

# A Virtual Reservoir Electricity Market Design Applied to the Brazilian System Using an Agent Based Model

Felipe Alves Calabria

CNPq Brazil and FEUP

Fac. de Engenharia da Univ. do Porto

Porto, Portugal

felipecalabria@gmail.com

João Tomé Saraiva

INESC TEC and FEUP/DEEC

Fac. de Engenharia da Univ. do Porto

Porto, Portugal

jsaraiva@fe.up.pt

Ana Paula Rocha

FEUP/DEI/LIACC

Fac. de Engenharia da Univ. do Porto

Porto, Portugal

arocha@fe.up.pt

**Abstract**— This paper discusses the current Brazilian electricity market, brings out some dilemmas that should be examined in order to implement a more market-oriented approach, and describes a new market design to overcome these issues. The proposed market design is based on virtual reservoirs and aims at enhancing the flexibility to enable market participants to comply with their contracts, while still ensuring the efficient use of the water and maintaining the current level of the security of supply. In addition, to simulate the behavior of the market participants in this new framework, an Agent-Based Model - ABM where agents use reinforcement Q-Learning - is developed and applied to a case study that includes a centralized dispatch as currently exists in Brazil. The results obtained so far show that this new design is suitable to allow hydros managing their commercial contract commitments with extra flexibility.

**Index Terms** - Electricity market design, hydro power generation, Brazilian electricity market, virtual reservoir model, agent-based simulation.

## I. INTRODUCTION

The electricity sector in several countries has been changing in order to ensure fair competition, higher efficiency, declining prices and reliability of operation. The Brazilian electricity market is characterized by having a continental interconnected transmission system, a large and growing demand, a total installed generation capacity around 137 GW [1], from which around 70% comes from hydropower plants with multiple owners coexisting in hydro cascades. This electricity market has gone through two large institutional and regulatory reforms in the last twenty years, and nowadays it contains some aspects that distinguish it from other markets.

In this market two contracting environments coexist: the *ACR* - Regulated Contracting Environment (based on a single buyer model), encompassing around 70% of the market, and the *ACL* - Free Contracting Environment (competition in the retail market), responsible for 30% of the demand. In the *ACR*, prices are determined by national public auctions conducted by the government while the *ACL* prices are freely negotiated.

The total amount of energy that can be traded in the market by each power plant is called “physical guarantee”. This amount can be interpreted as a certificate assigned by the Ministry of Mines and Energy and it represents the maximum energy generation that can be provided almost continuously over the years. Through a process called “seasonalization”, once a year market participants are allowed to monthly distribute the annual amount of energy coming from their physical guarantee and bilateral contracts.

Power plants are dispatched by the system operator (ONS) in a centralized way to optimize the hydrothermal system and to ensure the efficient operation of the hydro cascades. Since generators are not allowed to decide their own generation in order to comply with their contracts (there is not a bid based market), and the ISO decides their outputs without considering their commercial commitments, there is a mechanism known as *MRE* – Mechanism for Energy Reallocation. Shortly, aiming to offset this centralized dispatch, the *MRE* reallocates energy, transferring the surplus generated from those that produce beyond their monthly allocated physical guarantee to those that produce below. Finally, all contracts must be registered at the market operator (CCEE), who also measures the energy produced/consumed by each market participant and automatically uses the *MRE* rules in the settlement process of the Brazilian short-term market (*MCP*).

The above market design poses several problems to market agents as detailed in the next sections. The recognition of these problems is the main focus of the research reported in this work in which we propose a new market design based on virtual reservoirs. This design aims at enhancing the flexibility of market agents to comply with their commercial commitments while not destroying the optimized operation of the entire system. According with these ideas, this paper is organized as follows. After this Introduction, Section II details the dilemmas and problems associated with the current market structure in Brazil, Section III details the proposed virtual reservoir based design, Section IV describes the implemented ABM simulation approach, Section V presents the results obtained so far and Section VI outlines the main conclusions.

## II. PROBLEMS AND DILEMMAS

In the current Brazilian market design, the conciliation between commercial commitments and the physical dispatch is not smooth: there is a lack of “trading opportunities” to encourage hydro companies to comply with their contracts (namely related with the unique annual “window” of the seasonalization process) and there is no flexibility for hydros to better address their exposition risk according to their own risk perception and market strategy (since the *MRE* mechanism is automatically activated). Moreover, the Brazilian short-term market just corresponds to a mechanism to settle differences rather than a true market given that there is not a bid based dispatch. On the other hand, neither the short-term price (*PLD*) nor the dispatch schedules are determined by the market given that they are both a result of the application of a chain of software models that are operated by the Brazilian ISO.

If a more market-oriented approach is to be adopted, several dilemmas have to be faced:

- 1) Efficiency of the use of energy resources - putting into perspective the dichotomy between a centralised dispatch (based on a cost hierarchy) and a decentralised one (based on a market approach), as discussed in [2], it becomes clear how important it is to coordinate the use of water stored in the reservoirs to safeguard the efficiency of using the energy resources, i.e. to take advantage of the all potential energy stored in the cascades. A decentralised dispatch, e.g. a scheme of bids in a pool market, can essentially correspond to a short term process and the inter-temporal features of river chain operations will not be fully represented, if a pure single-period market clearing mechanism is adopted. Likewise, the presence of several owners in the hydro cascades, as is the case in Brazil, endorses a market design based on a centralised dispatch;
- 2) Security of supply - for the time being, the *PLD* has an average value of 37.72 €/MWh, which is close to the hydro average bid price in public national auctions (38.16 €/MWh) [3]. Nevertheless, the *PLD* standard deviation of the entire set of data is around 55.36 €/MWh. With an average of 37.72 €/MWh, considering the hypothesis of a Brazilian electricity market entirely based on a short-term mechanism, this price fluctuation brings a huge risk to the business. Thereby, it is recognised the need for a capacity mechanism. Nowadays, this concern is addressed via both the contracting scheme where the demand must be fully contracted ex-ante and contracts physically backed, and the ISO dispatch, either through the mechanism of risk aversion embedded in the software package, or using a dispatch out of the merit order authorised by the entity that monitors the supply adequacy in the country;
- 3) Flexibility to comply with bilateral contracts - if in one hand a market design having an ISO central dispatch is to be kept, on the other hand the mechanism to share the short-term risk exposition (*MRE*) is automatically performed. This imposes a kind of “strait jacket” on hydros, especially regarding their risk perception and

commercial commitments. Every time there is a water shortage, the *PLD* increases, there are more dispatched thermal stations and less hydro. Depending on the amount of dispatched thermal units, hydros can be displaced in such a way that *MRE* will not have the extra energy to be shared among its participants. In this case, the *MRE* is not able to cover the risk of generators that have to buy electricity in the short-term market to fill the energy committed in their contracts.

In this context, reference [4] states the most consistent path to deal with these problems: to structure the market design such that risks are allocated to those who take decisions and who hold the responsibility for putting them into force.

## III. A MARKET DESIGN BASED ON VIRTUAL RESERVOIR

In order to overcome the described problems, a new market design is conceived partially bearing in mind the proposal of the Revitalization Committee of the Brazilian Electricity Sector [5]. This market design is based on the concept of energy right accounts conceived as virtual reservoirs and it aims at enhancing the flexibility to enable market participants to uphold their contracts, while still ensuring the efficient use of the energy resources and maintaining the current level of the security of supply. In this new market, each hydro is modeled as an agent that has a virtual reservoir representing how much energy is virtually stored in his hydro plant.

For each accounting period, each account is fed by the corresponding fraction of the total natural affluent energy that flowed to the hydro cascade where the station is located. The ISO continues defining the amount of generation assigned to each power plant, which means that it maintains its responsibility as it currently happens in Brazil: it defines the physical dispatch as a way to optimize the use of the resources, dispatching the hydro and thermal units. Nonetheless, a hydro short-term market is established based on bids for the total demand to be supplied by hydros. In this market, hydros will have the opportunity to withstand their bilateral commercial commitments whereas they try to have successful bids. The result of this market is a virtual dispatch conceived for financial settlement purposes. To do that, hydro agents can offer bids considering:

- a price between zero and a regulatory ceiling price; and
- a quantity available within the balance of his account.

Then, the final short-term price is calculated as a weighted average considering the most expensive successful hydro price bid and the variable cost of the last non-hydro resource dispatched by the ISO.

In this new market design two worlds would then coexist: the real one, associated with the power system considering physical effects, and where the dispatch is performed by the ISO in a centralized way; and the virtual one, related to the settlement system. This virtual mechanism has only commercial effects and hydro agents participate in the mentioned short-term market. Both worlds simultaneously operate, and the link between them is the total affluent natural energy that flowed along the hydro cascades during the

accounting period. Thus, the settlement process will be done considering the successful quantity bid by each hydro participant, and the exposed position will be valued by the new final short-term price, which comes from the combination of thermal costs based on the ISO dispatch and the short-term price arising from the hydro short-term market.

#### IV. SIMULATION OF THE NEW MARKET RULES

To simulate the behavior of the hydropower plants in this market design, an Agent-Based Model - ABM is adopted with the reinforcement Q-Learning. According to [6], among the modeling alternatives for bidding strategy analysis in the electricity short-term markets, ABM is highlighted because it allows designing complex electricity market as collections of rule-based agents interacting with one another dynamically and intelligently, simulating human beings' behavior to build optimal bidding strategies. Generally, the ABM procedure can be described as follows [7]: (1) define the research questions to be addressed; (2) construct a model comprising an initial population of agents; (3) specify the initial model state by defining the agents attributes and the structural and institutional framework of the electricity market within which the agents operate; (4) have the model evolve over time without further intervention; (5) analyze simulation results and evaluate the regularities observed in the results.

Considering the proposed virtual reservoir market design, the ABM research question is as follows: will agents be able to uphold their ex-ante bilateral contracts? Agents represent hydro companies that prepare their bids depending on the amount available to honor their contracts, the level of their (virtual) reservoir, and month of the year, which is associated with the water inflows into the hydro cascade. Their goal is to avoid negative exposures in case of end up providing through their successful bids an amount of energy lower than the contracted. Agents learn how to act optimally using the Q-learning algorithm, which focuses on the impacts of rewards and punishments on their choices to seek their goal.

An agent using Q-learning is a goal-oriented learner that, for a giving Markov decision process, continually interacts with his environment, receives feedback from it, and searches for the most profitable action considering the past experience. In other words, at each time step  $t$  the agent is in some state  $s_t$ , chooses any action  $a_t$  that is available in state  $s_t$ , receives a corresponding reward  $r_t$ , and moves into a new state  $s_{t+1}$ . Thus, in Q-learning it is imperative to carefully structure the set of states  $S = \{s_1, s_2, \dots, s_n\}$ , actions  $A = \{a_1, a_2, \dots, a_n\}$  and rewards  $r_t(s_t, a_t)$ . Lastly, the capability of learning to act optimally in Markovian domains by experiencing the consequences of actions is given by an value function  $Q(s_t, a_t)$ . This function provides the expected utility of taking a given action in a given state [8], and it is given by (1).

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \{r_t(s_t, a_t) + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\} \quad (1)$$

So, for each admissible pair (s, a), the value function is defined as the Q value. The parameter  $\alpha$  is the learning rate, which reflects the degree to which recently learned information will override the old one ( $\alpha$  equal to 0 will make the agent not learn, while equal to 1 will induce the agent to consider only the most recent information), and the parameter

$\gamma$  entitles the discount factor that determines the importance of future reinforcements ( $\gamma$  equal to 0 will make the agent myopic by only considering current rewards, while values closer to 1 turn distant rewards more important). Additionally, the expression  $\max_a Q(s_{t+1}, a)$  represents the best the agent thinks it can do in state  $s_{t+1}$ .

In addition, in order to adequately balance the capacity not to converge to local optima and the acceleration of the learning process, a Simulated Annealing (SA) mechanism is included into the developed algorithm as detailed in [9]. An overview of the entire algorithm is present in Figure 1.

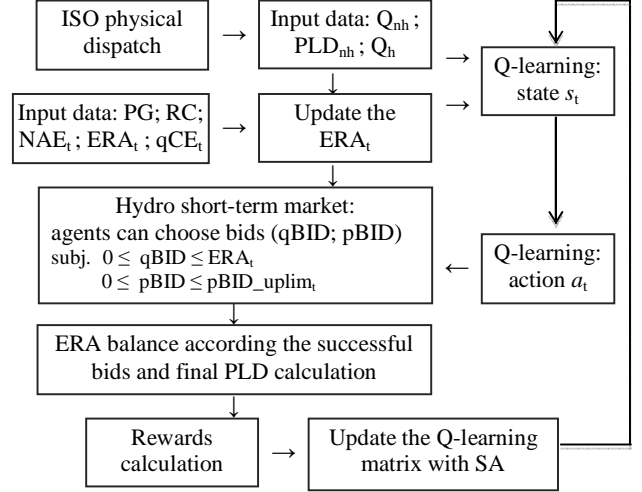


Figure 1. Overview of the developed algorithm.

In brief, the algorithm gets from the ISO physical dispatch procedure the total demand to be supplied by non-hydros ( $Q_{nh}$ ) and by hydros ( $Q_h$ ), and the variable cost of the last non-hydro dispatched unit ( $PLD_{nh}$ ). It also requires information about ex-ante bilateral contracts ( $qCE_t$ ) for each account period  $t$ , i.e. the amount of energy that must be endured in the hydro short-term market by the bids. Other input data are as follows: the natural affluent energy that flowed in period  $t$  to the hydro cascade ( $NAE_t$ ), from which a fraction is allocated to each hydro giving their physical guarantee ( $PG$ ); the reservoir capacity ( $RC$ ), and the virtual reservoir level or energy right account ( $ERA_t$ ) in its initial stage.

The hydro short-term market takes place considering  $Q_h$ , and bids can be done once respected the following constraints: the quantity bid ( $qBID$ ) is limited by the ERA balance in each period  $t$ , and the price bid ( $pBID$ ) is limited by a maximum regulatory price ( $pBID_{uplim_t}$ ). The bids are sorted considering the  $pBIDs$ , the successful bids are then identified, and the ERA updated. The clearing price rule is adopted to define the final price of the hydro short-term market ( $PLD_h$ ), and the final short-term price ( $PLD_{final}$ ) is determined by (2).

$$PLD_{final} = (PLD_h \cdot Q_h + PLD_{nh} \cdot Q_{nh}) / (Q_h + Q_{nh}) \quad (2)$$

The Q-learning is then structured bearing in mind the purpose of the simulation: an analysis over the hydros' flexibility to comply with bilateral contracts. So, the state space is divided in three: state  $s_1$ , when the amount available into the ERA is higher than the ex-ante contract ( $qCE_t$ ); state

$s_2$ , when ERA balance is equal or lower than  $qCE_t$ ; and state  $s_3$ , if the  $qCE_t$  is equal to zero. Moreover, it is possible to offer twelve different bids (actions), each one formed by a pair (qBID, pBID). The action space is formed by qBIDs equal to zero, 100% of the ERA, or 100% of the qCE, depending on the state. The pBID can be 100%, 75%, 50%, 25% or 0% of the maximum allowed price bid for the account period (pBID\_uplim<sub>t</sub>). Both the state and action space are illustrated in Table I, which corresponds to the Q-learning matrix.

TABLE I. ILLUSTRATION OF THE Q-LEARNING MATRIX

States	ERA > qCE	$s_1$												
	ERA ≤ qCE	$s_2$												
	qCE = 0	$s_3$												
			a1	a2	a3	a4	a5	a6	a7	a8	a9	a10	a11	a12
pBID (% of the pBID_uplim <sub>t</sub> ) =	0	100	100	75	50	25	0	100	75	50	25	0		
qBID =	0													
			100% . ERA				100% . qCE							
			Actions											

Some of the actions are just allowed for a specific state (white cells in Table I), and the reward is given by the difference between the successful quantity bid (qBID<sub>suc<sub>t</sub></sub>) and the amount committed through contracts for that period (qCE<sub>t</sub>), as shown in (3).

$$\text{reward} = \text{qBID}_{\text{suc}_t} - \text{qCE}_t \quad (3)$$

## V. RESULTS

A case study that includes four hydro power plants with the characteristics in Table II was designed to test the aforementioned flexibility. In this example, these hydros compete to (virtually) supply a variable load  $Q_h$  aiming to achieve their goal: have successful bid to avoid having to buy energy in the short-term market to complete their contracts. In Brazil the market operator compares every month the commercial commitments with the produced/consumed energy in order to set the exposed position of each market player. Accordingly, the current version of the simulated hydro short-term market also operates in monthly basis.

TABLE II. CHARACTERISTIC OF THE HYDROS

Hydro	H1	H2	H3	H4
Physical guarantee (PG)	125	100	50	225
Reservoir capacity (RC)	500	400	200	900
ERA in $t=1$	0	0	0	0

The simulation of the system considers three different weather patterns (year with *high*, *medium* and *low* water inflows) and market conditions (level of ex-ante contract – qCE equal to 100%, 50% and 0% of the physical guarantee). Furthermore, two types of seasonalization are performed on the bilateral contracts: flat seasonalization (*FLAT<sub>seaso</sub>*), where the annual energy committed through the ex-ante contracts are distributed in equal monthly amounts; and a seasonalization according to the NAE (*NAE<sub>seaso</sub>*), in which the annual amount of the contracted energy is allocated considering weights determined by the natural affluent energy that flowed to the hydro cascade.

With the aim of covering all conditions described above, 18 scenarios were run. The algorithm converges in average after simulating 2,500 years and considering a tolerance of 0.5% on the values that are included in the Q-learning matrix built in each iteration regarding the matrix of the previous one. Table III details the scenarios, and shows the optimal bidding strategy obtained for the three states and for each hydro.

TABLE III. THE OPTIMAL BIDDING STRATEGY

Scenarios		Best action policy													
Water flow	seaso	qCE level	n°	State $s_1$				State $s_2$				State $s_3$			
				Hydro				Hydro				Hydro			
				1	2	3	4	1	2	3	4	1	2	3	4
high	FLAT	100%	1	a12	a12	a12	a12	a3	a3	a5	a6	-	-	-	-
		50%	2	a12	a12	a12	a12	a6	a7	a6	a6	-	-	-	-
		0%	3	-	-	-	-	-	-	-	-	a1	a1	a1	a1
	NAE	100%	4	a12	a12	a12	a12	-	-	-	-	a1	a1	a1	a1
		50%	5	a12	a12	a12	a12	-	-	-	-	a1	a1	a1	a1
		0%	6	-	-	-	-	-	-	-	-	a1	a1	a1	a1
medium	FLAT	100%	7	a12	a12	a12	a12	a3	a5	a4	a4	-	-	-	-
		50%	8	a12	a12	a12	a12	a5	a3	a3	a3	-	-	-	-
		0%	9	-	-	-	-	-	-	-	-	a1	a1	a1	a1
	NAE	100%	10	a12	a12	a12	a12	a7	a7	a3	a3	a1	a1	a1	a1
		50%	11	a12	a12	a12	a12	-	-	-	-	a1	a1	a1	a1
		0%	12	-	-	-	-	-	-	-	-	a1	a1	a1	a1
low	FLAT	100%	13	a12	a12	a12	a12	a7	a7	a7	a7	-	-	-	-
		50%	14	a12	a12	a12	a12	a4	a7	a4	a3	-	-	-	-
		0%	15	-	-	-	-	-	-	-	-	a1	a1	a1	a1
	NAE	100%	16	a12	a12	a12	a12	a7	a7	a7	a4	a1	a1	a1	a1
		50%	17	a12	a12	a12	a12	-	-	-	-	a1	a1	a1	a1
		0%	18	-	-	-	-	-	-	-	-	a1	a1	a1	a1

### A. High water flow scenarios

In all *FLAT<sub>seaso</sub>* scenarios, ERA lower than qCE occurs during dry months (June, July and August). When this happens it is not possible to withstand all the energy committed by contracts (state  $s_2$ ). Nevertheless the chosen action adopts the strategy of using all the available energy in the virtual reservoir to mitigate the punishment (actions from a3 to a7). In wet months (state  $s_1$ ), the best action is a12, i.e. to bid a quantity equal to qCE at zero price.

Regarding *NAE<sub>seaso</sub>* scenarios (no. 4, 5 and 6), no episode registered an ERA lower than qCE (state  $s_2$ ) even when hydros are entirely committed, i.e. their physical guarantee is 100% committed through bilateral contracts. In dry months qCE can be at zero (state  $s_3$ ) since NAE is equal to zero and consequently the seasonalized qCE is also equal to zero. In scenarios no. 4 and 5, the best strategy when qCE is different from zero is to offer 100% of the qCE at a price equal to zero (action a12). It can be noted that choosing a12, there will be no negative financial exposition. Additionally, due the available water typical of the good hydrological year, the chosen price bid is zero. If there is an attempt to push price closer to the ceiling price (i.e. to choose a pBID different from zero), owing to the competition in the short-term market there will be a risk of not being (virtually) dispatched to uphold the ex-ante contract. Among these scenarios, no. 4 is the one that better illustrates the flexibility to comply with contracts given

that hydros are 100% committed. For this scenario, Figure 2 shows that qBIDs (dots) are always equal to qCE values given in the lines for all hydros ( $i = 1, 2, 3$  and 4).

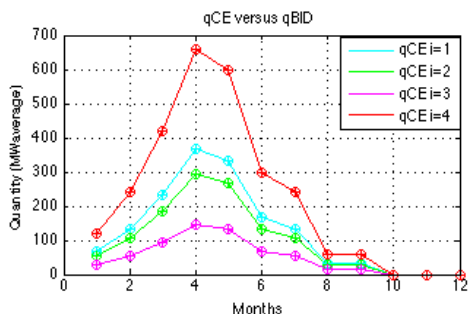


Figure 2. Quantity bids versus ex-ante contracts in scenario no. 4.

### B. Medium water flow scenarios

As with the high water flow scenarios, in the medium water scenarios the same pattern of bidding strategy also occurs. However, some changes are noticed. In *FLAT\_seaso* scenarios, once there is no so much water flows as in the equivalent previous scenarios, the best actions have higher price bids. So, when the level of bilateral contracts are not so high (e.g. at 50%), by comparing actions from scenario no. 2 (a6, a7, a6 and a6) and from scenario no. 8 (a5, a3, a3 and a3) during dry months (state  $s_2$ ), we can notice that as inflows begin to decrease, prices tend to rise. This was the expected result as resources become scarce.

Focusing on *NAE\_seaso* scenarios, by analyzing scenarios no. 4 and no. 10 we can conclude that state  $s_2$  started to occur because we are considering medium water flow scenarios. This means that there were months in which the energy available in the virtual reservoir (ERA) was lower than the amount needed to comply with the contracts (qCE).

### C. Low water flow scenarios

Lastly, the worst scenario occurs when the hydrological year is bad, the seasonalization process is performed in its flat mode, and the energy from hydros is fully committed through bilateral contracts (scenario no. 13). In this case the best bid strategies are: a12 for state  $s_1$  and a7 for state  $s_2$  (there are no events in state  $s_3$ ). Figure 3 presents the qCE x qBID results for scenario no. 13 (*FLAT\_seaso*). As it can be observed, qCE cannot be kept during all months by qBIDs because there is not enough energy available in the ERA at several periods along the year.

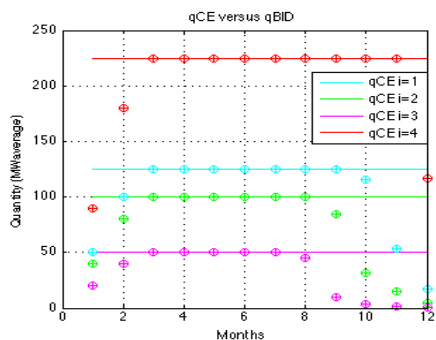


Figure 3. Quantity bids versus ex-ante contracts in scenario no. 13.

## VI. CONCLUSION

The reported results indicate that under several different scenarios agents are capable to act on their own in order to seek the best strategist to sustain their bilateral contracts. And they are allowed to that through the hydro bid based market. Notably, this market design contains a greater degree of freedom than the rules of the seasonalization of the physical guarantee (and its annual “window” for the monthly allocation) and the MRE (where energies are automatically shared). By replacing the *MRE* and the seasonalization process by the virtual reservoir model maintains the current levels of efficiency and security remain the same, but there is an increase in the level of flexibility regarding the commercial behavior of the agents. This is because the management of (virtual) reservoirs is under the responsibility of each hydro, which could (virtually) save water according to their own risk perception, while the operation of the physical system is not affected, ensuring the efficiency of the hydro cascade and maintaining the current level of the security of supply.

Other advantages of this market design can be noted as follows. First, it promotes an increased transparency related with computational models used by the ISO to run the centralised dispatch. Once the codes associated with this software package are under intellectual property rights, inconsistencies in these algorithms have a huge impact within the entire sector. A mix of centralized dispatch and market based on bids can, therefore, increase confidence of the electricity market. Second, with both the physical and virtual dispatch operating in parallel, it can be possible to promote a monitoring of the ISO performance based on comparisons of decisions, namely the ISO decisions versus agents decisions.

Moreover, as future improvements in the algorithm extra bids will be implemented aiming at allowing the agents to manage their reservoirs in order to optimize the leftover stored energy. That is, apart from the management of their bilateral contracts, agents would be allowed to get extra profit in the short-term market when there is more energy into the reservoir than the need to comply with the bilateral ex-ante contracts.

## REFERENCES

- [1] ANEEL – Brazilian Electricity Regulatory Agency. Database of electricity generation [Online]. Available: <http://www.aneel.gov.br>
- [2] A. Philpott, Z. Guan, J. Khazaei, G. Zakeri, “Production inefficiency of electricity markets with hydro generation,” *Utilities Policy*, vol. 18, pp. 174-185, Dec. 2010.
- [3] CCEE - Electric Power Commercialization Chamber. Prices. [Online]. Available: <http://www.ccee.org.br>
- [4] IEA - International Energy Agency, *Energy market experience: Lessons from liberalized electricity markets*. Paris: OECD/IEA, 2005.
- [5] Revitalization Committee of the Brazilian Electricity Sector. Final Report. Brasilia: Ministry of Mines and Energy, 2002.
- [6] Li G., Shi J., Qu X. Modeling methods for GenCo bidding strategy optimization in the liberalized electricity spot market: A state-of-the-art review. *Energy*, vol. 36, pp. 4686-4700, 2011.
- [7] Weidlich, A.; Veit, D. A critical survey of agent-based wholesale electricity market models. *Energy Economics*, 30, pp. 1728-1759, 2008.
- [8] Watkins, C.; Dayan, P. Technical Note: Q-Learning. *Machine Learning*, 8, pp. 279-292, 1992.
- [9] M. Guo; Y. Liu; J. Malec, “A new Q-learning algorithm based on the metropolis criterion”, *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 34, n. 5, pp. 2140-2143, Oct. 2004.