Risk Analysis in Electricity Markets by using Decision Trees

N. Mosquera, J. Reneses, J. Barquín, E.F. Sánchez-Úbeda

***Abstract*—This paper describes a procedure for medium- and long-term risk analysis by using decision trees. A market equilibrium model is presented in order to assess the impact of the different sources of uncertainty. Decision trees are defined and applied in a study case, showing the advantages of these techniques for medium-term operation and planning. The paper analyzes different scenarios of five main risk factors: Hydro-inflows, fuel (coal and gas) costs, system demand, and CO_2 emission price.**

***Index Terms*— Decision Trees, Electricity Markets, Gas Options, Market Equilibrium, Risk Factors, Risk Management.**

I. Introduction

Competitive environments in electricity generation markets offer new opportunities and bring new risks to electricity companies. The companies are responsible for taking decisions that would affect their economic results. Taking into account new risk factors, some of the most notable mid-term decisions they must address are: the hydro-thermal coordination plans type, duration and coverage of signed contracts, strategies for risk hedging (e.g. options and forwards), etc. Therefore, new medium- and long-term models development is indispensable to take adequate decisions in order to maximize the expected company's profits.

Among the main variables subject to uncertainty, the system demand, hydro conditions and fuel costs are the most notable.

Arguably, the most theoretically rigorous and powerful method to model uncertainty is by using stochastic programming, at least if the uncertainty can be represented by a stochastic process. Meaningful robust decisions could be obtained for the short-term, as shown in [1]. However, it is a methodology with high computational cost, because of the exponential growth in size of the scenario tree, unless a simplified modeling of the driving stochastic process is included. However, this can be unacceptable in risk analysis applications, where it is often required to address low-probability scenarios of potentially huge impact.

Hence, this paper presents an alternative method to include uncertainty by using a methodology, in which several risk factors scenarios are simulated using a market equilibrium model. Outputs obtained by the different scenarios are analyzed using a decision tree, which provides useful statistic information.

Section II introduces the market-equilibrium model, which is solved through an equivalent optimization problem based on a conjectured price response approach. Risk factors models, and the definition and uses of decision trees, are the subject of Sections III and IV respectively. In order to show the advantages of the proposed methodology, a study case is analyzed in Section V. Finally, the main conclusions are summarized in Section VI.

II. Market Equilibrium Model

A market equilibrium is defined as a set of prices, generator outputs, consumption and other relevant numerical quantities, which any market agent could not modify unilaterally, without a decrease in its profit.

In this paper, market equilibrium is solves by using the model described in [2]. The model computes a conjectured-price-response equilibrium problem by means of an equivalent quadratic optimization problem:

$$\min_{P_{cp}} \quad \sum_{c=1}^{C}\sum_{p=1}^{P}\overline{C}_{cp}\left(P_{cp}\right)$$
$$\left[\begin{array}{llll} s.t. & \sum_{c=1}^{C}\left(P_{cp}\right)=D_p & : & \eta_p \\ & \text{Technical Constraints} & & \end{array}\right. \tag{1}$$

where P_{cp} represents the production of company c at period p, D_p is the demand of period p, C is the number of companies, P is the number of periods, and $\overline{C}_{cp}(\bullet)$ denotes a term called *effective cost function*:

$$\overline{C}_{cp}\left(P_{cp}\right)=C_{cp}\left(P_{cp}\right)+\frac{\left(P_{cp}\right)^2\cdot\theta_{cp}}{2} \tag{2}$$

being $C_{cp}(P_{cp})$ the cost function for company c at period p and θ_{cp} the conjectured price response; defined as the variation of the clearing price λ_p with respect to each firm production:

$$\theta_{cp}=-\frac{\partial\lambda_p}{\partial P_{cp}} \tag{3}$$

These conjectured-price responses are assumed to be known. The system marginal price can be computed from the dual variable of the power-balance constrain:

All the authors are with the Instituto de Investigación Tecnológica, School of Engineering, Universidad Pontificia Comillas, c/ Alberto Aguilera 23, 28015 Madrid, Spain. Corresponding author is N. Mosquera (e-mail: natalia.mosquera@iit.upco.es)

$$\lambda_p = \frac{\eta_p}{l_p} \tag{4}$$

being l_p the duration of period p.

III. RISK FACTORS

Risk analysis requires taking into account several sources of uncertainty, also called risk factors [3]. The five main risk factors considered in this paper are hydro-inflows, fuel costs (gas and coal costs), system demand and CO_2 emission price. All of them are assumed to follow independent stochastic process, which is a suitable approximation for medium-term studies.

Theoretically, different methods have been developed to generate random and logical scenarios for this kind of variables subject to uncertainty. Hydro-inflow, demand and fuel costs can be modeled by AR processes that include a GARCH component, relevant for fuel costs evolution modeling, as shown in [4].

Model estimation of the above mentioned risk factors uses historical information. However, this approach is not valid for the CO_2-emission-price risk factor, as emission trading has only been recently established on the EU. So, any CO_2 emission price evolution model must be based mainly on theoretical considerations. A possibility is to use the Black-Scholes model, as it is simple, robust and not mean-reverting.

However, in the study case of this paper, the methodology applied to risk factors scenarios generation has been simplified in comparison with the methods above described. The values of the scenarios for each risk factor have been estimated in order to obtain logical ranges.

IV. APPLYING DECISION TREES

In order to enhance the risk analysis task in electricity markets, we propose a methodology consisting of two main steps (Fig. 1): (i) data base generation of possible simulated scenarios, and (ii) data base analysis by using decision trees. Note that this approach has been applied successfully in other problems (e.g. [5][6]).

A. Generation and simulation of scenarios

The objective of the first stage is to obtain a sufficiently rich data base (DB), which both contains plausible cases and covers all relevant situations. These scenarios are randomly drawn from previous risk factor distributions. Thus, each case in the DB consists of a particular combination of risk factor scenarios and the equilibrium model results obtained according to section II.

Note that the CPU time required for simulating each scenario is not despicable. Thus, the overall task could be time-consuming if real-sized cases are simulated. For example, in the case study of section V, the simulation of each real-sized case requires about 5 minutes in a standard 2.8 GHz PC.

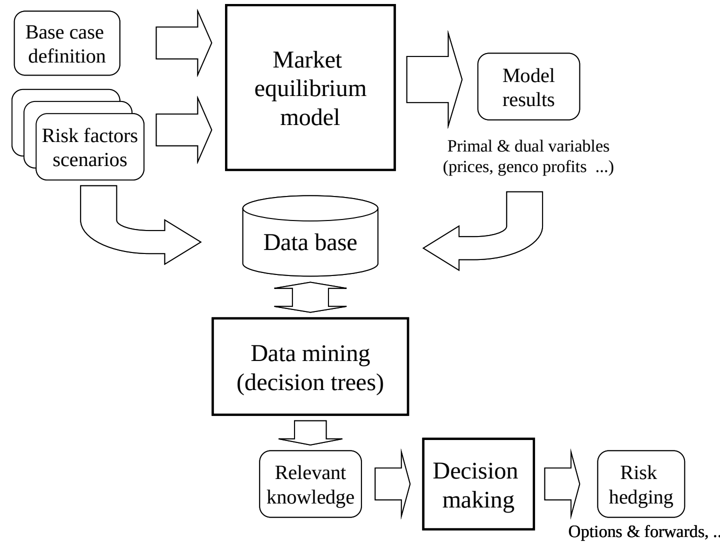

Fig. 1. Methodology proposed for risk analysis.

B. Analysis of scenarios by using Decision trees

Once the DB has been generated, decision trees are applied in order to extract relevant knowledge from the set of pre-analyzed cases. In particular, we are interested on the relationships between the output variables (spot prices and agent profits) and the input ones (hydro-inflows, fuel costs, system demand, CO_2 emission price, …).

In order to find this input-output mapping, several supervised learning techniques could be used, such as neural networks or decision trees [7]. However, decision trees are particularly well-suited for the problem at hand. The main reasons for selecting this technique are: the input-output relationship modeled by a decision tree is fully interpretable (i.e. it provides, if exists, direct knowledge), they can deal with both numerical (continuous) and nominal (discrete) input variables, the existence of high-level learning algorithms (i.e. very few parameters must be set for building a "good" decision tree), and very fast computer times for creating and evaluating a decision tree.

A standard decision tree models the input-output relationship by partitioning the input space in non-overlapped regions (hyper-cubes). Each region is bounded by a combination of one-variable partitions, and it has associated an output value. For example, the region given by "gas cost < 4.00" and "demand<300000" could correspond to a "low" company's profit (the output equals "low"). This is usually expressed in terms of rules like "if gas cost is lower than 4.00 and the demand is lower than 300000, then the company's profit is low", which shows the power of decision trees in terms of interpretability.

From a practical point of view, a decision tree, as stated in [7], is defined as a model where the extracted knowledge is sorted hierarchically by the use of connected nodes. These nodes can be of two types: Internal nodes (also called test or non-terminal nodes), and the terminal nodes (also called leaves). Internal nodes check the value of a single input variable and represent the forks of the tree, whereas terminal ones have not tests and they represent the outputs' value associated to each input region. These outputs' values are derived from the set of pre-analyzed cases contained in each region, usually by computing the majority class or the mean value.

There exist many variants of decision tree learning algorithms (see e.g. [8][9]). However, the typical strategy used to obtain the tree model is quite similar. A tree is obtained from the learning data by repeatedly partitioning the input space in smaller regions. This data-driven approach tries to obtain a partition where the diversity of the output values in each region is small enough. In this paper ID3 learning algorithm [6][7][9] has been used.

V. STUDY CASE

A. *Case description*

The study case represents an oligopolistic electricity market with a time scope divided in twelve periods (one year). Each period is supposed to be split into *5* load levels, according to the demand. Each generation company has its own hydro-thermal units, as shows Table I. This study case is based on the estimated 2006 Spanish electrical system.

TABLE I
INSTALLED POWER (MW) AND NUMBER (IN PARENTHESES) OF GROUPS OWNED BY GEN COS

Company	Nuclear	Coal	Gas	Hydro	Pumping
1	3506 (4)	5526 (16)	3070 (7)	3825 (12)	1390 (6)
2	3180 (4)	1167 (5)	4970 (11)	7036 (6)	2280 (4)
3	707 (1)	1888 (7)	3442 (9)	1654 (4)	216 (1)
4	156 (1)	1488 (5)	385 (1)	327 (1)	115 (1)
5	-	858 (5)	-	554 (2)	360 (1)
6	-	-	3435 (8)	-	-
7	-	109 (1)	4172 (9)	-	-

Besides, the study has considered a number of scenarios per each risk factor as follows:

1) *System demand:* In Table II the annual demand is represented in three different scenarios: low, medium and high. These scenarios have been calculated based on future expectations.

TABLE II
SCENARIOS OF ANNUAL DEMAND (GWH)

High	321745
Medium	316261
Low	310249

2) Hydro conditions: Following the same methodology, Table III shows three scenarios corresponding to dry, medium and wet situations, In particular, this table shows the total annual hydro inflows and run-off-the-river (ROR) used for each hydro scenario.

TABLE IIISCENARIOS OF THE ANNUAL HYDRO INFLOWS AND RUN-OFF-THE-RIVER (GWH)

	Hydro-inflows	Run-off-the-river
Wet	13591	27182
Medium	9514	19028
Dry	7557	15114

3) Fuel costs: Alike the last risk factors, three alternative scenarios (high, medium and low) have been defined in table IV for the two fuels considered: gas and coal. These values are the average costs of the groups that use gas and coal as fuels.

TABLE IV
SCENARIOS OF FUEL COSTS (C€/KWH)

	Gas	Coal
High	4.12	2.914
Medium	3.433	2.428
Low	2.747	1.943

4) CO_2 emission price: Based on recent statistic data, three different scenarios have been considered, as shown in Table V. These high, medium and low values have been proved as suitable in this recently CO_2 market.

TABLE V
SCENARIOS OF CO2 EMISSION PRICE (€/TON)

High	25
Medium	15
Low	20

B. Obtained results

1) Market equilibrium model results: As a result of considering three different scenarios per each risk factor, 243 equilibriums have been computed by the market model. The model has been coded in GAMS 21.6 language and solved by using the QP algorithm provided by the CPLEX 9.0.2 solver. Each execution takes about 5 minutes in a 2.8 GHz PC with Pentium IV processor and 504 MB of RAM memory.

The average electricity prices and one company's profits for each scenario combination have been obtained. Fig. 2 shows the histogram of these electricity prices. The minimum and maximum values are shown, as well as their frequency.

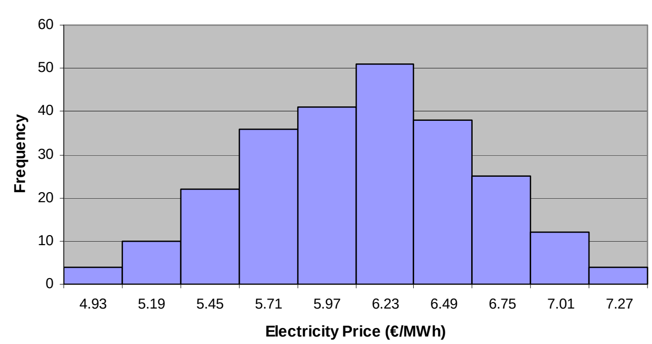

Fig. 2. Histogram of the annual average electricity price (c€/kWh)

Other interesting data that the study can provide is the company's operation profit. Fig. 3 shows the histogram of this variable for one of the biggest companies. Using this graphic, it is possible for the company to determine those ranges that suppose risky situations for its economic interests. Furthermore, histograms of the real output variables are of great value to define the ranges of the discretized output variables for the decision trees.

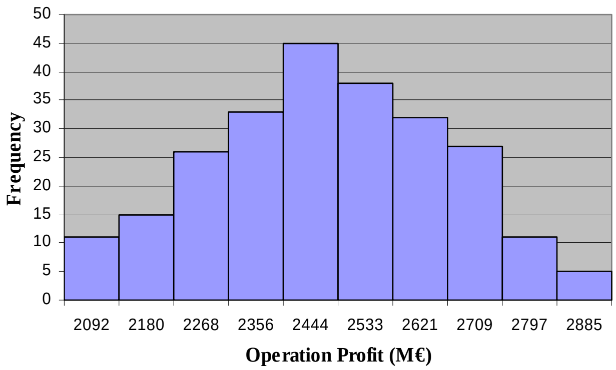

Fig. 3. Histogram of the company's annual operation profit (M€)

2) Decision trees: In order to build a decision tree, a set of candidate inputs should be selected to explain appropriately the output variable under study. In our particular case the candidate input variables are the total hydro inflows, the total system demand, the average gas costs, the average coal cost and the CO_2 emission price. These variables could be used to explain the two outputs of interest, which in this case are the electricity price and the profit of a particular company.

A decision tree per output is required to study each output's behavior. To carry out the assessment, each of the output variables has been discretized splitting them in three logical ranges. In particular, a new discrete electricity price variable has been created, whose values can be 'low' (which corresponds to real prices less than 5.7), 'medium' (prices within the range [5.7, 6.5], or 'high' (greater than 6.5). A similar partition has been applied to the company's operational profit, where 'low' means values less than 2342, 'medium' refers to values in [2342, 2639], and 'high' corresponds to real profits greater than 2639.

Fig. 4 and Fig. 5 show the decision trees obtained for the (discretized) annual average of the electricity price and the company's annual operation profit, respectively. Note that the decision tree learning algorithm determines automatically in each node the test '*var* < *thd?*', where *var* is the selected variable and *thd* is the optimal threshold obtained from data. Depending on the answer, the right or the left child of the current node will be visited.

To illustrate the decision trees, we have used a graphical representation based on the notation proposed by Wehenkel [6]. Basically, each node is represented by a box, showing the proportion of examples of each class relative to this node (lower part), as well as the proportion of correct classifications of test examples for the corresponding sub tree (upper part). Moreover, the size of the node represents the proportion of learning examples in that node (the root has the larger size).

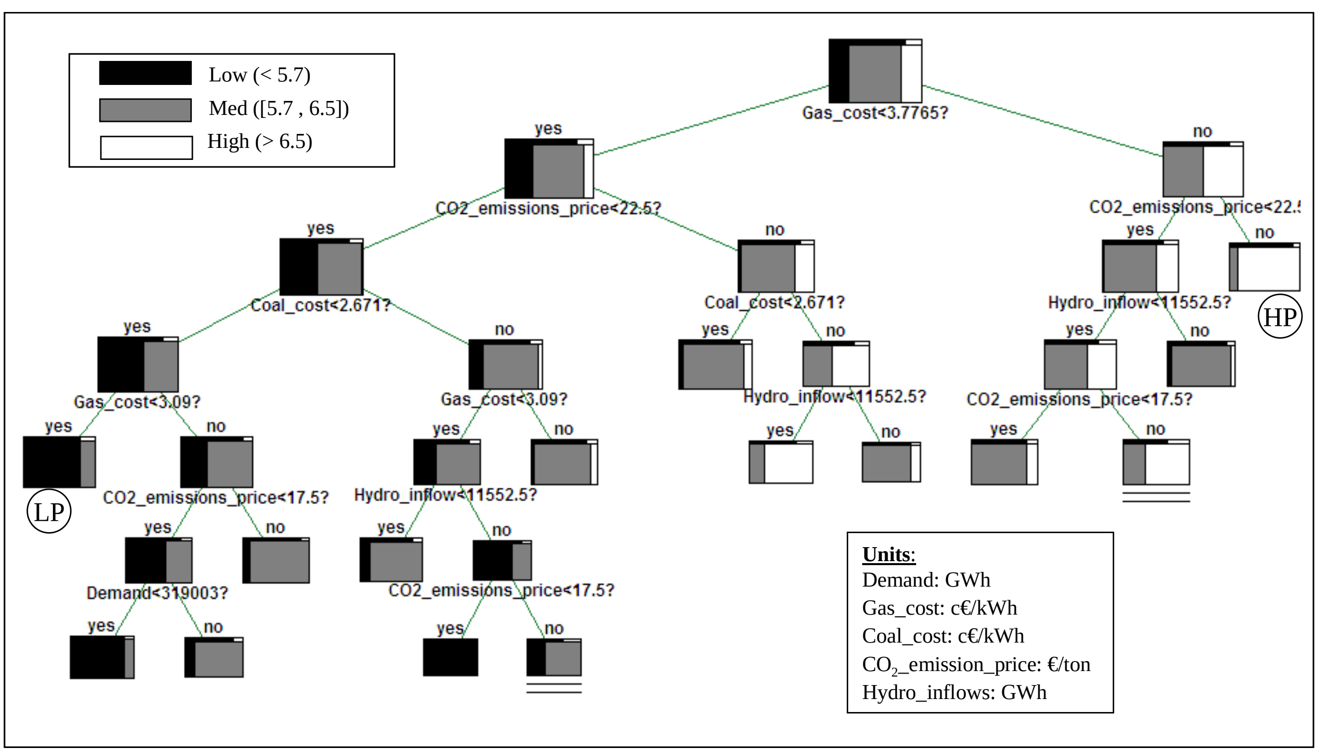

Fig. 4. Decision tree of the electricity price (c€/kWh)

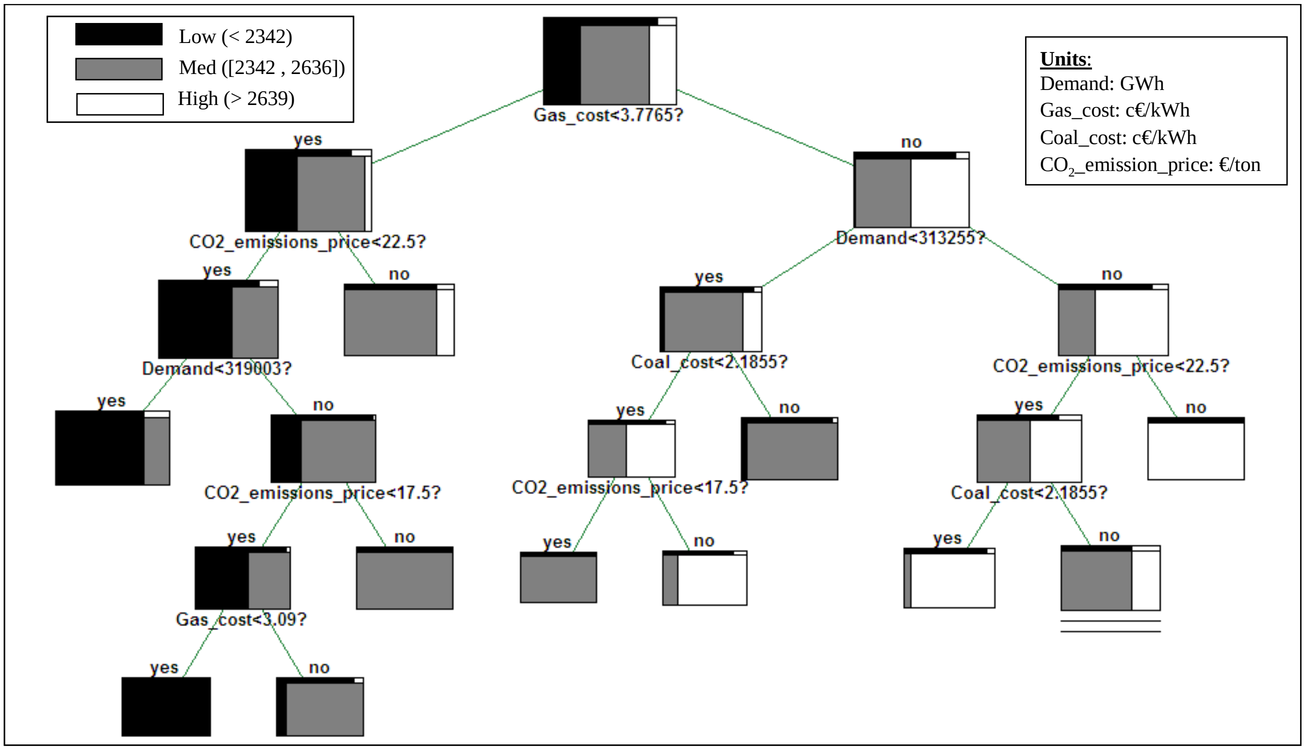

Fig. 5. Decision tree of the company's annual profit (M€)

Decision trees shown in Fig. 4 and Fig. 5 are both simple and accurate, i.e., having a small number of nodes; they are able to explain correctly more than 85% of the considered scenarios. Note that the resemblance among these decision trees yields to the fact that there is a notable correlation among the electricity price and companies' profit. Fig. 6 shows this strong dependence .The profit depends strongly on the price, but, for a given price, there exists a large uncertainty due to factors such as fuel costs.

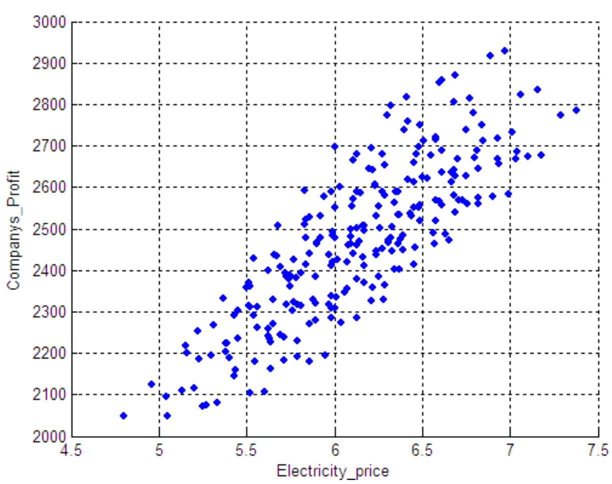

Fig. 6. Relationship between company's profit & Electricity price

Let us discuss the physical interpretation of the obtained trees. According to Fig. 4, the main factors influencing the electricity price are the gas cost, the CO_2 emission price and the coal cost. For example, node labeled with "HP" (high prices) corresponds to scenarios where the gas cost is high (>3.77) and the CO_2 emission price is also high (>22.5). On the other hand, if gas cost is low (<3.09) and the CO_2 emission price is not high (<22.5) and the coal cost is low (<2.67), then the electricity price is low (node labeled with "LP").

The decision tree of the company's annual profit (Fig. 5) throws interesting knowledge about the factors related to low and high profits. For example, if gas cost is high (>3.77) then the profit is not low (>2342). Furthermore, there exist a significant number of low-profit scenarios characterized by non-extreme values of gas costs, CO_2 emission price and demand. In other words, this company has significant losses when the demand is not high (<319003), the CO_2 emission price is not high (<22.5) and the gas cost is not high (<3.77).

Finally, note that the knowledge acquired from decision tree of Fig. 5 can be used to decide the correct strategy in order to hedge against risk. In particular, situations where the operation profit is low can be avoided.

3*) Risk hedging*: Once defined the gas price limit, that can lead to lower operation profits, it is possible to reduce the risk incurred by using some risk mitigation mechanisms.

One of them is a gas option. The use of oil or gas options can provide the project company with the hedge needed, and leave it with the full benefits associated with a favorable development of the commodity price. Fig. 7 shows the pay off of a gas option example:

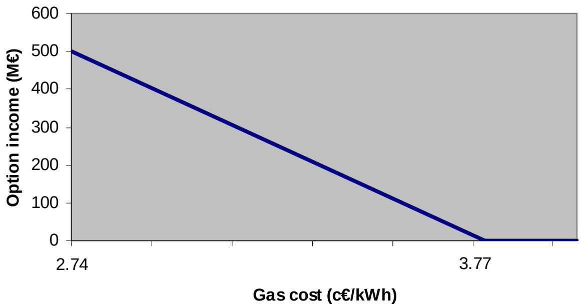

Fig. 7. Gas option income

Including the price and possible benefits of this particular gas option the new expected company's profit would be:

$$p' = p + oi(gp) - op \qquad (5)$$

Where p is the expected company's operation profit without considering the gas option, the option income oi is a function of gas pricing gp, (as seen in Fig. 7), and the last constant is the option premium(price) op.

In order to quantify the above option, it is assumed that the company acquires a portfolio of monthly put options, with a total notional of 500 M€ and a strike price of 3.77 c€/kW-h. It is furthermore assumed that gas price follows a Black-Scholes process [10] with volatility of a 20% per year, and an initial price of 3.43 c€/kW-h. Short-term interest rate is assumed to be 0. Direct application of Black-Scholes formula provides a total option price (i.e., op) equal to 204 M€.

Fig. 8 shows the decision tree of the company's annual profit including the option gas of Fig. 7. This tree is as robust as previous ones, explaining correctly more than 85% of the 243 scenarios. It can be observed how the gas option has hedged part of the risk of incurring in low benefits for the company. In this new situation, the root node (Gas_cost<3.09?) divides the decision tree in two main branches where the low gas cost generates the higher profits. In this new situation the lowest operation profit appears when the gas cost is not low (>3.09), the CO_2 emission price is not high (<22.5), and the demand is low (<313255). To make even less risky the new situation, other mechanisms can be applied, for instance, invest on options over other fundamentals.

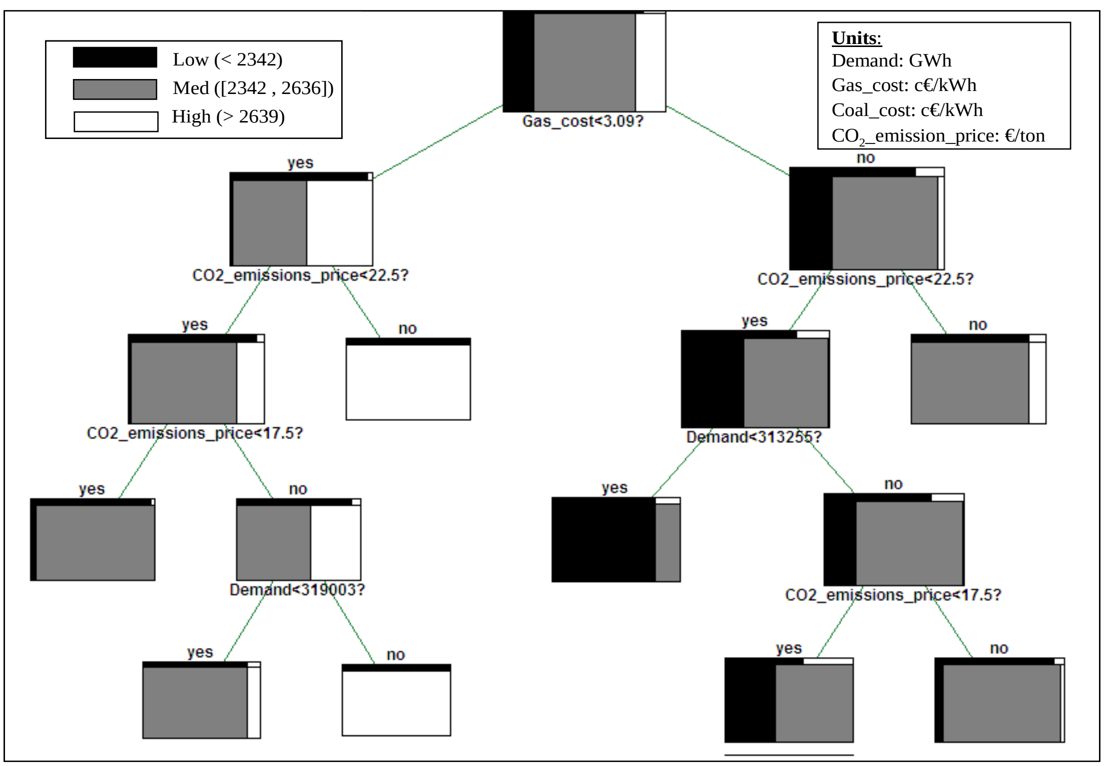

Fig. 8. Decision tree of the company's annual profit including an option gas (M€)

VI. CONCLUSIONS

Electricity markets offer new opportunities and also new risks. Therefore a risk analysis has to be carried out and medium-term decisions have to be taken to help achieve economic targets and risk hedging.

It has to be taken into account that risk analysis requires to consider several sources of uncertainty. The main ones are: hydro-conditions, fuel costs, system demand.

It is necessary to develop models that include the factors subject to uncertainty to help decision making. One of the methodologies for the risk analysis is decision trees.

A real-sized study case is presented in order to show the main advantages of the approach, mainly the use of a market equilibrium model and decision trees to obtain significant computational savings. This study case represents an oligopolistic electricity market, based on the estimated 2006 Spanish electrical system. The studied variables subjected to uncertainty (coal cost, gas cost, CO_2 emission price, hydro inflows and demand) lead to risk scenarios for the electrical companies.

Decision tree approach helps to gain information for wise decision making and to hedge against risk. It helps to identify the values of the studied variables that produce risk on the electricity price and in the operation profit (the considered outputs in the study case) and therefore help the electrical companies on the decision making.

It can also be inferred that if the number of considered scenarios for each risk factor is higher, then the number of computed executions by the market equilibrium model would be higher. Therefore the group of training data that feed the decision tree would be higher and diverse leading to more reliable results.

Over time many mechanism have been created for risk mitigation. The main transactions used can be classified into two mayor groups: forwards-based derivatives and options-based derivatives. In the study case, investing in gas options has been proved as an alternative that diminish the risk of low profits.

VII. REFERENCES

[1] J. Barquin, E. Centeno, J. Reneses, “Stochastic market equilibrium model for generation planning computing,” *Probability in the Engineering and Informational Sciences.* Vol. 19. pp. 533-546, October 2005.

[2] J. Barquin, E. Centeno, J. Reneses, “Medium-term generation programming in competitive environments: A new optimization approach for market equilibrium computing,” in *IEE Proceedings – Generation, transmission and distribution.* Vol. 151(1). Pp. 119-126, January 2004.

[3] C. Batlle, "A model for electricity generation risk analysis," PhD Dissertation, Department of Industrial Organization, Universidad Pontificia Comillas, Madrid, Spain, 2002.

[4] C. Batlle, J. Barquin, “Fuel prices scenario generation based in a multivariate GARCH model for risk analysis in a wholesale electricity market”, *in Proc. of the 7th International conference on Probabilistic Methods Applied to Power System (PMAPS)*, Naples, 2002.

[5] E. F. Sánchez-Úbeda, J. Peco, P. Raymont, T. Gómez, S. Bañales, A. L. Hernández, “Application of Data Mining Techniques to Identify Structural Congestion Problems under Uncertainty,” *in Proc. IEEE Power Tech Conference , Porto*, Portugal, September 2001.

[6] L. Wehenkel, *Automatic Learning Techniques in Power Systems*, Kluwer Academic, Boston, 1997.

[7] E. F. Sánchez-Úbeda, "Models for data analysis: contribution to automatic learning," PhD Dissertation, Department of Electrotecnic and Systems, Universidad Pontificia Comillas, Madrid, Spain, 1999.

[8] J. R. Quinlan, “Learning decision trees classifiers”, *ACM Computing Surveys*, pp. 71-72, March 1996.

[9] J. R. Quinlan, “Induction of decision trees”, *Machine learning*, pp. 81-106, 1986.

[10] J. C. Hull, *Options, futures, and other derivatives*, Prentice Hall., Third edition, pp. 572. New Jersey, U.S.A, 1997.

VIII. BIOGRAPHIES

Natalia Mosquera received a degree in Industrial Engineering (2004) and a degree in Industrial Organization Engineering (2004) at the Universidad Pontificia Comillas in Madrid, Spain.

Currently, she is a Researcher at the Technological Research Institute [*Instituto de Investigación Tecnológica*]. Her areas of interest include planning of power systems, international electricity markets and operational research.

Javier Reneses received a degree in Industrial Engineering (1996) and a PhD in Industrial Engineering (2004) at the Universidad Pontificia Comillas in Madrid, Spain, and a degree in Mathematics from the Open University in Spain [Universidad Nacional de Educación a Distancia (UNED)] in 2005.

At present, he is a Research Fellow at the Technological Research Institute [*Instituto de Investigación Tecnológica*]. His areas of interest include operation, simulation models and planning of electric energy systems and risk management strategies in electricity markets.

Julián Barquín (M’94) obtained a degree in Industrial Engineering (1988) and a PhD in Industrial Engineering (1993) at the Universidad Pontificia Comillas, Madrid, and a degree in Physics in 1994.

Currently, he belongs to the research staff at the Technological Research Institute [*Instituto de Investigación Tecnológica*]. His present interests include control, operation and planning of power systems.

Eugenio Fco. Sánchez Úbeda received a degree in Industrial Engineering (1991) and a PhD in Industrial Engineering (1999) at the Universidad Pontificia Comillas, Madrid.

He is a Research Fellow at the Technological Research Institute [*Instituto de Investigación Tecnológica*]. His areas of interest include data mining, Artificial Intelligence and its applications.