MODELING OF BIDDING PRICES IN POWER MARKETS USING CLUSTERING AND FUZZY ASSOCIATION RULES

MODELAMIENTO DE PRECIOS DE OFERTA EN MERCADOS DE ELECTRICIDAD USANDO ALGORITMOS DE AGRUPAMIENTO Y REGLAS DE ASOCIACION DIFUSAS

LUIS EDUARDO GALLEGO VEGA

Department of Electrical and Electronic Engineering, Universidad Nacional de Colombia –Bogotá, lgallegov@unal.edu.co

OSCAR GERMAN DUARTE VELASCO

Department of Electrical and Electronic Engineering, Universidad Nacional de Colombia –Bogotá, ogduartev@unal.edu.co

Received for review September 21 th, 2009, accepted October 3th, 2010, final version October, 3 th, 2010

ABSTRACT: In this paper, a strategy for discovering patterns over the continuous domain of bidding prices is proposed. In particular, the proposed method represents bidding functions as points in a multidimensional space where a clustering algorithm is applied. Also, as a result of this method, a dramatic reduction over the search space of bidding strategies is achieved. In addition, some relations of dominance over bidding strategies are found, improving the pattern recognition process of agents' bidding behavior. This method is applied on the bidding prices database for some GENCOs of the Colombian power (electricity) market. Furthermore, an application of some data mining algorithms is presented with the purpose of quantifying some hypothesis formulated on the effect of hydrology over both spot and bidding prices.

KEYWORDS: Bidding prices, electricity market, pattern recognition.

RESUMEN: Este artículo propone una metodología para el descubrimiento de patrones sobre el dominio continuo de los precios de oferta en mercados de electricidad. El método propuesto representa las funciones de oferta como puntos en un espacio multidimensional donde un algoritmo de agrupación es aplicado. Como resultado se obtiene una gran reducción sobre el espacio de búsqueda de estrategias al igual que algunas relaciones de dominancia sobre las mismas, mejorando el reconocimiento de patrones de comportamiento de la oferta. Este método es aplicado para los 10 agentes generadores más grandes del mercado eléctrico colombiano. Adicionalmente, se presenta la aplicación de algoritmos de minería de datos con el propósito de cuantificar la veracidad de algunas hipótesis tradicionalmente formuladas sobre el efecto de la hidrología tanto en los precios de oferta como en los precios de bolsa.

PALABRAS CLAVE: Precios de Oferta, Mercados de electricidad, Reconocimiento de patrones

1. INTRODUCTION

During the last two decades, the electric industry has experimented with several regulatory changes evolving from vertically integrated industries to deregulated systems where competition has been introduced, especially in the electric generation sector of the industry. In Colombia, this deregulatory process occurred in 1995 with the creation of a power market mainly based on two mechanisms: a spot market and a bilateral financial contract market working as short and long-term mechanisms, respectively.

Based on this competitive framework, Generation Companies (GENCOs) face a new problem which consists of the modeling of their competitor's bidding behavior, and in particular, the behavior of bidding prices. In addition, this problem seems to be more difficult to solve according to the most recent regulatory rules in Colombia (January 2009) [1] which state that bidding prices must now be kept confidential during a period of 3 months—with the aim of improving the market's competition.

A major problem commonly faced in modeling the behavior of bidding prices consists of determining the most representative bidding strategies of market agents. In this regard, several attempts to model this behavior have been reported in the literature during the last five years. In [2] a fuzzy regression model is used to search for bidding price patterns in electricity markets strictly related to a given demand value. Reference [3] presents an application of data mining over the power market data generated by an auction simulator in order to reduce the epistemic uncertainty in VaR/PaR inferences by using the information gap theory. In [4] the authors use a data mining approach to search for both normal spot price levels and spot price spikes which are based on the bidding prices provided by Queensland, Australia's electricity market. Reference [5] uses a decision tree-based classifier called SLIQ in order to represent the knowledge in the bidding decision system of a generation unit subject to the market's demand and unit capacity.

The continuous nature of the bidding prices domain makes it difficult to assess and predict a competitor's behavior; therefore, a methodology for discretizing this behavior seems to be mandatory. Also, it is common sense to think that Generation Companies (GENCOs) design their strategies on their own portfolio of generating plants, therefore, the construction of bidding functions that represent the aggregated behavior of GENCOs seems to be logical.

Following this aggregated approach, three possible questions need to be properly addressed: (i) What bidding prices are assigned by GENCOs to each one of their generating plants?, (ii) Do these bidding prices correspond to some particular bidding strategies? and (iii) Is there a common knowledge about the main variables that drive these bidding strategies? If so, Can this knowledge be quantified by any means? These are the questions that may be solved by using the methodology proposed in this paper.

2. THE PROPOSED METHODOLOGY

With the aim of addressing the above mentioned issues, it is necessary to find a strategy of pattern recognition among bidding prices that are assigned to the different portfolios of generating plants. This search of patterns allows assessment of bidding functions that are typically used by GENCOs during their daily routine.

The proposed methodology is based on a popular clustering algorithm called k-means which has been extensively used in unsupervised learning tasks. However, the main contribution of this methodology is not the application of the k-means algorithm, but the representation of bidding functions such as points in a multidimensional space where clustering algorithms might be applied. In addition, a new data mining algorithm is used to quantify the confidence in common knowledge about some variables that are currently assumed to affect bidding strategies.

2.1 The *k*-means algorithm

Broadly speaking, the *k*-means algorithm consists of a clustering method that divides a space of n objects into *k* partitions, where each partition represents a cluster. These clusters are formed following the idea of optimizing a similarity function, so that objects within a cluster maximize this function and, on the contrary, objects of different clusters minimize this function. In the case of *k*-means algorithm, this function is commonly defined as a distance from any object that belongs to a particular cluster, to a point that is defined as the centroid of this particular cluster. Consequently, this algorithm returns two arguments as follows:

(i) A vector of length *k* representing the centroids for the *k* clusters.

(ii) A vector of length n representing the assignment of each object to one of the *k* clusters.

The centroids are obtained by an iterative process attempting to minimize the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances. Therefore, the algorithm converges to a local optimum that consists in a partition of points in which moving any single point to a different cluster increases the total sum of distances [6]. This sum is defined as [7]:

$$D = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2 \qquad (1)$$

where:

m_i is the assigned centroid of cluster C_i.

p is an object assigned to cluster C_i.

D represents the sum, over all clusters, of the within-cluster sums of point-to-cluster-centroid distances.

Algorithm 1 Clustering algorithm k-means

1: Initialize k // Number of partitions
2: Select randomly k objects as initial centroids
3: (Re)assign each object to a cluster whose centroid is nearest.
4: (Re)calculate centroids in each cluster no change in centroids is observed

Finally, this simple clustering algorithm follows the 4 steps that are shown in algorithm 1.

2.2 Clustering methods over the space of bidding prices

The application of *k*-means algorithm over the continuous domain of bidding prices starts from a suitable representation of bidding functions as points in a u-dimensional space, where u corresponds to the number of generating plants that each GENCO owns.

In mathematical terms, an agent might own *n* generating plants of *q* different technologies, and might assign a distinct bidding price p_{ij} to each one of the plants in his portfolio. Now, bidding functions (C*a*) of agent *a* must be a function of bidding prices as follows:

$$C_a = f(p_{ij}) \ \ / \ i \in [1,n] \ , \ j \in [1,q] \qquad (2)$$

It is possible to think that bidding functions also depend on quantities (MW) that are declared available to the System Operator (ISO). However, in the case studied here, it is assumed that agents have no incentives to declare them unavailable. This assumption is based on the behavior exhibited by GENCOs in Colombian market data, showing that there is not a significant variation in plants availability. On the contrary, as a result of the Price-Based Unit Commitment (PBUC) scheme of the Colombian power market, bidding prices exhibit a great dynamics.

On the other hand, a meaningful difference is commonly observed among bidding prices of different generation technologies, probably explained by the distinctive operational costs associated to thermal and Hydro technologies. This condition might cause the application of clustering algorithms to result in fake patterns that fall in a space between bidding prices of different technologies. To avoid these unexpected results, a single bidding function is constructed for each generation technology in the portfolio of agents. Following this approach, a bidding function for each technology (Ca_j) is represented by a vector of length u, formed by bidding prices p_{ij} of the *u* generating plants that belongs to a particular technology *j*. Thus,

$$Ca_j = [p_{1j}, p_{2j}, p_{3j}, ...p_{uj}] \qquad (3)$$

Bidding functions represented in the form of equation 3, might be interpreted as points in a *u-dimensional* space, where each dimension represents a bidding price of a particular plant and where a clustering algorithm may be applied with the aim of finding certain regularities over this new space. It is important to highlight that this approach allows one to summarize the strategic bidding behavior of an agent in a single point over a multidimensional space. Moreover, by following this new approach, bidding behavior is not considered to be an isolated process for each generating plant, but to be a combined bidding strategy of the agents over their diversified portfolios.

2.3 Data mining association analysis on power market databases

Broadly speaking, data mining consists of a large collection of algorithms to extract or "mine" knowledge that might be *interesting* and *innovative* to the analyst. Among all the available data mining techniques, the application of association analysis should provide interesting results in power market databases**.** In general terms, *association analysis* is the discovery of *association rules* showing some data attributes that occur frequently together in a given set of data [2]. In this paper, this algorithm is used with the purpose of quantifying several hypotheses about the effect of some variables over bidding and spot prices. In particular, this algorithm is used to quantify the confidence about the effect of hydrological conditions on bidding and spot prices in a market *strongly dependent* on Hydro plants (as is the Colombian case).

These association rules are of the form X => Y in such a way that the hypothesis to be verified may also be represented by rules. In our case, rules may be in the form:

IF Hydrology is High => Spot Price is Low

IF Hydrology is High => Bidding Price is Low

In brief, the search for meaningful rules is based on the estimation of two basic parameters called *support* and *confidence*. The support (Sup) of a rule in the form X => Y measures the frequency of this rule on the data set, and the confidence (Conf) of X => Y measures the conditional probability of a *consequent* given a particular *antecedent*. Formalizing these relations:

$$Sup(X) = Prob(X)$$
$$Conf(X) = Prob(X \mid Y)$$

However, given the continuous domain of bidding prices and reservoir levels (hydrology), it is necessary to find a strategy of discretization over these domains. In this proposed methodology, a *fuzzy* discretization is used by applying a fuzzy version of *k*-means called fuzzy *k*-means. This type of clusters follows the idea of fuzzy sets [8,18], where an element belongs partially or to a certain degree to a specific set. Therefore, the assignment of one object to a particular cluster is defined by a "membership" degree to that cluster, taking values in the interval [0,1]. Figure 1 shows the type of fuzzy sets considered over the variables domain in the mined association rules.

On the other hand, the search of association rules following this new fuzzy approach becomes a little bit different because now *counting* must consider the different membership degrees of objects. *Fuzzy Association rules* is a relatively new data mining algorithm. References [9] [10] introduce the concept of Fuzzy Association Rules and present some applications on relational databases. Reference [11] develops this algorithm in a multidimensional databases model and reference [12] applies this methodology to search possible correlations between lightning parameters and some geographical and meteorological characteristics. In this context, with the purpose of evaluating the mined fuzzy association rules, a new parameter called *certainty factor* is defined. The *certainty factor* takes values in the range of [1,-1] and tries to measure not only the *presence* of the elements in a rule, but also the *absence*. To illustrate this situation it is possible to evaluate a rule in the form:

IF Hydrology is Low THEN Bidding Price is High

But we may also ask about the rule that implies the absence of the consequent and the antecedent as being:

IF Hydrology isn't Low THEN isn't Bidding Price High?

In this sense, the rule must be considered to be *very strong* if both rules are also strong. On the other hand, the sign of the *certainty factor* explains the dependence relation between the *antecedent* and the *consequent:* a positive sign appears when the dependence between the antecedent and the consequent is also positive, and a negative sign appears when the dependence is negative; zero means no dependence [13]. In other words, a positive certainty factor means that the *presence* of the antecedent *implies* the *presence* of the consequent, and on the contrary, a negative certainty factor means that the *presence* of the antecedent *implies* the *absence* of the consequent.

Finally, this data mining algorithm was applied to the Colombian power market database in order to search for association rules between Hydrology and Bidding prices. The values of *confidence*, *support*, and *certainty factors* are used to *quantify* the confidence and truthfulness of some hypothesis about the effect of hydrology on bidding prices.

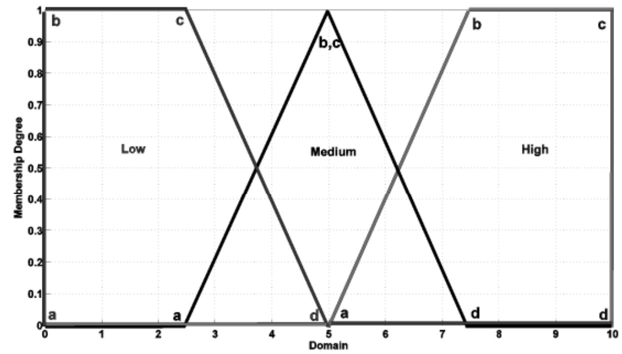

Figure 1. Type of fuzzy sets considered in the mined association rules

3. AN EXAMPLE OF APPLICATION

The methodology described above was applied to ten of the most representative GENCOs of the Colombian power market. With the purpose of illustrating the proposed methodology, a particular bidding of one of the analyzed agents is shown in Figure 2. This bidding function is constructed based on the bidding prices of *three* hydro plants portfolios of this particular agent which can be represented as a point in a *three*-dimensional space. This particular point representing the bidding function is shown in Figure 3.

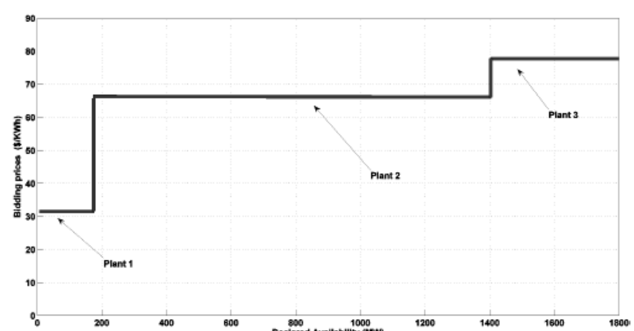

Figure 2. An example of a Hydro bidding function of a Colombian GENCO

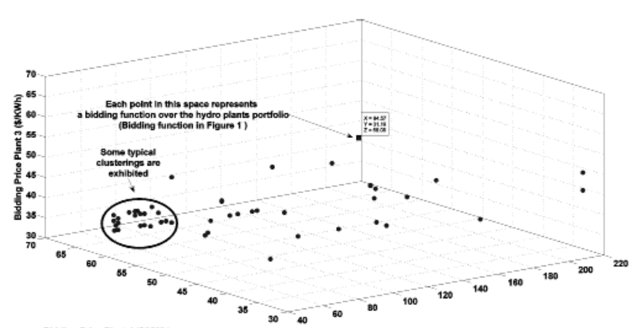

Figure 3. Representation of bidding functions in a n-dimensional space for a particular agent

On the other hand, the rest of the points are also a representation of different bidding functions, capturing the behavior of this particular agent in the period from January to February of 2004 [14,15]. In addition, some typical clusters may be identified in this figure's space, representing that some bidding strategies are used more often than others in a distinctive manner. Precisely, the proper identification of these clusters is the main goal of the *k*-means algorithm and furthermore, the centroids of these clusters may be considered as the bidding patterns of the agent in consideration. Therefore, as a result of the application of the *k*-means algorithm, *k* bidding patterns are obtained for each type of plants portfolio (Thermal, Hydro, Nuclear) of the agent.

To illustrate the application of the *k-means* algorithm, the obtained clusters for the agent in Figure 3 are shown in different colors in Figure 4 for a value of $k = 6$.

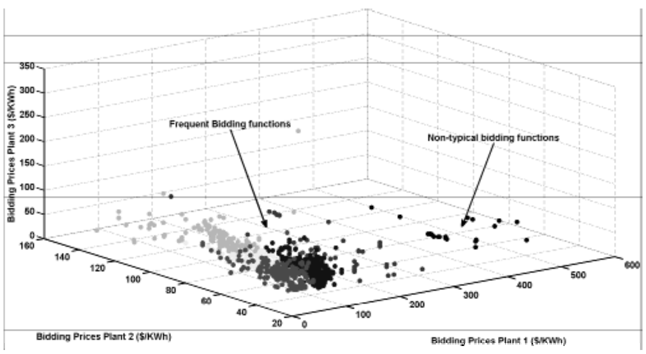

Figure 4. Application of the *k*-means algorithm over a three-dimensional space of bidding functions

As was mentioned previously, the methodology was applied to ten of the most representative agents of the Colombian power market in a sequence of steps that are shown in Figure 5 and may be summarized in four steps as follows:

1) Achieve SQL queries about the bidding prices of all agents' portfolios on a new market database. A new Colombian power market database was constructed with the purpose of organizing and combining several sources of information about this power market. This new database allows us to correlate essential information about agents that is commonly dispersed.

2) Classify bidding prices by different technologies in order to obtain the desired bidding functions.

3) Apply the *k*-means algorithm in an *iterative* manner. In this case, one *iteration* ends when a partitioning of the space in *k* clusters is obtained. However, this application is iterative in the sense of progressively increasing the number *k* of clusters until any stop criterion is reached. With the proposed methodology, this stop criterion consists of a *inter-centroids distance,* which is calculated on each iteration until 10% of these *inter-centroid distances* are less than 10% of the *maximum inter-centroid distance* calculated in the current iteration. The underlying idea of this criterion is to search for a number of *k* partitions over the space, so that centroids are not close enough to be considered as a single cluster.

4) If the *stop criterion* is fulfilled, current centroids are taken as the *k* bidding function patterns.

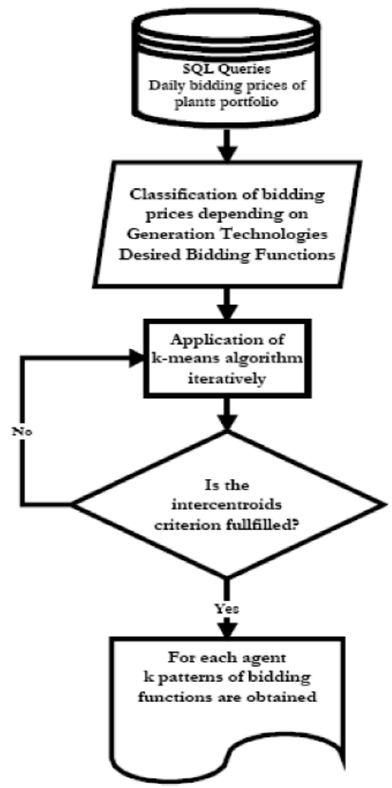

Figure 5. Flow chart of the application of k-means to obtain bidding pattern functions

4. RESULTS

Data mining Association Analysis on power market databases

As was mentioned earlier in the paper, the proposed clustering algorithm was applied to ten of the most representative GENCOs in the Colombian market. As a result, each GENCO was characterized by *k* bidding pattern functions that must be interpreted as the most representative and distinctive bidding strategies of the agent considered. An example of these bidding patterns is shown in Table 1 for a Hydro plant portfolio of one of the most complex GENCOs in the Colombian power market.

Table 1. Details of bidding function patterns for a hydro plant portfolio. *All prices are in ($/KWh) at constant prices of December, 2004

Strategy	Plant 1	Plant 2	Plant 3	Plant 4	Plant 5	Plant 6
1	60.8	70.9	57.3	60.2	71.5	471.3
2	47.1	281.2	118.8	43.9	258.0	482.5
3	42.8	37.6	37.4	38.1	39.0	52.4
4	55.2	57.3	46.6	54.6	56.9	80.4

Furthermore, it is important to notice that the proposed clustering methodology finds bidding patterns for each technology of the generating plants (thermal or hydro). Consequently, agents owing hydro and thermal plants are characterized by bidding patterns in each one of the owed technologies. Now, these types of agents may offer a *combination* of these patterns to the power market. These types of bidding strategies are considered here to be *mixed strategies*. This is not to say mixed strategies in the sense of a Nash Equilibrium, but in the sense of a combination of bidding patterns resulting from different technologies or portfolios. Table 2 shows mixed strategies for a GENCO with both portfolios as specific combinations of a particular bidding pattern in hydro and thermal portfolios. For example, Mixed strategy # 8 is the combination of bidding pattern # 1 in a hydro portfolio and bidding pattern # 4 in a thermal portfolio.

However, one of the most interesting results shows that *not necessarily* all the possible combinations of *mixed strategies* are used by agents and also, not all the strategies are used with the same frequency. This result shows that it is possible to establish an *ordering* or *dominance* over the strategies space in terms of their frequency of use. To illustrate this *dominance relation*, Table 2 shows a description of used strategies and their frequency of use for one of the most complex GENCOs of the Colombian market.

First of all, it is important to say that in spite of the diversity of its portfolio, a big reduction is achieved over the continuous domain of bidding prices (16 pattern strategies for this particular agent). Secondly, not all the possible combinations are used (just 8 of 16 possible mixed strategies are used). Third, a set of 4 strategies are employed most of the time (90% of the time). Therefore, from an infinite set of possibilities in the continuous domain of bidding prices, it is possible to model the bidding behavior of this agent by 4 strategies, which is an exceptional reduction over the bidding prices space. Finally, Table 3 summarize the reduction mentioned, describing the number of bidding pattern strategies over the *possible*, the *used*, and the *dominant* space where in some cases this reduction may take values around 50 %.

Table 2. Mixed strategies for agents with diversified portfolios

Mixed Strategy	Hydro Portfolio	Thermal Portfolio	Frequency
1	1	3	1
2	4	3	5
3	2	4	26
4	4	2	32
5	3	1	55
6	3	2	207
7	4	1	289
8	1	4	370

Table 3. Comparison between possible and dominant strategies for the agents considered in the Colombian power market

GENCO	No. Possible Strategies	No. Used Strategies	No. Dominant Strategies (90%)
EPM	16	8	4
ISAGEN	24	24	14
EMGESA	30	25	15
EPSA	24	22	12
BETANIA	4	4	3
CHIVOR	7	7	6
EBSA	5	5	4
T/JERO	4	4	4
URRA	5	5	5

4.2 Fuzzy association rules

A data mining algorithm based on fuzzy association rules was applied to the Colombian power market database in order to *quantify* the *confidence* and *truthfulness* of some hypothesis about the effect of hydrology on bidding prices. In particular, this hypothesis has been traditionally formulated in the Colombian power market due to the great dependence of the system on hydro plants [16,17,19]. In fact, it has been considered historically that an inverse relation between bidding prices and hydrology exists, probably supported in the distinctive operational costs associated with thermal and hydro technologies.

In this section, the parameters of *confidence*, *support*, and *certainty factors* are used to *quantify* the veracity of this hypothesis by two different approaches. The first approach consists of analyzing this hypothesis in an aggregated form, "mining" the relation between the daily spot price and the national bidding reservoir. The second approach analyzes this hypothesis in a disaggregated form, "mining" the relation between bidding prices for a particular plant and its bidding reservoir. As usual, this hypothesis is formulated in the form of a rule X => Y.

4.2.1 Spot prices and national bidding reservoir

The results of this relation were obtained with the tool Fuzzy Text Data Miner developed in reference [13]. The results are shown in Table 4. Given that a natural trend in association analysis is to select rules that exhibit high values of *support*, these rules are shown in blue color on Table 4. As mentioned previously, these are the rules that are candidates for being *strong rules*. Therefore, in this subset some important rules have been disregarded. For instance, the rule bidreservoirlow => pricehigh has *low* values of *support*. However, it is interesting to observe the *confidence* and *certainty factor* values for this rule, which are also *low* (24.23% and 0.15, respectively). These values may be interpreted as a low confidence in the belief that a necessarily low national bidding reservoir implies high spot prices. Hence, this information quantified in that form may become an important tool in the decision making process of GENCOs.

On the contrary, focusing on rules with high *support*, it can be seen that some rules also exhibit a high value of *confidence*, which are highlighted in red color. Regarding this subset, two rules seem to be interesting.

The first rule is bidreservoirmedium => pricemedium with confidence and certainty factor values of 53.48% and 0.17, respectively. At first sight, it seems that if the national bidding reservoir is at a medium level, there is a probability of 50% that the spot price is also at a medium level. Furthermore, checking the certainty factor values, it seems that there is a *weak* relation which implies the *presence* of both medium prices and medium reservoirs at the same time.

Table 4. Resulting Fuzzy Association Rules for daily spot prices and national bidding reservoir

Rule	Sup (%)	Conf (%)	CF
natreservoirlow ⇒ *pricelow*	1.82	7.15	-0.84
pricelow ⇒ *natreservoirlow*	1.82	2.92	-0.81
natreservoirmedium ⇒ *pricelow*	14.37	29.09	-0.36
pricelow ⇒ *natreservoirmedium*	14.37	25.89	-0.30
natreservoirhigh ⇒ *pricelow*	34.84	70.05	0.45
pricelow ⇒ *natreservoirhigh*	34.84	74.47	0.51
natreservoirlow ⇒ *pricemedium*	11.90	65.54	0.38
pricemedium ⇒ *natreservoirlow*	11.90	21.86	0.08
natreservoirmedium ⇒ *pricemedium*	24.11	53.48	0.17
pricemedium ⇒ *natreservoirmedium*	24.11	44.40	0.12
natreservoirhigh ⇒ *pricemedium*	17.10	30.76	-0.31
pricemedium ⇒ *natreservoirhigh*	17.10	31.72	-0.33
natreservoirlow ⇒ *pricehigh*	5.36	24.23	0.15
pricehigh ⇒ *natreservoirlow*	5.36	33.09	0.21
natreservoirmedium ⇒ *pricehigh*	7.30	18.84	0.09
pricehigh ⇒ *natreservoirmedium*	7.30	65.55	0.45
natreservoirhigh ⇒ *pricehigh*	0.87	1.35	-0.87
pricehigh ⇒ *natreservoirhigh*	0.87	3.72	-0.92

The second rule is bidreservoirhigh => pricelow with confidence and certainty factor values of 70.05% and 0.45, respectively. Apparently, this rule is the strongest rule mined, implying that if the national bidding reservoir is at a high level, there is a probability of 70% that the spot prices are also at a low level. However, certainty factors are not as high as might be expected (0.45), which may suggest that the presence of high national bidding reservoir levels does *not necessarily* imply the presence of low spot prices. These values may be supported in the data behavior shown in Figure 6 where most spot prices may also be medium when national reservoirs are low. To sum up, despite the high levels of confidence, this rule is not as strong as has been historically stated. Again, the main idea is that by using an association analysis, the *truthfulness* of this hypothesis may be *quantified*, which is a huge contribution to market data analysis.

In conclusion, an asymmetry seems to exist in the hypothesis historically formulated about the effect of hydrology in spot prices. Apparently, there is more confidence in the belief that "*high bidding reservoir levels imply low spot prices*" that in the belief that "*high spot prices are caused by low bidding reservoir levels*". These asymmetries found by applying the proposed methodology give more information about the bidding behavior of agents in power markets.

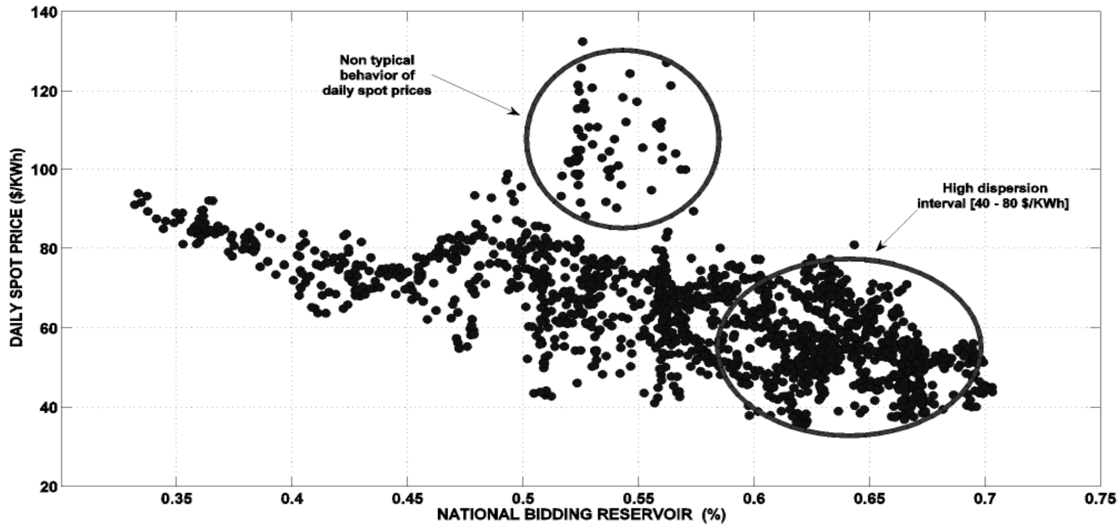

Figure 6. Relation between National Bidding reservoir and daily spot prices

4.2.3 Bidding prices and Bidding reservoirs

In like manner, fuzzy association rules were used to "mine" the disaggregated relation between Bidding prices and Bidding reservoirs. The results are shown in Table 5. It is important to distinguish that, in this case, the bidding reservoir corresponds to a particular plant as well as bidding prices. The results shown in Table 5 belong to the bidding behavior of one of the biggest hydro plants in Colombia.

Table 5. Resulting Fuzzy Association Rules for bidding prices and bidding reservoir

Rule	Sup(%)	Conf(%)	CF
bidreservoirlow ⇒ pricelow	1.41	5.40	-0.72
pricelow ⇒ bidreservoirlow	1.41	3.12	-0.73
bidreservoirmedium ⇒ pricelow	12.18	25.40	-0.24
pricelow ⇒ bidreservoirmedium	12.18	26.40	-0.28
bidreservoirhigh ⇒ pricelow	26.70	41.50	0.37
pricelow ⇒ bidreservoirhigh	26.70	43.20	0.35
bidreservoirlow ⇒ pricemedium	13.10	32.30	0.35
pricemedium ⇒ bidreservoirlow	13.10	23.50	0.09
bidreservoirmedium ⇒ pricemedium	26.40	34.30	0.27
pricemedium ⇒ bidreservoirmedium	26.40	42.40	0.15
bidreservoirhigh ⇒ pricemedium	15.60	33.10	-0.33
pricemedium ⇒ bidreservoirhigh	15.60	36.20	-0.36
bidreservoirlow ⇒ pricehigh	6.20	27.10	0.21
pricehigh ⇒ bidreservoirlow	6.20	35.60	0.24
bidreservoirmedium ⇒ pricehigh	8.10	17.40	0.03
pricehigh ⇒ bidreservoirmedium	8.10	53.20	0.52
bidreservoirhigh ⇒ pricehigh	1.40	1.35	-0.89
pricehigh ⇒ bidreservoirhigh	1.40	3.72	-0.76

In general terms, the results of Table 5 show that there are no clearly distinguishable rules, evidenced in the lower values of *confidence* and *support*. However, within the rules that exhibit higher values of support and confidence (but not high enough), two rules seems to be interesting. The first rule is bidreservoirhigh => pricelow with confidence and certainty factor values of 41.5% and 0.37, respectively.

The second rule is bidreservoirhigh => pricelow with confidence and certainty factor values of 34.3% and 0.27, respectively. In conclusion, it seems that the hypothesis about the effect of bidding reservoirs on bidding prices is really *weak*, finding, at most, values of confidence around 42%. This quantification may suggest that at a disaggregated level by plants, the agents have other incentives different from hydrological conditions as can be supported by the great dispersion of data shown in Figure 7.

5. CONCLUSIONS

The main contribution of this paper is a methodology to reduce the search space over the continuous domain of bidding prices. The proposed representation of bidding functions as points in a multidimensional space, allows us to find bidding patterns by applying a clustering algorithm. As a result, a dramatic reduction over the search space of bidding strategies is achieved. This approach allows us to summarize the strategic bidding behavior of an agent, in a single point over a multidimensional space. In addition, this bidding behavior is no longer considered to be an *isolated* process for each generating plant, but as a *combined bidding strategy* over its diversified portfolio.

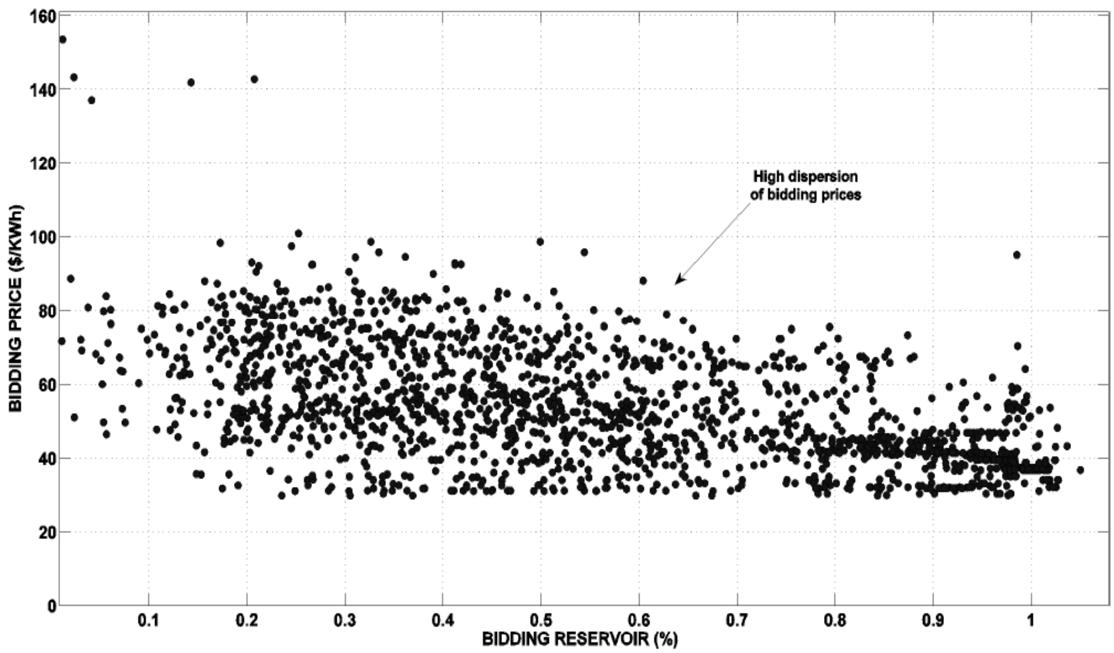

Figure 7. Relation between Bidding reservoir and bidding prices for a major hydro plant in the Colombian market

On the other hand, some relations of *dominance* over bidding strategies are found by establishing an ordering in terms of the frequency of use of these patterns. This ordering over bidding strategies may be the first step in the construction of some analytical models supported in empirical evidence (e.g., non-cooperative games with mixed strategies). Moreover, these patterns may be used in a multi-agent model of power markets that involves learning the capabilities of the agents, where the problem of dimensionality is always present [14].

Furthermore, an application of a data mining algorithm based on fuzzy association rules is presented. In this regard, the main contribution consists in a methodology to *quantify* the *confidence* and *truthfulness* of some hypotheses about the effect of hydrology on bidding prices. In particular, this methodology allows us to analyze rules over a partitioned domain of the variables involved. Consequently, some asymmetries in these hypotheses may arise from the association analysis, which gives more significance to the relations studied.

Specifically, the relationship between spot prices and national hydrology exhibits an apparent asymmetry: there is more confidence in the belief that "high bidding reservoir levels implies low spot prices" than in the belief that "high spot prices are caused by low bidding reservoirs levels." In addition, the association analysis on a disaggregated level shows that the hypothesis historically stated about the inverse relation of bidding prices and bidding reservoir levels is too weak in the Colombian case.

REFERENCES

[1] CREG006/2009 National Gas and Energy Regulatory Commission. January 30 2009.

[2] NIIMURA T, TAKASHIMA T, Deregulated Electricity Market Data Representation by Fuzzy Regression Models. IEEE Transactions on Systems, Man And Cybernetics—Part C: Applications and Reviews, Vol. 31, No. 3, August 2001. p 320-326.

[3] CHEONG M. P., SHEBLE G. B., BERLEANT D., Knowledge Extraction and Data Mining for the Competitive Electricity Auction Market. 9th International Conference on Probabilistic Methods Applied to Power Systems KTH, Stockholm, Sweden - June 11-15, 2006

[4] LU X, DONG Z, LI X, Electricity market price spike forecast with data mining techniques. Electric Power Systems Research, Volume 73. Issue 1, January 2005, p 19-29.

[5] YAN H, MA R, TONG X, SLIQ in data mining and application in the generation unit's bidding decision system of electricity market. International Power Engineering Conference 2005, IPEC 2005. December 2005. Singapore.

[6] SEBER G., *Multivariate Observations*. Wiley, 2004.

[7] HAN J., *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2001.

[8] ZADEH L., "Fuzzy sets," *Information and control*, vol. 8, pp. 338–353, 1965.

[9] SANCHEZ D., DELGADO M., "Fuzzy association rules: General model and applications," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 2, pp. 214–225, April 2003.

[10] SÁNCHEZ D., "Adquisición de relaciones entre atributos en bases de datos relacionales," Ph.D. dissertation, Department of Computer Science and Artificial Intelligence University of Granada Spain, 1999.

[11] MOLINA C., "Imprecisión e incertidumbre en el modelo multidimensional: Aplicación a la minería de datos," Ph.D. dissertation, Univ. de Granada, 2005.

[12] YOUNES C., DUARTE O., "Metodologías para la correlación de parámetros del rayo con características geográficas y meteorológicas caso Colombiano," Ph.D. dissertation, Universidad Nacional de Colombia, Junio 2006.

[13] SERRANO J. AND VILA M., "Mining web documents to find additional query terms using fuzzy association rules," *Fuzzy Sets and Systems*, vol. 148, no. 1, pp. 85–104, November 2004.

[14] GALLEGO L, DELGADILLO A., "Agent learning methodology for generators in an electricity market,". IEEE PES General Meeting, July 2008.

[15] GALLEGO L., "Modelamiento del Comportamiento de la oferta de la Energía Eléctrica en el mercado Colombiano," Ph.D. dissertation, Universidad Nacional de Colombia, Junio 2008.

[16] ARANGO S, DYNER I and LARSEN E, Lesson from the deregulation in South America: understanding hydro-based electricity markets. Utilities Policy, 2006, 14:196-207.

[17] LARSEN E, DYNER I, BEDOYA L and FRANCO C, Lessons from deregulation in Colombia: successes, failure and the way ahead. Energy Policy, 2004,32(15):1767-1780.

[18] DUARTE O, Técnicas difusas en la evaluación de Impacto ambiental. PhD. dissertation, Universidad de Granada, 2000.

[19] SMITH R, MONTOYA S, Modelamiento de la evolución de la oferta del sector eléctrico Colombiano, Revista Energética (19), 39-46.