  and

$$P2 = \frac{p_{t+1} + p_{t+2} + p_{t+3}}{3} \quad (3)$$

The evolution of the Colombian energy pool price monthly mean ( $p_j$ ) and the corresponding variation ( $PV_j$ ) are shown in Figure 1.

As the output variable, this value represents the expected medium term percentage vari-

ation of the energy pool price due to the considered input variables.

### ANFIS model

The Adaptive-Network-based Fuzzy Inference System (ANFIS) was proposed by Jang and Sun [4] and it combines the features of fuzzy logic and neural networks. This system has been used in numerous modeling and forecasting problems, including specifically some works on energy sector [5, 6]. The conceptual and mathematical details of ANFIS are not presented in this paper but we invite readers to review the documents which are cited in this section for further information.

### Model structure

As it is show on Figure 2 this model has five layers where each one corresponds to a different part of the inference system. As a summary an ANFIS is a set of IF <antecedents> THEN <consequents> rules, where the antecedents are fuzzy sets and the consequents are linear functions of the type  $a_i x + b_i$ . Shape of fuzzy sets and functions parameters are calibrated by the model minimizing an error measure. In other words, an ANFIS uses a learning rule to optimize the parameters of a first order Takagi-Sugeno system [7-9].

### Training and validation data sets

In order to optimize the parameters of the model it is necessary to divide the available sam-

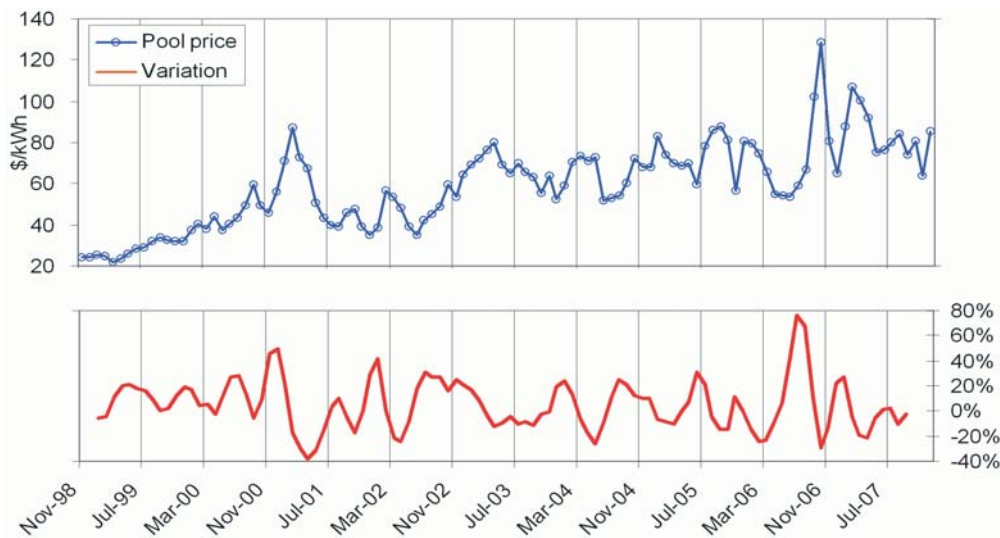


Figure 1. Evolution of the pool price and its variation during 1999-2007.

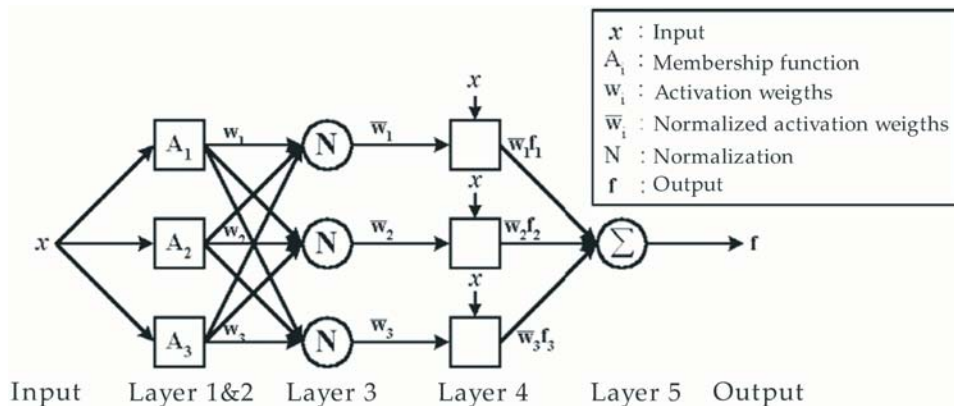


Figure 2. ANFIS structure.

ples (collections of input and output values) into two separated sets: the training set and the validation set. The first one is used to calibrate the model using a supervised learning algorithm, and the second one is used to “validate” such calibration comparing the outputs values of this set with the resulting outputs of the model using the corresponding inputs. This comparison shows the generalization capacity of the calibrated model and allows determining when it is over trained.

In this work the 90% of samples were used for training and the remaining 10% were used for validation. To extract that 10%, several criteria may be used: Taking the samples from the end of the samples, taking them from equally spaced intervals, or taking them randomly. In this case, the second option was chosen taken care of avoiding seasonal effects. The other ones were discarded because they may bias the results, although a combination of second and third may be a better option.

### Calibration

To implement the model presented in this paper we used Matlab, specifically the fuzzy toolbox and the *anfisedit* command. In this software, the first step in calibration is to generate the initial fuzzy sets defining the number of sets for each input variable and their shapes. In the first of these tasks two aspects must be considered: Large number of sets may produce better fitness in training process but a poor validation due to insufficient samples; on the other hand, few sets may produce a poor fitness but a more adequate

generalization. To solve this dilemma and to achieve both goals, fitness and generalization, we select these numbers according to relation among input and output variables (statistically, physically and economically) and keeping the number of resulting rules short. Following this criterion, and after some trials, the selected sets numbers was: 4 for the ENSO phenomenon expectation (SST forecast), 2 for the reservoir level and 2 for the contributions of the rivers.

Now, in order to do define the shapes of the sets there are three alternatives: to select an arbitrary shape, using a uniform distribution within the domain of the variables, or using some clustering method. In any of these alternatives some of the standard representation shapes must be used: triangle, trapezoidal, Gauss curve, etc. The best results in this work were obtained when the uniform distribution was used, and the shape of sets was defined using Gauss curves whose formula is presented in (4).

$$e^{-\frac{(x-c)^2}{2\sigma^2}} \quad (4)$$

Once the initial fuzzy sets are selected and the training and validation sets are loaded, it is necessary to choose the calibration algorithm. This toolbox in version 7.4.0.287 (R2007a) has implemented a back propagation and a hybrid algorithm. In this case, the second one was selected.

From calibration results, the fuzzy sets for one of the input variables are shown in Figure 3, being the vertical axis the membership value.

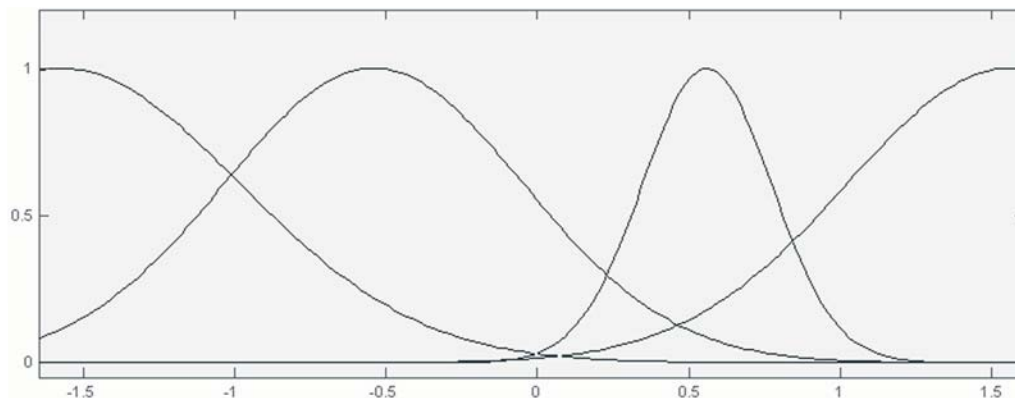


Figure 3. Fuzzy sets for ENSO phenomenon expectation variable (in °C).

## Results

The fitness of the model was measured with regard to the actual output as shown in Figure 4 in the two cases: training and validation. During training the fitness was 72%, that means that in 72% of the samples (months) the output of the system matched the actual tendency of the prices considering a variation percentage of 10% in both cases to define the three output sets: fall, stable or rise. This value may not seem pretty high considering that such sample was used to extract the inference rules of the system. However it is important to highlight that the movements of the prices do not obey just to hydrological factors, but also to other issues as economic, social and regulatory, etc.; in this way it is possible to think in such fitness as a measure of the contribution of only the factor studied in this article to the price formation process.

In validation the fitness was 50%, and the five wrong matches corresponds to months: October 99, June 01, August 05, June 06 and April 07. In the first, third and fifth cases there seems to be a jump in model's output that does not correspond with actual data. These cases suggest that model may be improved including other factors and heuristics about cautious or speculative behaviors of the generation agents. On the other hand, in second case (among with previous three training samples) there seem to be a lag between model and real data that highlights the importance of agents expectation (over actual conditions) under certain market conditions.

As an interesting point, it can be observed in Figure 4 that there is an outlier on April 02 as

result of an over sizing of the output value, however the tendency of price (in this case a fall) is correctly modeled.

Another kind of analysis that can be done is to visualize the relationship between input variables and their contribution to the output variable through fuzzy surfaces. These surfaces allow us to watch in a 3D view the possible combinations of two input variables with regard to output variable while the other is fixed around its mean, as show in Figure 5.

According to the fuzzy surfaces and the outputs given by the model in different conditions, it is possible to deduce, among others, the next considerations.

About the market:

- Low levels of the reservoir in a specific period can be compensated with high levels of the contribution of the rivers under certain conditions because it can be inferred that such contribution will raise the dams in the near future.
- The ENSO phenomenon is an important factor for energy price formation in countries located in its influence area not only because of its relation with other variables that determine the generation capacity, but also because it produces speculations among the agents of the market.

About the model:

- Relationship between input and output variables is non linear.
- The range of the output variable compared to real data clearly shows that training and



Figure 4. Actual price variation versus model's output.

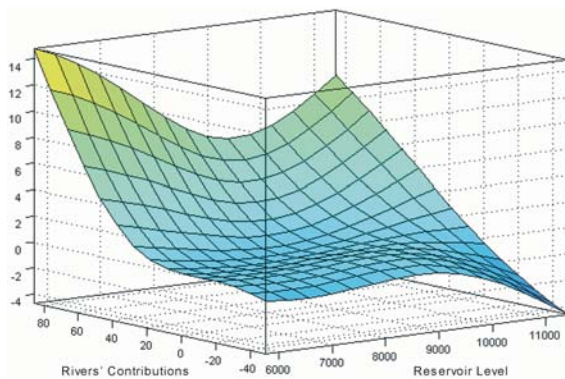


Figure 5. Fuzzy surface for the Rivers contributions (%) - Reservoir level combination (GWh).

validation data sets are insufficient. Some oversized output values like the ones shown in Figure. 5 are result of the lack of samples for some combinations of input data. For instance the high levels of output values can be observed in this figure when *RC* has high values, *Reservoir level* has low values and *expected ENSO* has a mean value. In this case saying the model is wrong may be inadequate, instead we may argue that this specific combination of variables was not present (or was not statistically significant) in history.

## Conclusions

An ANFIS model is proposed in this paper in order to measure the expectations of the pool price variation in the Colombian energy market with regard to hydrological factors. However, such model may not be used as a kind of predictor or expert system for the pool price evolution; instead it can be used, joined with the analysis of other factors that alter the price, as a tool in the making decision process for the trading agents. Among these factors there are economic, political, social, financial, risk and regulatory conditions. What this last statement indirectly means is that, even if the output of the proposed model does not match the actual tendency of the energy price in some periods as results showed, it does not mean that its formulation is wrong, instead it means that price is also affected by a set of other variables. For example, in the first months of

2006, the electricity pool price was in high levels even when an irregular winter was presented. For this period, this "anomaly" may have diverse causes like the proximity of the presidential elections, the changes on the regulations in such period, among others. All those other factors may be considered in a more robust model that is beyond the scope of this paper but that could be an interesting research.

The proposed model considers specifically as input variables the ENSO phenomenon expectation, the reservoir and the contribution of the rivers to the generation system. According to the analyses that were made, all the input variables have a significant relation with the price movements. The last two because they reflect the country hydraulic generation capacity and the former one because it directly affects the other two and may cause speculations within the market.

Even if the obtained results are promising, it is obvious that a better calibration may be achieved with more available data. However in the Colombian case, where actual market structure is relatively young this can not be done yet. As an alternative, a latter model with heuristic criteria may be implemented in order to smooth output values flattening fuzzy surfaces. Another improvement that may be done is using some rules to separate the cases when previous prices are already too high or low (historically speaking).

Finally, it is important to highlight that the model presented in this paper was built for Colombian energy market; however it can be extrapolated to other markets with similar features, especially if their generation resources are mainly hydraulic plants.

## References

1. Medina, S. and Moreno, J. Risk evaluation in Colombian electricity market using fuzzy logic. *Energy Economics*, Volume 29 (5), (2007) 999-1009.
2. Mesa, O., Smith, R., Salazar, J. y Carvajal, L. Modelos de predicción de caudales para el sector eléctrico Colombiano. Parte I: modelos mensuales y modelo semanal. *Revista avances en recursos hidráulicos*, N° 3. 1995.

3. National Oceanic Atmospheric Administration of the United States - NOAA. <http://www.cpc.ncep.noaa.gov/>
4. Jang, J.-S.R. and Sun, C.-T. Neuro-fuzzy modeling and control. *Proceedings of IEEE* 83 (3), (1999) 378-406.
5. Alturki, F., A Fuzzy Neural Approach For Forecasting Peak Power Demands. *Journal of King Saud University*, Volume 14, (2002) 29-40.
6. Mora, J., Carrillo, G. and Pérez, L. Fault Location in Power Distribution Systems using ANFIS Nets and Current Patterns. *Proceedings of IEEE PES Transmission and Distribution Conference and Exposition Latin America*. 2006.
7. Wang, L. Fuzzy systems are universal approximators. *Proceedings of IEEE International Conference on Fuzzy Systems*. 1992.
8. Bojadziev, G. and Bojadziev, M., *Fuzzy logic for business, finance, and management*. World Scientific Publishing Co. 1997.
9. Kasabov, N. *Foundations of Neural Networks, Fuzzy Systems and Knowledge Engineering*. The MIT Press. 1998.

Recibido el 9 de Julio de 2009

En forma revisada el 6 de Diciembre de 2010