

Generation of Realistic Scenarios for Multi-Agent Simulation of Electricity Markets

Francisco Silva¹, Brígida Teixeira¹, Tiago Pinto^{1,*}, Gabriel Santos¹, Zita Vale¹,
Isabel Praça¹

¹ *GECAD - Knowledge Engineering and Decision Support Research Centre - Polytechnic of Porto (IPP)
R. Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal*

ABSTRACT

Most market operators provide daily data on several market processes, including the results of all market transactions. The use of such data by electricity market simulators is essential for simulations quality, enabling the modelling of market behaviour in a much more realistic and efficient way. RealScen (Realistic Scenarios Generator) is a tool that creates realistic scenarios according to the purpose of the simulation: representing reality as it is, or on a smaller scale but still as representative as possible. This paper presents a novel methodology that enables RealScen to collect real electricity markets information and using it to represent market participants, as well as modelling their characteristics and behaviours. This is done using data analysis combined with artificial intelligence. This paper analyses the way players' characteristics are modelled, particularly in their representation in a smaller scale, simplifying the simulation while maintaining the quality of results. A study is also conducted, comparing real electricity market values with the market results achieved using the generated scenarios. The conducted study shows that the scenarios can fully represent the reality, or approximate it through a reduced number of representative software agents. As a result, the proposed methodology enables RealScen to represent markets behaviour, allowing the study and understanding of the interactions between market entities, and the study of new markets by assuring the realism of simulations.

KEYWORDS: Data-Mining; Electricity Markets; Knowledge Discovery; Machine Learning; Multi-Agent Simulation, Scenarios Generation.

* Corresponding author: Tiago Pinto is with GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, R. Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal; Tel.: +351 22 8340500; Fax: +351 22 8321159, Website: <http://www.gecad.isep.ipp.pt> / E-mail: tmcfp@isep.ipp.pt

1. Introduction

The restructuring of Electricity Markets (EM) has created a new design of the electricity sector, becoming a new global trend. The monopoly model gave way to a non-vertical structure that allows the development of a free competition market, as it allows the liberalization of potentially competitive segments, such as production, transportation and distribution [1]. These changes provide conditions for the establishment of a more competitive market but also more complex, where the laws of supply and demand stimulate the quality of service, reduce costs and increase efficiency. The growth of complexity creates new challenges, namely the difficulty in decision making due to markets unpredictability and the high number of associated variables [2].

Simulation started to play a key role in supporting decision-making in the electricity sector. It allows studying and analysing the profiles and strategies of market participants and, thus, model market behaviour and forecast proposals [3]. Such characteristics enable the establishment of conditions conducive to decision making, turning the simulators into essential tools to support the necessary decisions on major investment environments.

Simulation tools applied to EM are designed to help market entities to respond efficiently to market unpredictability. Each one of the surrounding entities may present different behaviours resulting from the context and its goals, having an immediate impact on the results obtained in the market [4]. Therefore, simulators allow to perform a study of new types of participating entities, their behaviour, market models, types of market trading mechanisms, among others, to understand the impact of these variables on system performance and, therefore, to react in order to achieve the best possible results.

Although multi-agent simulation tools present many advantages, most are not able to present results that can be applied to reality [5]. In order to deal with this problem, it is necessary to create realistic scenarios that consider real data about the behaviour and characteristics of market participants as well as the specifications of the markets [6,7]. Although data is available on the website of the different market players (e.g. online platform of Nord Pool market operator [8]), the extraction of these data is a very laborious procedure due to their size, diversity and availability mode. When this problem is surpassed, the real challenge arises: the analysis and extraction of useful information based on these data so that it becomes possible to create realistic scenarios and thereby improve the performance of multi-agent simulators.

This paper presents a novel methodology that generates realistic scenarios for multi-agent simulation of EM. Artificial Intelligence (AI) and Data Mining (DM) techniques are used to perform data analysis and extract the

information needed for modelling the characteristics and behaviour of market participants. In order to simplify the simulation, reducing the required computational effort, when necessary, an adjustment of the size of the desired scenario is performed, using fuzzy logic and clustering techniques. The fuzzy logic process determines the number of representative agents of the entities participating in the market, while the clustering approach decides how to divide the players by the determined number of representative agents, taking into account each of their characteristics and respective importance. Artificial Neural Networks (ANN), Support Vector Machines (SVM) and a Simple Average technique are also used in order to forecast, based on the history of the player, the price and quantity of energy that each market player will negotiate in the market. In addition, there is a VPP (Virtual Power Player) creation mechanism that allows agents to form coalitions, by grouping together in order to increase their competitiveness in the market.

The proposed scenario generation model is integrated in RealScen (Realistic Scenario Generator), a tool that is directed to supporting the creation of realistic simulation scenarios for different electricity markets multi-agent simulators. A very important contribution to this tool's development was one of the previous works that is related to a tool able to handle, store and dynamically update real electricity markets data [9]. With access to that data, it was possible to develop a preliminary version of RealScen that was focused in the data aggregation component [5]. The next step was to evolve the tool in order to make it more complete and accurate, which resulted in a new version, presented in [10]. One of the main aspects of the new release was the inclusion of a graphical user interface that eases the whole process of creating a scenario.

After this introductory chapter, chapter 2 presents a contextualization of EM – the target environment that is modelled by the proposed method, including the presentation of the main models, with special focus to the Day-ahead Spot Model. Chapter 2 also introduces some of the main multi-agent EM simulators. Chapter 3 describes the proposed scenario generation model, and the RealScen tool. This includes the model architecture and operating mode, where the main steps in the generation of a scenario are defined. Chapter 4 presents a case study that demonstrates the quality of results of the proposed model, by presenting a comparison between the scenarios generated by the proposed model and the actual market results for the same days. Finally, chapter 5 presents the main conclusions of the developed work.

2. Electricity Markets and Simulation

The conjuncture in the electricity sector has been changing over the years. Currently, most countries have their own market or participate jointly with other countries in a regional markets [2]. Each EM is usually composed by several market types, which are usually made from three basic models: Day-Ahead Spot, Intraday or Balancing Market and Bilateral Contracts model. The differences between them lie under their rules, objectives and how they operate. The Day-Ahead Spot Market aims to negotiate energy for the following day, the Intraday/Balancing Market allows players to negotiate almost in real-time in order to adjust their needs and compensate possible forecast errors. Negotiation by means of Bilateral Contracts is based on trading power for an extended time horizon, directly between players [11].

The Spot Market is usually where the larger amount of energy is traded, and it is characterized by an enhancement of competitiveness, given that the likelihood of selling in this market increases when the bid value is lower, which gets more buyers to compete in the bidding. This process lowers the selling prices and motivates buyers to submit bids with higher values to increase their chances of buying as well.

2.1. Electricity Market Simulators

Due to the great complexity and dynamism inherent to EM and to their constant evolution, the use of multi-agent simulators applied to EM is increasing. There are several simulators available, each one of them with their own goals and limitations. The market regulators are especially interested in this kind of tools in order to test new rules and detect inefficiencies that are occurring in the system. On the other hand, the players also benefit from these tools because they allow the study of players' behaviour, in order to maximize their profit [12]. Some of the most relevant simulators are EMCAS (The Electricity Market Complex Adaptive System) [13], AMES (Agent-based Modelling of Electricity Systems) [14], MASCEM (Multi-agent Simulator for Competitive Electricity Markets) [15, 16] and MAN-REM (Multi-Agent Negotiation and Risk Management in Electricity Markets).

The Electricity Market Complex Adaptive System (EMCAS) [13] uses an agent based approach with agents' strategies based on learning and adaptation. Different agents are used to capture the restructured markets heterogeneity, including generation, demand, transmission, and distribution companies, independent system operators, consumers and regulators. It allows undertaking Electricity Markets simulations in a time continuum ranging from hours to decades, including several Pool and Bilateral Contracts markets.

Agent-based Modelling of Electricity Systems (AMES) [14] is an open-source computational laboratory for the experimental study of wholesale power markets restructured in accordance with U.S. Federal Energy Regulatory Commission (FERC)'s market design. It uses an agent-base test bed with strategically learning electric power traders to experimentally test the extent to which commonly used seller market power and market efficiency measures are informative for restructured wholesale power markets. The wholesale power market includes independent system operator, load-serving entities and generation companies, distributed across the busses of the transmission grid. Each generation company agent uses stochastic reinforcement learning to update the action choice probabilities currently assigned to the supply offers in its action domain.

Simulator for Electric Power Industry Agents (SEPIA) [17] is a Microsoft Windows platform oriented simulator. It is based on a Plug and Play architecture, allowing users to easily create simulations involving several machines in a network, or in a single machine, using various processing units. SEPIA allows specifying the number of participating agents, as well as their behaviours, interactions, and changes during the simulation. The simulation developments can be followed and oriented by mechanisms for that purpose.

Power Web [18] is a Web-based market simulator, allowing the various participants to interact from very distinct zones of the globe. It is a rather flexible system that allows the definition of simulations with a large set of scenarios and rules. This simulator includes a centralized agent, acting as an independent system operator, to guarantee the reliability of the system, according to a defined group of entities, acting in various markets. It also allows users to enter an open market, competing against producers controlled by other users or computational algorithms.

The Short – Medium run Electricity Market Simulator (SREMS) [19] is based on game theory and is able to support scenario analysis in the short-medium term and to evaluate market power, in some situations. Some of its main features are: short-medium run (a time horizon of one month or multiples of it) simulation of electricity markets based on game theory, calculating price-makers optimal hourly bids; inelastic load, defined hour by hour and zone by zone; tree-like network with interzonal transit limits; monthly scheduling of reservoir hydro pumping storage plants; highly realistic representation of thermal plants; possible quota appointed to physical bilateral contracts, depending on producers share and risk attitude. It is particularly suitable to study the Italian electricity market.

The Genoa Artificial Power Exchange (GAPEX) [20] is agent-based framework for modelling and simulating power exchanges. GAPEX is implemented in MATLAB and allows the creation of artificial power exchanges reproducing exact market clearing procedures of the most important European power-exchanges.

MASCEM is a multi-agent tool that enables the study of the dynamics and complexity of EM, in competitive environments [16, 21]. For this, its participating entities are modelled, including Producers, Consumers, Market Operators and ISO, as well as their behaviour, characteristics and objectives. MASCEM is able to simulate the Day-Ahead Spot Market, Balancing Market and Bilateral Contracts. Important examples of markets at an European level that are supported by MASCEM are MIBEL – Iberian market [6], EPEX – central European market [7], and Nord Pool – northern European market [8]. Each market includes different variations in market mechanisms, such as the possibility of presentation of complex offers, in the case of MIBEL; possibility of submitting block offers in EPEX market; and flexible offers in the Nord Pool market. In order to enhance the simulation capabilities, MASCEM is integrated with ALBidS (Adaptive Learning for strategic Bidding System), which supports the behaviour of agents in their negotiations so that players are able to adapt their strategic behaviour by learning from previous transactions [22]. Further information about the way MASCEM's agents interact and how the market mechanisms deal with the players' behaviour, is presented in [23].

MAN-REM is focused on bilateral contracts and respective risk management [24]. It allows the simulation of various types of bilateral contracts (futures contracts, forward contracts, option contracts and contracts for difference) and is currently being integrated with MASCEM in order to form a more complete simulation tool. The forward contracts are energy trading contracts between a buyer and a seller in which they agree to exchange a certain amount at a future date, with trading price negotiated at the time. This type of trading has a great uncertainty involved given that at the agreed time for exchange, the agreed price may be lower or higher than the current price of energy. The futures contracts have the same behaviour as forward contracts, with the exception of being managed by a third party responsible for ensuring compliance with the agreement. Option contracts are similar to the previous ones, with the difference that the two entities only guarantee an option to purchase / sell. The exchange is carried out in the stipulated time if the two sides intend to exercise their option trading. Contracts for difference are contracts that allow the protection of the two entities involved from energy price change between the agreement establishment date and the agreed date of trade.

These simulators are important contributions; however, all of them share a common limitation: the inexistence of an automatic realistic simulation scenarios generator. This means that simulation scenarios used by these simulators are often created using past, outdated data, which takes a large amount of time and effort to treat and make available for large scale scenarios with a large number and diversity of different players and market rules. An automatic methodology that enables real data to be extracted in real time from EM operators' websites, and then be treated and transformed into realistic simulation scenarios, is highly valuable to allow simulators to become able to provide realistic EM simulation capabilities, with highly updated data. RealScen is the scenarios generator tool that allows the automatic generation of such scenarios, by using the methodology proposed in this paper, which is presented in section 3.

3. Realistic Scenarios Generation

RealScen (Realistic Scenarios Generator) is a tool developed with the purpose of automatically creating realistic EM simulation scenarios, using real data. The simulation scenarios are characterized by a list of bids to buy/sell energy, with the specification of the players (who sell or purchase), including their characteristics, quantity in MWh and the price in €/MW. This software, when combined with a multi-agent simulator of EM, becomes a powerful tool for obtaining market models that, through the use and processing of data provided by various MO's, are highly representative of the reality. RealScen has been developed in JAVA language and its results can be presented in XML and / or XLS formats, in order to be easily used by any EM simulator.

During the scenario creation process, RealScen uses several AI and DM methodologies to treat data, in order to suit it to the goals of each scenario (and simulation). These methodologies include clustering [25], classification [26], forecasting [27] and fuzzy logic [28].

RealScen has the following features for the purpose of facilitating the integration with any EM multi-agent simulator:

- Independence. Although the results produced by RealScen can be integrated by a market simulator, the scenario generation does not depend on a simulator. The application itself performs all necessary operations in its intermediate steps.
- Integration. The fact that the results are stored in generic formats allows the tool to be used by any simulator in a simple way. In order to use these results, the simulator should only include a small program translator / reader of the used formats and its data structure.

- **Adaptability.** The scenario is generated according to the purpose of the simulation; specifically, the user can set the scenario using data from different markets, even if different from the specific EM for which the scenario is intended. There is a wide range of values which the user may choose to customize in order to achieve a more targeted setting to its goals.
- **Succinctness.** The program is provided with clustering mechanisms that reduce the amount of data to be used in the scenario without losing the quality of results. This feature allows faster processing by the simulator by reducing the computational effort of large scale simulations, and simplify the study of markets, including new rules and mechanisms.
- **Forecasts by history.** When performing a scenario referent to a certain date, RealScen includes forecasts of both the bid price and quantity of energy of each agent, based on its historical offers.

Figure 1 presents a general overview of RealScen, including the proposed intelligent methodologies that allow the automatic generation of realistic simulation scenarios.

RealScen architecture

Fig. 1.

As can be observed in Figure 1, RealScen considers five steps for generating a scenario: Data Selection, Market Definition, Forecasting, Clustering and VPP's Creation. The Data Selection phase is characterized by selecting the data source to be used (section 3.1); Market Definition comes down to the choice of markets, market types and the date that the scenario is intended to represent (section 3.2); The Forecasting step is composed by the process of forecasting players offers after setting forecast parameters according to the objectives (section 3.3); the Clustering phase refers to the grouping of players in smaller samples according to their similarities (section 3.4); VPP's Creation is an optional step consisting on the creation of players coalitions in order to be able to increase their competition power (section 3.5).

3.1. Data Selection

The data used to generate scenarios is automatically collected from the website of various MO, specifically the MIBEL, EPEX, Nord Pool among others, and stored in a database [29]. These data refer to real transactions, covering players' offers and their characteristics. So when a scenario creation is started, the user must indicate

the source of information that is desired for the generation of the scenario. The used data may also cover market types of different MO (e.g. Day-Ahead from Nord Pool, Day-Ahead from EPEX and Intraday from MIBEL). This option allows the consideration of data from other scenarios and studying particular cases, allowing performing experiences considering hybrid markets with data from different sources.

Still at this stage, the user can choose the initial date of the data to consider. This step is important, since the forecast of amount of energy and price is made through the history and therefore the user may wish to consider extended periods of time or not. The amount of data used at this stage will have a direct impact on the forecasting process because the methods have to process all the values chosen for each player.

3.2. Market Definition

The Market Definition phase considers the configuration of the market scenario, i.e. dates, markets and respective market types that the scenario is intended for. Similarly to the data origin, the scenario may have as objective various market types from distinct markets. Chosen markets do not need to match the setting made in the Date Selection phase, i.e. the user can, for instance, use data from the Day-Ahead Spot market of Nord Pool, to run it using the Day-Ahead market of MIBEL. This allows studying what would happen if different market types, for instance from Iberia, were implemented in Northern Europe. This allows enables the creation of hybrid markets, including market types from different market operators, e.g. Day-Ahead market from MIBEL, Intraday market from EPEX and pool for Bilateral Contracts from Nord Pool. Such experiences are very useful for studying and assessing the impacting of different possible market configurations.

3.3. Forecasting

Forecasting is the first step to generate the scenario. For this, it is necessary to supply the date for which the scenario will be generated. Considering this date, RealScen will analyse the data loaded in the Data Selection phase and forecast the player's behaviour for the day in question, i.e., obtain a price and amount of energy for each period of the desired date.

RealScen has three mechanisms to perform the forecasts: simple average, ANN (Artificial Neural Networks) [30] and SVM (Support Vector Machines) [31]. The use of each of the three alternatives depends on the user preferences on the balance between execution time and quality of the results. In this way, forecasts (and thus,

the creation of the scenario) can be executed very quickly, but with a certain loss in results' accuracy; or on the other hand, require a longer execution time, but provide a much more accurate quality of results.

3.3.1. Artificial Neural Network

The considered ANN [30] is characterized as a feedforward neural network, receiving as inputs the market prices and total amount of negotiated energy in the market, referring to: the day before the desired forecasted day, one week before, two weeks before, and three weeks before. The ANN considers four nodes in the intermediate layer, and one output – the forecasted market price. This topology has been presented in [30], and has been defined after some analysis and experiences. The structure of this ANN is presented in Figure 4.

ANN topology

Fig. 2.

Backpropagation using the gradient descent method [31] has been used as training algorithm for the ANN. This requires calculating the derivative of the squared error function with respect to the weights of the network. The squared error function E for the single output neuron is defined as in (1).

$$E = \frac{1}{2}(t - y)^2 \quad (1)$$

where t is the target output for a training sample, and y is the actual output of the output neuron.

For each neuron j , its output o_j is defined by feedforward calculation, as in (2).

$$o_j = f\left(\sum_{k=1}^n w_{kj}x_k\right) \quad (2)$$

where n is the number of input units to neuron j , and w_{kj} is the weight between neurons k and j . Hence, the input for the activation function f of a neuron is the weighted sum of outputs o_k of the previous neurons. The used activation function f is the logistic function, a log-sigmoid function, which can be defined as in (3).

$$f(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

The backpropagation algorithm is used as the training method of the designed artificial neural network. The backpropagation algorithm includes the following steps [27]:

1. Initialize weights as small random numbers;

2. Introduce training data to the NN and calculate the output by propagating the input forward through the network using (2);
3. Calculate the error using (1)
4. Propagate the sensitivities backward through the network by simply taking the derivative of the activation function (3) with respect to the network parameters;
5. Calculate w_{kj} updates;
6. Update the values of w_{kj} ;
7. Repeat steps 2 to 6 until all examples are classified correctly.

Since this ANN provides only one price as output, related to a single negotiation period of a day, and days usually have several periods (e.g. 24 periods, when these are hourly), it is necessary to execute this ANN several times to forecast the whole day, thus potentially leading to a high execution time for some utilization contexts.

3.3.2. Support Vector Machines

Similarly to the ANN, which considers each trading period independently of the others, the considered SVM assumes the same principle [32]. This means that in the price forecast for a certain day and time period, the SVM will use only historical data for the same time period in the training process. The used historical data relate to the days immediately preceding the date in question for which the forecast is required. The number of previous days for which data is used in the SVM training process is determined by the *Training Limit* variable. This means that, if the value of *Training Limit* is e.g. equal to 7, the prices of the desired period of every day of the previous week will be used as training of SVM. The information to use in an SVM must follow the format suggested in (4):

$$(y_1, x_1), \dots, (y_l, x_l), x \in R^n, y \in R, \quad (4)$$

where each example x_i is a space vector example; y_i has a corresponding value; n is the size of training data. For classification: y_i assumes finite values; in binary classifications: $y_i \in \{+1, -1\}$; in digit recognition : $y_i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 0\}$; and for regression purposes, y_i is a real number ($y_i \in R$).

After exhaustive preliminary testing, it has been found that the most applicable kernels for the specific problem approached in this paper (forecast of electricity market prices as basis for the decision support in an

electricity market negotiation process) are the Radial Basis Function (RBF) and the Exponential Radial Basis Function (eRBF). Although these two kernels can be applied to both classification and regression problems, they show their best potential when directed to regression in time series data. RBF have received significant attention by the scientific community, most commonly with a Gaussian of the form (5).

$$K(x, y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \quad (5)$$

Classical techniques utilizing radial basis functions employ some method of determining a subset of centres. Typically a method of clustering is first employed to select a subset of centres. A smart feature of the SVM is that this selection is implicit, with each support vectors contributing one local Gaussian function, centred at that data point. By further thoughts it is possible to select the global basis function width, or angle (σ) using the Structural risk minimization (SRM principle) [33].

The development of the SVM approach was performed in MATLAB, which is a framework highly directed to mathematical calculations, such as required by this problem. The SVM approach for regression of the electricity market prices, takes as parameters:

- *Training Limit* - limit number of training days;
- *kernel* – kernel function that will be used in the regression process;
- *e_Val* - Value of ϵ -insensitive (sensitiveness to error);
- *C* – limit of kernel function;
- *p1* – angle of the kernel function (σ);
- *p2* – offset of the kernel function.

More details on the used SVM can be found at [30].

3.3.3. *Simple Average*

Finally, a very simple average forecast algorithm is also provided as option. This algorithm performs a forecast for a certain day considering the average offer values of the agents in question. This is a very simple, but also quick method to run, which is essential when the main requirement to be fulfilled is the execution time.

3.4. Clustering

Simulation using large quantities of data generates some drawbacks, mainly due to difficulties in analysing the results obtained by the simulation and the time necessary to achieve good results. However, there are numerous entities that participate every day in the EM. A promising solution for this issue is the data reduction. Regardless of the number of participating entities, many of them will exhibit similar behaviours in their offers. Thus, it is possible to extract a behavioural pattern and consequently categorize them as entities with similar characteristics and/or behaviour. This is hence the role of the clustering phase, which aims to simplify the simulation process and, in a subsequent step, the analysis of the results, when evaluating their performance.

The clustering process is divided into three parts: (i) selection of the size of the scenario, (ii) choice of the considered characteristics and respective weight assignment, and (iii) the clustering algorithm.

3.4.1. Scenario Dimension

The user can define the size of the scenario manually or dynamically. In the manual definition, the user indicates the exact number of agents that will be present in the simulation. This number can be exactly the same as the number of real players participating in the market, and, in this case, each software agent will represent a real player. If the chosen number of agents is smaller, the real players are grouped using the clustering process for producers and consumers. This clustering process will define the same number of groups (clusters) as the number of desired agents, so as to have one agent defined as result from each cluster.

The dynamic definition aims to facilitate the process for the user, and simply asking for a scenario dimension between 1% and 100%. The most appropriate number of agents, considering the preference dimension is achieved through fuzzy logic [28]. The smallest possible scenario - 1% - is characterized as having to group all players in one cluster (1 cluster to n players). The scenario with larger dimension - 100% - is where each cluster corresponds exactly to a real player (n clusters for n players). The fuzzy logic process allows the control of indeterminate states, transforming something undefined into something defined with an associated numeric value [34]. Thus, the method performs the intersection between the user-selected proportion with the actual amount of available data and provides as a result the number of clusters to be created by the clustering algorithm, i.e. the number of software agents to be considered in the simulation scenario.

3.4.2. Characteristics Choice

Each market player is defined through a list of characteristics. The identified characteristics are different for producers and consumers, as can be seen in Table 1.

List of considered characteristics for the clustering process

Table 1

The goal of the clustering is to divide the players into categories that define them. However, there may be characteristics that are not intended to be considered in the classification. The proposed model allows the user to define what characteristics to include in the process, for each of the different types of players (buyer and seller). The choice of the characteristics, in this step, has a significant impact on the clustering result, and in turn, in the result of the scenario, since choosing different characteristics originates very different scenarios.

In addition to the specification of characteristics used in clustering, it is also possible to assign a weight to each characteristic, i.e., define it to take a larger / smaller influence in the grouping process, when compared to the other characteristics used in this method.

3.4.3. Grouping

Two methods are considered to group the real players and create agents: K-Means clustering algorithm [35] and the Partitioning Around Medoids (PAM) clustering algorithm [36].

The K-Means clustering methodology considers a set of observations (x_1, x_2, \dots, x_n) , where each observation is a d -dimensional real vector, and n is the number of considered observations. The clustering process aims at partitioning the n observations into k ($\leq n$) clusters $C = \{C_1, C_2, \dots, C_k\}$ so that the Within-Cluster Sum of Squares (WCSS) is minimized (6).

$$\min \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (6)$$

where μ_i is the mean of points in C_i , i.e. the cluster *centroid*.

The dimension of the vector that characterizes each observation $x_p, p \in \{1, \dots, n\}$ is equal to the sum of the individual dimensions of n vectors, where each of these n vectors contains the data that is referent to a different data variable, e.g. electricity market prices, wind speed, solar intensity, etc.

With the objective of minimizing equation (6), the clustering process executes an iterative process between two steps: (i) the assignment step, where each observation x_p is assigned to the cluster $C^{(t)}$ whose mean value yields the minimum WCSS in iteration t , as presented in (7); and (ii) the update step, where the new means of each cluster are calculated, considering the newly assigned observations, determining the new *centroid* μ_i of each cluster, as in (8).

$$C_i^{(t)} = \{x_p : \|x_p - \mu_i^{(t)}\|^2 \leq \|x_p - \mu_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\} \quad (7)$$

$$\mu_i^{(t+1)} = \frac{1}{|C_i^{(t)}|} \sum_{x_j \in C_i^{(t)}} x_j \quad (8)$$

The execution of the algorithm stops when the convergence process is completed, *i.e.* when the assignments of observations to different clusters no longer change. By minimizing the WCSS objective, in equation (6), the K-Means clustering methodology assigns observations to the nearest cluster by distance. This means that each observation will be grouped in the same cluster as the ones that are more similar.

K-Means is, however, not prepared to deal with categorical data, so the adopted solution was to convert data to numeric values in order to be able to include this data. However, after exhaustive analyses it has been concluded that the conversion of categorical data prevented the total separation of clusters. A small example: by converting wind and nuclear technology types for numbers 1 and 2, respectively, the K-Means algorithm goes on to conclude that nuclear technology is the successor of wind technology and wind technology is half the nuclear. Such findings prevent separation of categorical characteristics since their values will always be related to each other. In order to get around this situation, the PAM clustering algorithm has also been implemented, thereby allowing the inclusion of categorical data [36].

The PAM algorithm computes k representative objects, called *medoids*. A *medoid* can be defined as an object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal. Such representative objects are usually called *centrotypes*. After finding the set of *medoids*, each object of the data set is assigned to the nearest *medoid*, *i.e.* object i is put into cluster v_i when medoid mv_i is nearer than any other medoid m_w , as in (9).

$$d(i, mv_i) \leq d(i, m_w), \forall w = 1, \dots, k \quad (9)$$

The k representative objects should minimize the objective function (O), which is the sum of the dissimilarities of all objects to their nearest *medoid*, as formalized in (10).

$$O = \sum d(i, mv_i) \quad (10)$$

The algorithm proceeds in two steps:

- *Build-step*: This step sequentially selects k "centrally located" objects, to be used as initial *medoids*;
- *Swap-step*: If the value of O can be reduced by interchanging (swapping) a selected object with an unselected object, then the swap is carried out. This is continued till the objective function can no longer be decreased.

The PAM algorithm allows to perform clustering with numerical and categorical data but, for this purpose, it needs to calculate the distances between observations of the input data. In order enable including categorical data, the Gower metric is used [37]. The Gower metric uses a distance measure between 0 and 1 for each variable $d_{ij}^{(f)}$. The distances are then aggregated as in (11).

$$d(i, j) = \frac{1}{k} \sum_{i=1}^k d_{ij}^{(f)} \quad (11)$$

The distances are calculated as in (12), where x_{if} is the value for object i in variable f , and R_f is the range of variable f for all objects.

$$d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{R_f} \quad (12)$$

Another important feature of the clustering process is the ability to set different weights for each characteristic in the clustering process, thus defining the impact that each one has on the clustering process.

3.5. Virtual Power Players Creation

Because of the multiplicity of Producer and Consumer players participating in the market, the variety of offers for the price and amount of power is also abundant. However, not all players are able to present the same influence on the market, and for smaller players, they may not even be allowed to participate in wholesale market negotiations, due to dimension restrictions. In this context aggregators play an important role as to enable smaller sized players to join together into coalitions in order to be able to compete in equal footing with larger players. With this arises the concept of Virtual Power Player (VPP), which consists in forming a group of agents operating in the market, as seen from the market perspective as a single Producer or Consumer entity [38]. In order to potentiate the study of these aggregator entities, and the impact they can have on the market,

RealScen also allows the creation of VPPs through the aggregation of several players, and their inclusion in simulated EM scenarios.

For the VPP formation, the user must indicate the agents to include in the group and, according to the chosen players, the VPP can be seen from the market perspective as a Producer, Consumer or Mixed. In the particular case of the Mixed type, it means that the VPP can take the behaviour of Producer or Consumer depending on its current needs in each negotiation period.

In order to achieve the aggregation of virtual agents a new application of clustering is required, however, this differs slightly from the previous one, since only one group of agents is required for each VPP. VPP acquires, for each characteristic, the one that is more common between the selected agents, with the exception of the quantity and price of energy, which is the result of the sum of the selected agents' amounts.

4. Experimental Findings

This section presents a case study that aims to demonstrate the scenarios generation process, and evaluate the quality of the generated scenarios. With this objective, 3 scenarios are generated using the proposed methodology: (i) a small scenario, considering a very small number of software agents to represent all real players that participate in the EM in a given day; (ii) a medium scenario, considering an intermediate number of agents; and (iii) a big simulation scenario, which includes a large number of software agents. The small scenario results from the grouping of real players taking into account only one of the characteristics that characterize them, the medium scenario performs the clustering process using two characteristics, and the big scenario uses four characteristics to group real players according to their similarity, in order to generate the software agents that represent the real players in the simulated scenario.

The chosen dimension of the scenarios is based on the desired amount of characteristics to be used to differentiate the players (e.g. dimension of small scenario that will be based on the amount of different values that a specific characteristic can have). However, this is just one of the experiments that can be performed in order to obtain the most suitable scenario. When choosing the dimension of a scenario, many criteria can be taken into account, being performance the main one. In this case, the performance to take into account is the MASCEM's performance, as it will be responsible for the simulation of the created scenario. Table 2 presents the average execution times of MASCEM with different numbers of players, as presented in [23].

The above table shows that the MASCEM's average execution time doesn't have significant difference between the scenarios with 48 or less players (about 1 second) and presents a very fast execution time (4-5 seconds). On the other hand, when simulating all the real players (1446), the execution time is much higher, about 1 minute, to simulate a single day (a month would take about 30 minutes). When defining the dimension of the scenario to create, the most suitable number of players would be a number that is capable of representing accurately all the real players while guaranteeing a fast execution time. In this case, a good test number of players would be somewhere between 12 and 48 players.

Another important criteria is the interpretability of the simulation results. For instance, in an academic level, when the aim of the study is the evolution of the market prices, a small scenario would be enough. On the other hand, if we want to analyse the influence of the different types of player (e.g. type of generation), a big scenario would be required. Sometimes a medium scenario is the most appropriate as it would guarantee a better execution time than a big scenario and better results than a small scenario. The medium scenario would be the most appropriate in the most cases unless if the desired execution time is the fastest as possible, without considering the quality of results.

This are just a few examples of criteria to define the number of players to consider. It is important to keep in mind that a good scenario is a scenario that suits each particular case study. In some case a number or percentage of players can give excellent results and in other case, the same value can fail to meet the study purpose. This is why RealScen is very important as it allows the creation of scenarios for different purposes. It can suit the needs of training, research, study from market operators and regulators or even study from participating entities in the market such as emerging entities or aggregators. Everyone gains from using this tool in a different and personalized way.

The evaluation of the scenarios quality is performed under two dimensions: (i) the level of representativeness of the scenario relatively to the reality, i.e. how well the reduced number of agents in the simulation scenario is able to represent the characteristics of all the real players; and (ii) the comparison between the real EM results and the results from simulations using the generated scenarios.

The data used to create the scenarios is real data, referring to the year 2014, of the Portuguese and Spanish reality, including a total of 1428 real players. With this information (available in [39]), it is possible to represent all players in the Iberian EM - MIBEL.

The generation of the three scenarios takes into account the following characteristics of the real EM players:

- Region – Region to which the player belongs (North / Central / South).
- Country – Player’s country (Portugal / Spain).
- Dimension – Player dimension based on their installed power. The scale is divided into three levels: small, medium and large.
- Installed power – Player’s installed power.
- Technology Type – Player’s type of technology. A total of ten different types of generation technology are considered, as shown in Table 3:

Technology types of the considered players

Table 3

4.1. Small Scenario

Given that the number of considered technology types is 10, the small scenario is defined setting as objective the creation of 10 seller agents, with the objective of having one agent representing all real players of a given technology type.

By grouping the real players into 10 clusters, according only to their type of technology, the players have been distributed as shown in Table 4.

Number of players per type of technology

Table 4

Table 4 shows that the process of clustering allowed the complete separation of players’ technology types, where each cluster contains only players of different types of technology. When comparing the clustering results with the real players’ data, it is possible to see that every player is allocated to a cluster with the same technology type as its own. The full separation of the players has been possible due to the use of the PAM clustering algorithm, which allows the clustering of categorical data, as presented in section 3.4.3.

Analysing the results it can be seen that the vast majority of producer players in the Iberian Peninsula are based on wind energy (78.6%). The second most represented generation type is small hydroelectric (9.7%), followed by large hydroelectric (3.2%) and solar photovoltaic parks (2.8%) as the most popular Iberian Peninsula, while nuclear (0.6%), cogeneration (0.3%) and fuel gas (0.1%) are less common. Note that this analysis refers to the number of players of each type of technology, and not to their installed power. The percentage of installed power per cluster generated by RealScen (i.e. per type of technology) and its comparison to the real data, as provided in [40] is shown by Figure 3.

Percentage of installed power per technology

Fig. 3

By analysing these results, it is possible to see that the proposed methodology is able to achieve exactly the same results when representing players per technology type, as occurs in reality. Regarding the obtained values, wind energy is the type of technology with more installed power in the Iberian Peninsula (32.2%), closely followed by CCGT (29.6%). The next technology type with most installed power is Coal with about half of the value of the previous two (16.3%). The Coal technology type is followed by two types with a number of installed power close to each other (Hydroelectric Power with 10.4% and Nuclear with 8.9%). At last there is Fuel Gas (0.8%), Photovoltaic Solar Park (0.8%), Small Hydroelectric Power (0.5%), Biomass (0.4%) and Cogeneration (0.2%) with a very low amount of installed power. Thus, it is possible to observe that the number of agents with a particular type of technology does not guarantee a greater or lesser amount of installed power. Figure 4 shows the percentage of installed power by agent of each technology type, i.e. what is the percentage of installed power represented by each player of each technology type.

Percentage of installed power per agent

Fig. 4

From Figure 4 it is easy to see that the technology types that have more installed power by agent are nuclear, CCGT and Coal (29.8%, 26.5% and 23.1%, respectively). In this scenario, wind is the major technology that, as can be seen, requires many more producer agents to generate the same amount of energy as other types of technology, such as nuclear (with a percentage of installed power per agent about 37 times higher than wind).

4.2. Medium Scenario

Departing from the previous scenario, the objective is to create a scenario with a higher number of agents in order to increase real players' separation in order to achieve a more comprehensive scenario. For this analysis, in addition to the technology type characteristic, the dimension characteristic is also considered. The dimension characteristic allows the separation of agents as Small, Medium and Big, given their amount of installed power. Generators with less than 50 MW of installed power are characterized as Small, the Medium plants have an installed power equal or greater than 50 MW and less than 100 MW, and the Big plants have 100 MW or more installed power.

Contrarily to the small scenario, it is not known beforehand how many clusters (software agents) should result for the desired separation. Theoretically, it is possible to know the number of clusters to use to obtain a total separation by multiplying the number of technologies by the number of dimensions that each can have. However, not all technologies have all dimensional levels, e.g. nuclear agents given that they all have installed power above 100 MW. It is, however, known that the suitable number of clusters is between 10 (in case there is only one dimension for each technology) and 30 (if all ten technologies have the three dimensions). The resulting number of clusters has been 17. Figure 5 shows the percentage of installed power by technology dimension.

Percentage of installed power per dimension of player

Fig. 5

From Figure 5 it is visible that, naturally, plants with greater installed power are the big ones except for the wind's small and medium centrals, given the large number of agents with this technology type. As in the previous scenario, the results obtained are similar to those verified in reality. Analysing in detail the case of big wind farms, there are 14 in Portugal and 2 in Spain, thus a total of 16 in the Iberian Peninsula, as can be seen in the data used [39]. The sum of the installed power results in 2331.7 MW which represents approximately 2.67% of the total power installed by all technologies (87138.19 MW).

4.3. Big Scenario

By studying the medium scenario it has been possible to include more types of agents; however, since not all have various dimensions, the size of the scenario did not increase significantly. With the generation of the big scenario the objective is to achieve a significant increase of clusters and attain better visualization of the influence of other players' characteristics in the clustering process. The previous scenario is increased with the addition of characteristics concerning the location of players, namely region and country to which they belong. The minimum number of possible clusters (software agents) is 10, while the maximum is 180 (Resulting from the multiplication of 10 technologies by 3 dimensions, 2 countries and 3 regions).

With the addition of two characteristics, regarding the localization of agents, in the clustering process, the ideal number of achieved clusters is 55. This value is well below the maximum of 180, since not all technologies are present in all regions of both countries; in addition to the conclusion taken from the medium scenario, that each technology type rarely has the three dimensions associated to it. Figure 6 shows the 19 clusters with the highest percentage of installed power.

Percentage of installed power per location and dimension of players of each technology type

Fig. 6

By analysing Figure 6 it can be easily seen that, in general, Spanish regions exceed the Portuguese in amount of installed power given the different geographic area of the concerned regions. In particular, the northern region of Spain that has the clusters with higher installed power (small dimension of wind and the big dimension of the coal and CCGT technologies).

Again, by verifying the real data [39], it is clear that nuclear power plants exist only in the north and centre of Spain, both areas with four plants and all of big dimension. From Figure 6 it can also be seen that the generation of the scenario by grouping the players by technology type, dimension, country and region, has originated the expected result as, considering the previous observation, there are two clusters (software agents) related to nuclear technology where one belongs to the north of Spain and another in the centre. The total percentage of installed power by these two clusters is equal to the value obtained in the small scenario by the nuclear energy cluster and also on the medium scenario for the cluster of nuclear energy of big dimension.

4.4. Simulation Results

The definition of the three scenarios has demonstrated the capability of the proposed methodology in generating scenarios that are fully representative of the real EM players. This section compares if EM simulations using these scenarios, using the software agents generated from each cluster, and complemented by the price and power forecasts proposed in section 3.3., are able to generate similar market results to those achieved in reality. For this, the MASCEM simulator (presented in section 2.1) is used to simulate the EM, using MIBEL day-ahead spot market mechanism, using each of the generated scenarios. From these simulations it is possible to compare the real market price that has been achieved, with the market price achieved in each hourly negotiation period, for each scenario. The considered simulation day is June, 1st 2014. Figure 7 shows the results and compares them with the real market price of each period.

Comparison between the real market price and the market price achieved using each scenario

Fig. 7

From Figure 7 it is visible that the market prices that resulted from the EM simulation using the Big scenario are exactly equal, in all 24 negotiating periods of the considered day, to those that resulted from the MIBEL day-ahead spot market. These results show that, by using the proposed scenarios generation methodology, MASCEM is able to represent the real EM players, with a much inferior number of software agents, flawlessly, given the accuracy of the simulated market prices. When reducing the number of represented players, the similarity of the results also decreases when compared to the real MIBEL outputs. When considering the simulation results using the Medium scenario, it is visible that the EM price curve follows the tendency of the real market price, accompanying the decrease in price in the first periods of the day, and the small peak in period 23. Reducing the number of players to 17 (Medium scenario) reflects into an increase of the price variation, distancing the resulting market price from the real MIBEL market price, although still following the real price tendency. Finally, when reducing the number of players to 10 (Small scenario), it becomes obvious that the simulation is no longer capable of fully reflecting the reality. An enormous reduction of the number of players (10, when the number of players participating in the MIBEL spot market is nearly 1500) leads to an incoherent representation of the market environment. Using a small number of agents and given that each has a different type of technology, it is verified that certain types of technology have more weight than the others and direct the results of the simulation. The influence of wind power, with a big generation power superiority and very

low selling prices, is visible in periods 10 to 15, where the achieved market price is 0 €/MWh. On the other hand, the Medium scenario presents results very close to reality, with only small errors on periods between 3 and 6 and on period 21 and 23. The best results were obtained by the Big scenario, which has resulted in simulation results equal to the reality.

Considering the comparison of the obtained results by the created scenarios, it can be seen that the higher the scenario is, the better the representation of reality. However, from a certain point on, it does not pay off to add more agents to the scenarios because the gain is not significant. For example, the Big scenario that was generated could fully represent the reality with only 55 agents (approximately 4% of the number of real players). By representing reality with so few agents, a lot of resources and time in the simulation process can be saved, which allows obtaining results more quickly, and facilitating their interpretation.

Taking into account the performance of the scenarios generation with the proposed methodology, it is also important to compare these results with those achieved without this method. As mentioned before, before the adoption of this methodology, the only way to obtain a realistic scenario was to create it manually, since no other scenarios generation methodology can be found in the literature. Figure 8 show the comparison between the market price and the price achieved by three scenarios (bid, medium and small, with the same number of agents as the previously presented scenarios), created manually. These scenarios are created by considering the larger players (larger amount of traded power), that have participated in the market in the considered simulation day.

Comparison of the real market price and the price achieved by three simulations using scenarios created manually Fig. 8

From Figure 8 it is visible that the market price values achieved by the simulations using the manually created scenarios are clearly less reliable than those verified using the scenarios created with RealScen (Fig. 7). The big and medium scenarios are able to follow the general market price tendency throughout most periods of the considered day. However, the quality of the scenarios is not enough to visibly accompany the market price fluctuations from one period to the following. Regarding the small scenario, it is clear that the null market price throughout the day is completely unsuitable to represent the electricity markets reality in the considered simulation day. Table 5 presents the comparison of the error, using the Mean Absolute Percentage Error

(MAPE), regarding the approximation of the market prices achieved by the different scenarios (created by RealScen and created manually) when compared to the real market price.

Comparison of simulated market price MAPE values between manual and RealScen's scenarios (%) Table 5

As it can be observed in Table 5, the scenarios generated by RealScen presents a much inferior error than the corresponding manually created scenario. The highest difference occurs in the small scenarios where the manually created one has an error of almost 100% while RealScen's scenario is about 37%. Although the difference of MAPE value between both methods is already a good indicator in favour of the proposed methodology, their most important difference is related to their time consumption and failsafe. The manual method consumes a much greater amount of time and is highly subject to human mistakes, which can compromise the results. On the other hand, RealScen is able to easily generate scenarios without taking too much time and, as it is automatic, it is safe from human errors.

With this analysis it's possible to conclude that both methods can deliver great results but certainly not with the same response time and the same reliability, which are two key factors. Taking this into account and the fact that there is no other known tool capable of automatic and quality generation of realistic scenarios, it's possible to recognize the importance that RealScen has in this area of study and its great contribution.

5. Conclusions

The electricity markets area is in constant evolution and has undergone major changes in its operation during the last few years. This evolution has dramatically increased its complexity and, consequently, has increased the difficulty in analysing its operation. In order to combat this trend, the electricity market simulators need to work with real data, to be able to perform simulations that reflect what occurs in real electricity markets. This paper presents a methodology that enables the generation of realistic simulation scenarios, using real electricity market data. Scenarios can fully represent the reality, or approximate it through a reduced number of representative software agents. The proposed methodology is integrated in RealScen, a simulation scenarios generation tool that, through the extraction and processing of real information, can provide quality information to simulators, to support their simulation success.

The presented study case proves the ability of the proposed methodology in generating realistic simulation scenarios, which are able to reflect the characteristics of real players. Results also show that simulations using the generated scenarios are able to achieve similar market results to those achieved in reality, thus demonstrating that reality can be modelled even through a reduced number of agents.

Acknowledgements

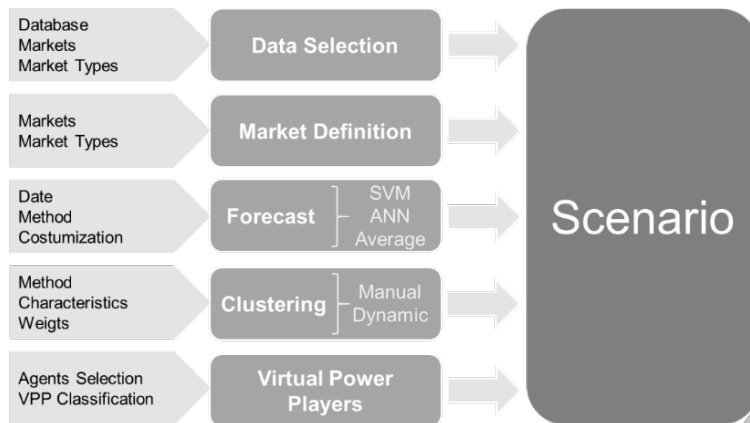
This work is supported by FEDER Funds through COMPETE program and by National Funds through FCT under the projects FCOMP-01-0124-FEDER: PEst-OE/EEI/UI0760/2015; PTDC/SEN-ENR/122174/2010; and PTDC/EEA-EEL/122988/2010.

References

- [1] M. Shahidepour, et al., "Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management", Wiley-IEEE Press, pp. 233-274, 2002.
- [2] F. P. Sioshansi, Evolution of Global Electricity Markets: New paradigms, new challenges, new approaches. 2013
- [3] H. Li, and L. Tesfatsion, "Development of Open Source Software for Power Market Research: The AMES Test Bed", Journal of Energy Markets, vol. 2, no. 2, pp. 111-128, 2009.
- [4] T. Pinto, et al., "Negotiation Context Analysis in Electricity Markets", Energy, Elsevier, vol. 85, pp. 78-93, June 2015
- [5] B. Teixeira, F. Silva, T. Pinto, I. Praça, G. Santos, and Z. Vale, "Data mining approach to support the generation of Realistic Scenarios for multi-agent simulation of electricity markets," in Intelligent Agents (IA), 2014 IEEE Symposium on, 2014, pp. 8–15.
- [6] MIBEL – homepage, Available: <http://www.mibel.com>, accessed on March 2016
- [7] EPEX – homepage, Available: <https://www.epexspot.com>, accessed on March 2016
- [8] Nord Pool – homepage, Available: <http://www.nordpoolspot.com>, accessed on March 2016
- [9] I. Praça, T. M. Sousa, A. Freitas, T. Pinto, Z. Vale and M. Silva, "Adaptive Tool for Automatic Data Collection of Real Electricity Markets," 2012 23rd International Workshop on Database and Expert Systems Applications, Vienna, 2012, pp. 345-349.
- [10] F. Silva, B. Teixeira, T. Pinto, G. Santos, I. Praça, and Z. Vale, "Demonstration of Realistic Multi-agent Scenario Generator for Electricity Markets Simulation," in Advances in Practical Applications of Agents, Multi-Agent Systems, and Sustainability: The PAAMS Collection: 13th International Conference, PAAMS 2015, Salamanca, Spain, June 3-4, 2015, Proceedings, Y. Demazeau, S. K. Decker, J. Bajo Pérez, and F. de la Prieta, Eds. Cham: Springer International Publishing, 2015, pp. 316–319.
- [11] D.R. Biggar, and M.R. Hesamzadeh (Eds.), "The Economics of Electricity Markets", Wiley, 1st edition, September 22, 2014
- [12] T. Pinto, et al., "Adaptive Learning in Agents Behaviour: A Framework for Electricity Markets Simulation", Integrated Computer-Aided Engineering, IOS Press, 21, 4, 399-415, September 2014
- [13] V. Koritarov, "Real-World Market Representation with Agents: Modeling the Electricity Market as a Complex Adaptive System with an Agent-Based Approach", IEEE Power&Energy, pp.39-46, 2004.
- [14] H. Li, and L. Tesfatsion, "Development of Open Source Software for Power Market Research: The AMES Test Bed", Journal of Energy Markets, vol. 2, no. 2, pp. 111-128, 2009.
- [15] I. Praça, C. Ramos, Z. Vale, and M. Cordeiro, "MASCEM: A Multi-Agent System that Simulates Competitive Electricity Markets", IEEE Intelligent Systems, vol. 18, No. 6, pp. 54-60, Special Issue on Agents and Markets, 2003.

- [16] G. Santos, et al., "Multi-Agent Simulation of Competitive Electricity Markets: Autonomous systems cooperation for European Market modelling", *Energy Conversion and Management*, 99, 387-399, July 2015
- [17] F. Lopes, H. Algarvio, and H. Coelho, "Bilateral contracting in multi-agent electricity markets: Negotiation strategies and a case study," *European Energy Market (EEM)*, 2013 10th International Conference on the , vol., no., pp.1,8, 27-31 May 2013
- [18] A. K. Jain, M. N. Murty and P. J. Flynn, "Data Clustering: A Review", *ACM Computing Surveys*, 31 (3). pp. 264-323, 1999
- [19] U. Fayyad, et.al., "From Data Mining to Knowledge Discovery: An Overview", in *Advances in Knowledge Discovery and Data Mining*, pages 1-34, AAAI/MIT Press, 1996
- [20] T. Pinto et al., "Adaptive Portfolio Optimization for Multiple Electricity Markets Participation," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1720-1733, Aug. 2016.
- [21] G. Santos, T. Pinto, I. Praça, and Z. Vale, "MASCEM: Optimizing the performance of a multi-agent system," *Energy*, vol. 111, pp. 513-524, 2016.
- [22] B. Wilamowski, and H. Yu, "Neural Network Learning Without Backpropagation", *IEEE Transactions on Neural Networks*, vol. 21, no. 11, pp. 1793-1803, 2010
- [23] Gomide, F and Pedrycz, W., "Notions and Concepts of Fuzzy Sets", *Fuzzy Systems Engineering: Toward Human-Centric Computing*, 2007
- [24] I. F. Pereira, et.al., "Data Extraction Tool to Analyse, Transform and Store Real Data from Electricity Markets" in *Advances in Intelligent Systems and Computing*, S. Omatu, et al. (Eds), pp. 387-395, Springer International Publishing, 2014
- [25] T. Pinto, T.M. Sousa and Z. Vale, "Dynamic artificial neural network for electricity market prices forecast," in *Proceeding of the 16th IEEE International Conference on Intelligent Engineering Systems*, Lisbon, Portugal, Jun. 2012, pp. 311-316
- [26] K. Rudd, G. Di Muro and S. Ferrari, "A Constrained Backpropagation Approach for the Adaptive Solution of Partial Differential Equations," *IEEE Trans. on Neural Networks and Learning Systems*, vol. 25, no. 3, pp. 571,584, March 2014
- [27] T. Pinto, T.M. Sousa, I. Praça, Z. Vale and H. Morais, "Support Vector Machines for Decision Support in Electricity Markets' Strategic Bidding", *Neurocomputing*, Elsevier, in press, August, 2015
- [28] H.M. Le, H.A.L.Thi, M.C. Nguyen, "Sparse semi-supervised support vector machines by DC programming and DCA", *Neurocomputing*, vol. 153, p.p. 62-76, 4 April 2015
- [29] Luo, M. and Zhou, X., "Robustness of reverse triple I algorithms based on interval-valued fuzzy inference", *International Journal of Approximate Reasoning*, 66, 16-26, November 2015
- [30] Anil K. Jain. "Data Clustering: 50 years beyond K-Means". *Pattern Recognition Letters*, Elsevier, Vol. 31, Issue 8, pp.651-666, June 2010
- [31] H.S. Park , C.H. Jun, A simple and fast algorithm for K-medoids clustering, *Expert Systems with Applications*, vol. 36, no.2, pp. 3336-3341, 2009
- [32] J. C. Gower, "A general coefficient of similarity and some of its properties", *Biometrics*, vol. 27 no. 4, pp. 857-871, 1971
- [33] H. Morais, et. al. "Multilevel Negotiation in Smart Grids for VPP Management of Distributed Resources", *IEEE Intelligent Systems magazine*, Special Issue "Sustainable Energy and Distributed AI", vol. 27, no. 6, pág. 8-16, November - December, 2012
- [34] MIBEL - Iberian Electricity Market Operator, homepage, <http://www.omie.es>, accessed on March 2016

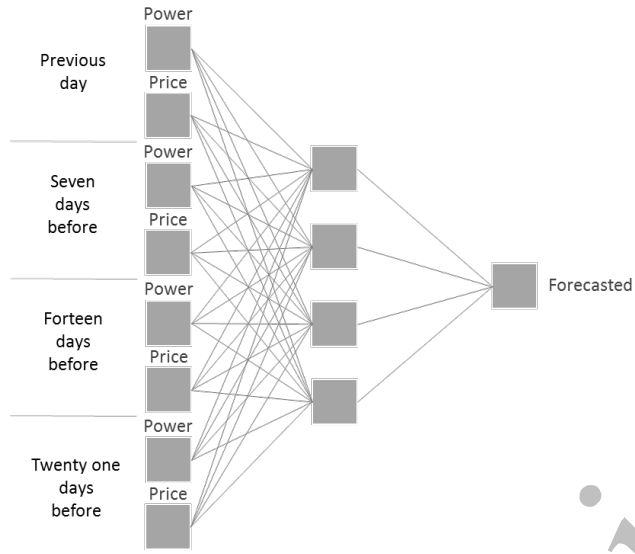
List of Figures



RealScen architecture

Fig. 1

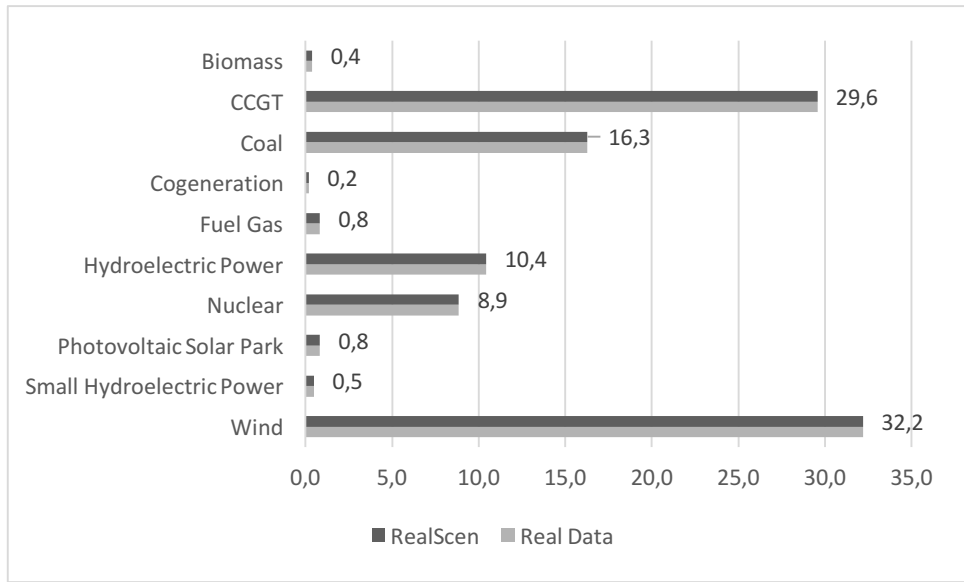
author's version



ANN topology

Fig. 2

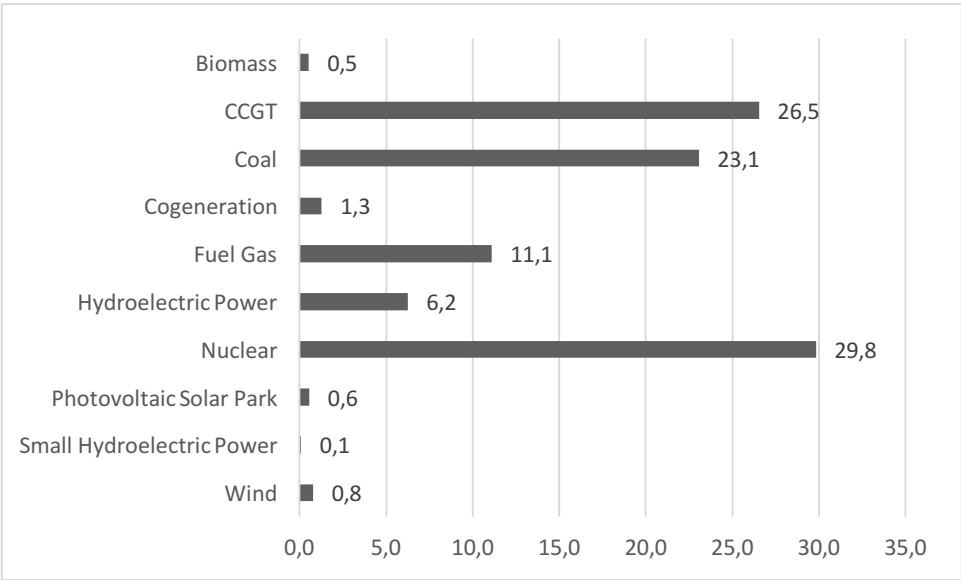
author's version



Percentage of installed power per technology

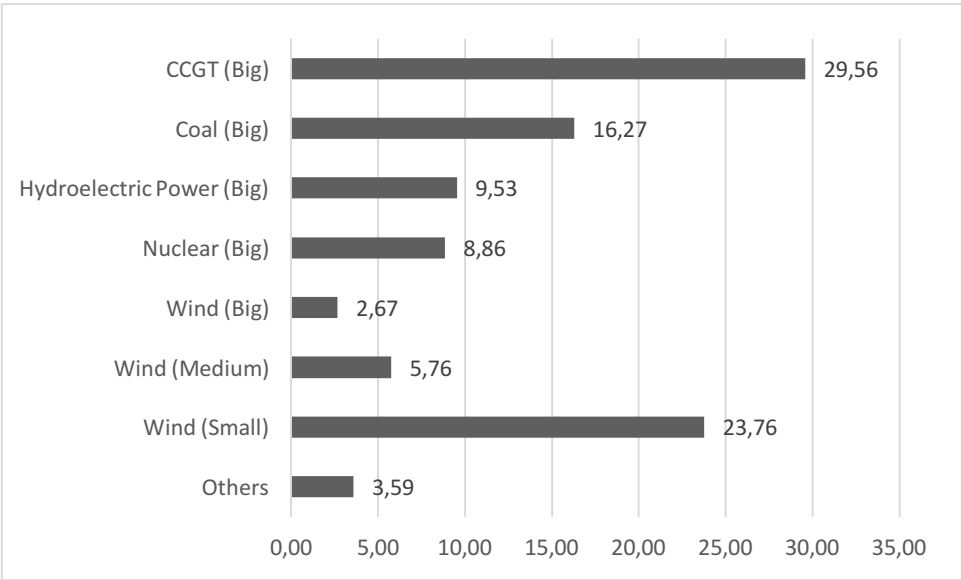
Fig. 3

author's version



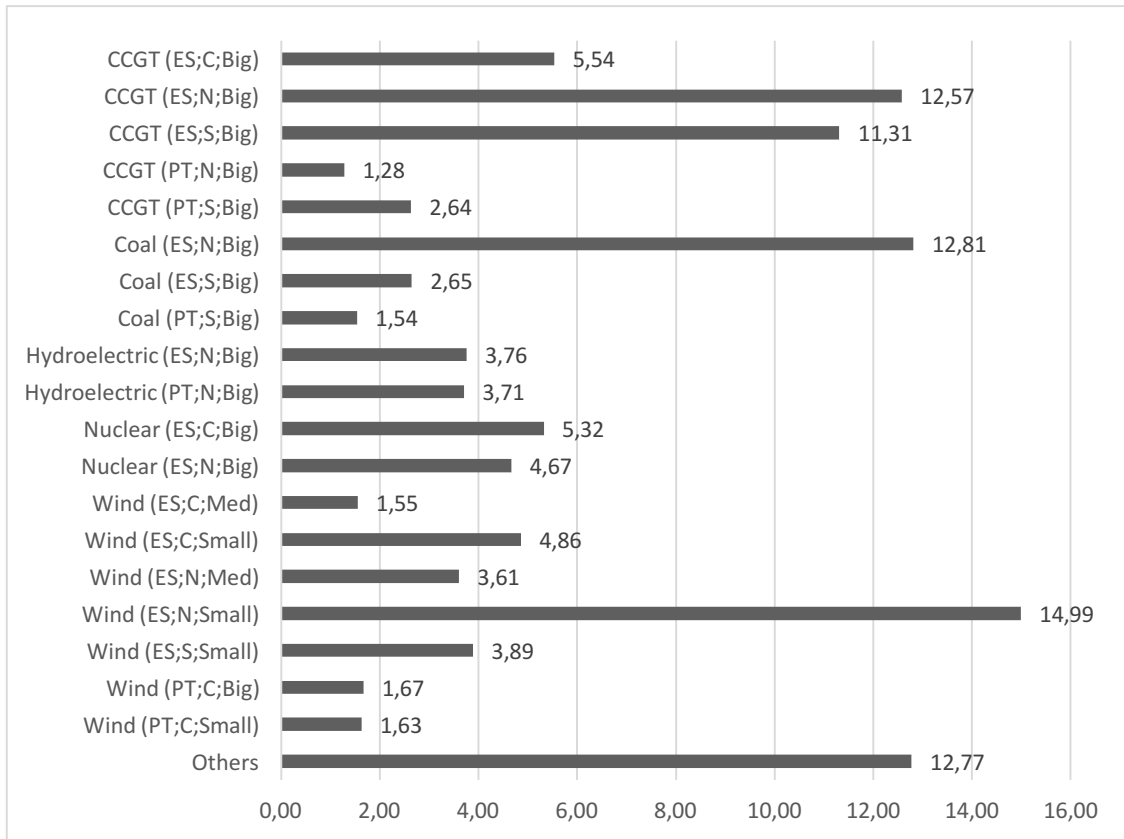
Percentage of installed power per agent

Fig. 4



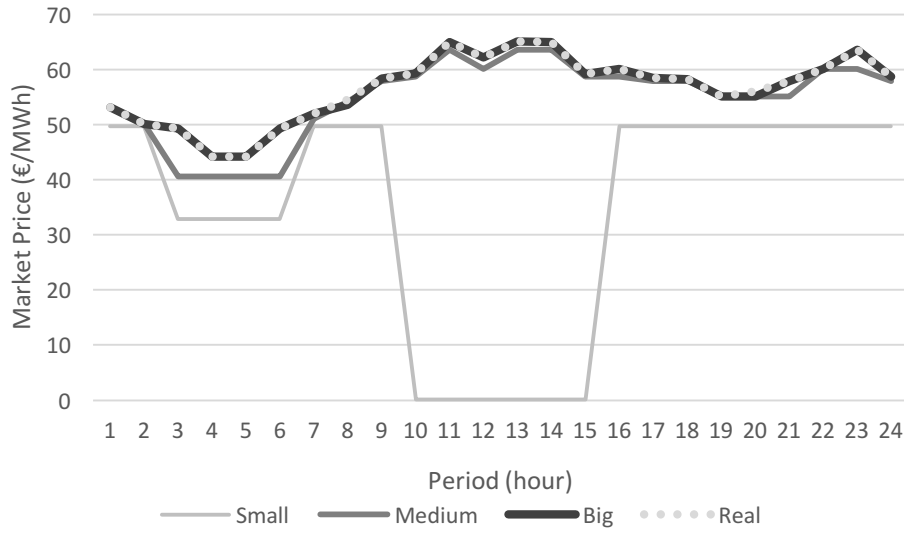
Percentage of installed power per dimension of player

Fig. 5



Percentage of installed power per location and dimension of players of each technology type

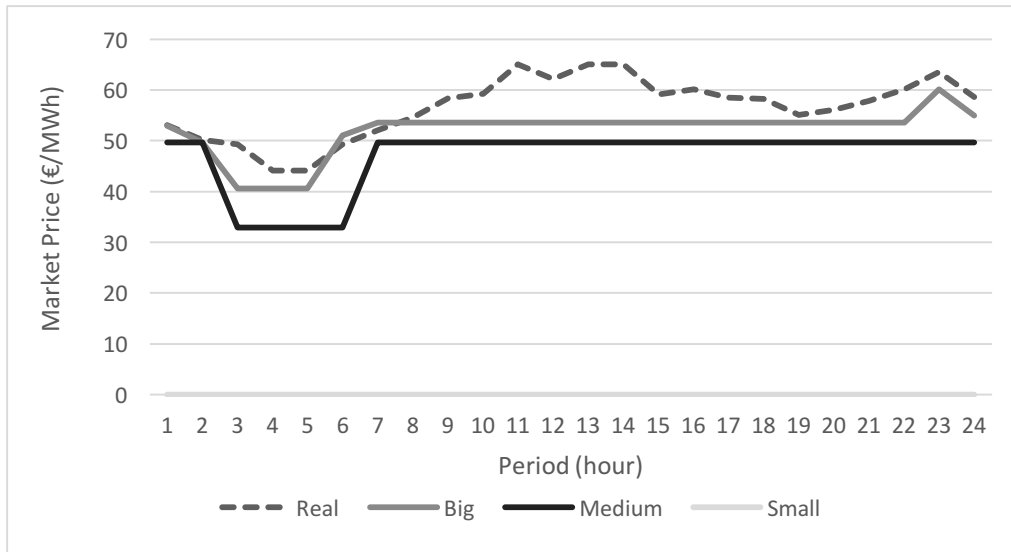
Fig. 6



Comparison between the real market price and the market price achieved using each scenario

Fig. 7

author's Vers



Comparison of the real market price and the price achieved by three simulations using scenarios created manually

Fig. 8

author's version

List of Tables

<i>Producers</i>	<i>Consumers</i>
Availability	Average power
Average power	Average price
Average price	Consumer type
Connection point	Enterprise
Emissions	Latitude
Enterprise	Longitude
Expectable cost	Maximum power
Latitude	
Lifetime	
Longitude	
Maximum power	
Minimum power	
Overload power	
Reliability	
Store capacity	
Technology type	
Voltage level	
Thermal capacity	
Volatility	

List of characteristics available for the clustering process

Table 1

<i>1446 Players</i>	<i>48 Players</i>	<i>24 Players</i>	<i>12 Players</i>
65327	5628	4841	4426

Average execution time of MASCEM (in milliseconds)

Table 2

author's version

<i>Technology Type</i>
Biomass
CCGT
Coal
Cogeneration
Fuel Gas
Hydroelectric Power
Nuclear
Photovoltaic Solar Park
Small Hydroelectric Power
Wind

Technology types of the analysed agents

Table 3

author's version

<i>Technology Type</i>	<i>No. Players</i>	<i>%</i>
Biomass	19	1,3
CCGT	30	2,1
Coal	19	1,3
Cogeneration	4	0,3
Fuel Gas	2	0,1
Hydroelectric Power	45	3,2
Nuclear	8	0,6
Photovoltaic Solar Park	40	2,8
Small Hydroelectric Power	139	9,7
Wind	1122	78,6
Total	1428	100

Number of agents per type of technology

Table 4

author's version

<i>Method</i>	<i>Small</i>	<i>Medium</i>	<i>Big</i>
Manual	0.9998	0.1714	0.0843
RealScen	0.3699	0.0356	0.0015
Difference	0.6299	0.1358	0.0828

Comparison of simulated market price MAPE values between manual and RealScen's scenarios (%)

Table 5

author's version