Strategic bidding of ancillary services for a hydro power producer

Hubert Abgottspon
Power Systems Laboratory
ETH Zurich, Switzerland
abgottspon@eeh.ee.ethz.ch

Göran Andersson
Power Systems Laboratory
ETH Zurich, Switzerland
andersson@eeh.ee.ethz.ch

***Abstract*—This paper presents an agent-based simulator for examination of a secondary control market dominated by hydro power producer as decision support for one of the market participants. Proposed is a Q-learning algorithm for determining possible strategic behavior. Adaptive learning is made possible by application of certain characteristics to agents quantity-price pairs bids. Considered are for each agent its portfolio of different hydro power plants with their water values estimated by a stochastic dynamic programming scheme. The simulator is applied to the Swiss system where strategic behavior will be shown. Additionally it is analyzed how single agents could make use of strategic behavior in case of special occurrences in the market.**

***Index Terms*—Ancillary services market, price-maker, hydro power, multiagent systems, stochastic dynamic programming, self-scheduling.**

I. Introduction

A. Motivation

In a deregulated electricity market environment the self scheduling of a hydro power producer typically is a bidding problem. Apart from selling of energy the offering of ancillary services is relevant for this problem. An optimal self scheduling optimization therefore must comply with at least two requirements:

1) Opportunity costs of stored water, the water values, have to be calculated considering the available market products, stochastic prices of these market products, stochastic water inflows, sufficient time horizon and relevant technical constraints. In literature this multistage stochastic problem is often solved by dynamic programming, reviewed e.g. in [1]. The offering of ancillary services is usually neglected if a non short-term perspective is considered. However for a pumped storage plant in Switzerland a medium-term time horizon of one year with hourly time steps have to be taken into account because of the importance of seasonal and daily water inflows. On the other hand a simplified optimization without the possibility to offer ancillary services leads to unrealistic results. In [2] we have shown how to tackle this problem and how to compute realistic water values.

2) For Switzerland the most interesting ancillary services market is the market for provision of secondary control power for frequency stabilization (spinning reserve, automatic one-minute control). This market can provide significant income for hydro power plants. However the market is relatively small and dominated by a few big players. So this market is of oligopolistic nature where strategic bidding is present and additional profit could be obtained if considered.

Agent-based modeling is able to model such a problem. [3]–[6] outline the research about this concept applied to energy markets. Agent-based modeling is mostly used for analyzing different market structures and not for self-scheduling. The few other works consider thermal or hydrothermal portfolios for an optimal bidding problem. This is similar to our problem, where opportunity costs can be minimized. For example in [7] game theory techniques were utilized to locate optimal Nash equilibrium solutions to the electricity market auction. Considered were apart from the energy market spinning reserves and reactive power markets with two players with thermal production.

Apart from agent-based modeling, mathematical optimization of estimated residual demand curves with market data can be used to find optimal strategic bidding. For example the authors in [8] use two estimated residual demand curves of one player and its competition in order to find optimal bidding considering day-ahead and intra-daily energy markets as well as secondary reserve market.

In our case one secondary control bid consists of several quantity-price pairs, from which the market operator can choose one of these pairs. In all mentioned approaches there is no solution to this problem. For standard agent-based modeling the algorithm either quickly gets intractable or the model has to be unrealistically simple, because of the large freedom of quantity-price pairs bids. For residual demand curves it would be possible to include such bids in the optimization. However it is first very difficult to model the demand curve of the competition and secondly it is even more difficult to model their strategic behavior, thus resulting again in an unrealistic setting.

If many quantity-price pairs would be present this could get approximated by a marginal cost curve. In [10] deviated slope and markup of such a marginal cost curve are used to model discrete strategic choices. However this method is not suitable here because there are not many pairs present, as it will be shown later. Apart from that this method would

This work was carried out within the scope of the project "Hydro Power Planning", supported by KTI (Swiss Innovation Promotion Agency), project 11635.2 and Axpo AG.

propose unrealistic strategic choices for this self scheduling problem.

B. Objective and proposed solution concept

The overall goal is the determination of an optimal bid of secondary control for a generation company with several pumped hydro storage power plants. This bid consists of quantity-price pair combinations. In order to comply with the two requirements for an optimal self-scheduling the following procedure is proposed: The secondary control market players are aggregated into agents, each agent representing one of the bigger market players. For all agents water values are calculated. The agents are modeled from public available data, water inflows are estimated based on typical inflows for the respectively locations and market prices are taken from an hourly price forward curve. The possibility to offer ancillary services is considered, however with a quantity only bid for a given estimated price. This multistage stochastic problem is solved by a dynamic programming scheme and is not further explained in this paper.[1]

After the water values are determined an agent based modeling approach is used to model strategic behavior of the agents. Depending on their production capabilities and water values optimal quantity-price pair bids are found.

A novelty within this approach is the direct consideration of several quantity-price pairs per bid without approximating it in a marginal cost curve. This results in a more realistic setting usable for a self scheduling of ancillary services in a hydro dominated secondary control market. In order to remain tractable the bids are characterized depending on marginal costs of the respective agent. The performance of each bid trial is determined by a profit simulation for each agent respecting estimated spot prices and again the water values. The clearing process itself is modeled as a binary linear problem.

With this novel approach of combining stochastic dynamic programming with agent based modeling a hydro power producer not only get an idea on how much secondary control he should offer but also on which quantity-price pairs he should focus. There should be stressed that although the bidding of energy is considered the algorithm gives no optimal energy bidding strategy in this respect. The idea is that this should be done as a second step daily or hourly and not weekly as it is the case for secondary control.

The paper is organized as follows: Section II explains the secondary control market in Switzerland and introduces the model. Section III outlines and explains the case study where the input data is briefly motivated and interesting results are shown. Finally section IV concludes the paper.

II. Model

The secondary control market in Switzerland (more details can be found here [11]) is operated by the transmission system operator (TSO). After a pre-qualification generation companies

[1]More details about this stochastic dynamic programming framework can be found in [2].

TABLE I
VARIABLES

Variable	Explanation
$i \in \mathrm{I}$	agents
$a^i \in \mathrm{A}$	action a taken by agent i
$Q^i : \mathrm{A} \rightarrow \mathbb{R}$	stored function of expected profit for given action
t	game round
$\alpha^i_t \in [0,1]$	degree of correction
$r^i(a^1_t, \ldots, a^n_t)$	reward for agent i for actions $a^1_t, \ldots, a^n_t$
$b \in \mathrm{B}_t$	bids, constructed out of k quantity-price pairs
$s_k \in \mathrm{S}_t$	price, demand charged per quantity [Euro/MW/h]
$q_k \in \mathrm{Q}_t$	offered secondary control power quantity [MW]
wv_{pp}	water value for respective PP [Euro/MWh]

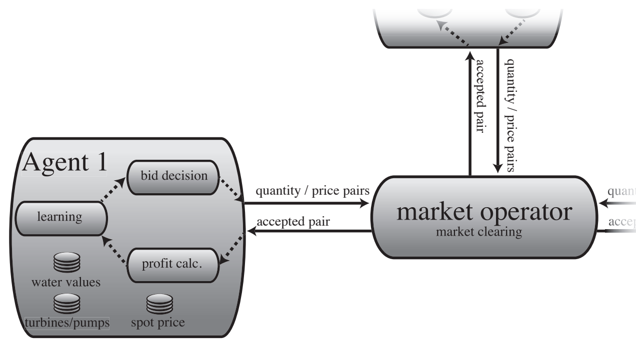

Fig. 1. Overview of the algorithm with agents deciding on their bids and the market operator who clears the market. The agents are each specified by their water values and production capabilities. Illustrated is agent 1 whereas two other agents only appear in outlines.

(GENCOs) are invited to bid provision of secondary control power on a weekly basis. The power has to be provided for the whole week in symmetrical capacity blocks of at least 5 MW with increments of ± 1 MW. It is allowed to fulfill this requirements from a pool of generating units. The bid $b \in \mathrm{B}$ itself is defined as a number of combinations k of the volume offered $q_k \in \mathrm{Q}$ and demand charged $s_k \in \mathrm{S}$, so-called quantity-price pairs. The number of combinations k are not limited.

The TSO can select at most one of the quantity-price pairs within each bid. This is done by a market clearing optimization where those pairs are selected that meet control demand with least costs.

The market participants are modeled as agents $i \in \mathrm{I}$. In order to simulate strategic behavior the agents should be able to learn by acquiring knowledge from past actions and decide for upcoming actions based on their experience. Reinforcement learning (reviewed e.g. in [12]) models the learning process through repeated interactions. Proposed in this paper is an algorithm based on the well-known Q-learning framework (introduced in [13], extended in various papers, e.g. in [14]).

Fig. 1 shows an overview of the proposed algorithm. The agents are characterized by their technical production capa-

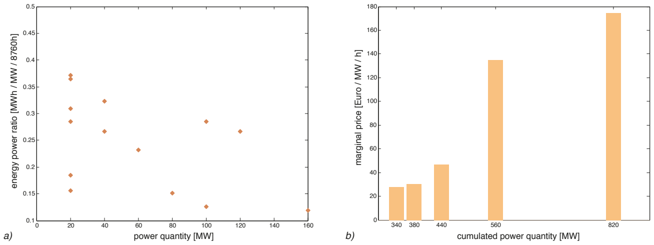

Fig. 2. a) Realizable amount of secondary control power per PP in an agent's portfolio vs. energy-power ratio. b) Cumulated power quantities and associated marginal prices for this portfolio.

bilities and water values. For each agent three tasks have to be done: bid decision, profit estimation for a given accepted bid and memorizing of relevant knowledge. Apart from the agents also the market clearing has to be simulated. The algorithm is repeated until sufficient learning has occurred and stable results are found. In the following all of these tasks are explained in more details.

A. Agent model

Each agent should represent one of the most influencing market participants. Typically those GENCOs have several different PPs or certain shares of them in their portfolios (see Fig. 2 a)). In order to get to the relevant information for bidding of secondary control, the following procedure is proposed:

1) Calculation of water values per PP
2) Determination of secondary control power per PP
3) Estimation of marginal costs per PP
4) Clustering of PP depending on marginal costs
5) Determination of power quantity for provision of secondary control per agent

First water values wv_{pp} have to be calculated for each PP in the agents portfolio. For this paper a stochastic dynamic programming approach was used for this calculation (see also [2]).
Secondly depending on technical properties a PP may provide a certain amount of secondary control power (see Fig. 3).
Afterwards marginal costs are calculated, which means the smallest remuneration for the provision of secondary control which make this provision still beneficial. For this calculation a profit estimation with provision of the secondary control bid is compared with estimated profit without it. This calculation is explained later. The PPs within a portfolio are then grouped together depending on the marginal costs.
Finally it is now possible to construct for each GENCO a list of secondary control power associated with their marginal costs. Note, that this list would lead to the most profitable bidding in a perfect market, which each quantity-marginal costs pair forming one bid. Note also that there are not enough quantity-price pairs in a typical portfolio in order to approximate a marginal cost curve (e.g. Fig. 2 a)).

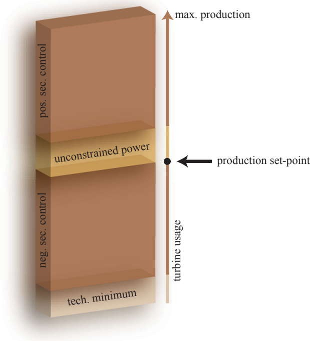

Fig. 3. Production set-point of a hydro PP with technical minimum, provided secondary control and unconstrained power range.

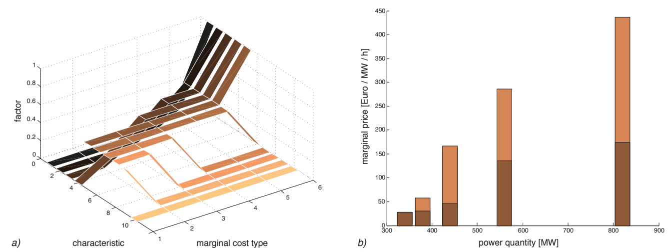

Fig. 4. a) Factors for each characteristic and marginal cost type. b) Example: increase of prices by application of characteristic 4.

B. Bid model

As already mentioned, secondary control bids are defined as a number of quantity-price pair combinations. For a GENCOs the question arises, what kind of bid they should focus on and how to adjust the bid if new knowledge about the performance of a bid is available. The idea proposed here is to adjust the list of cumulated power quantities with associated marginal costs (Fig. 2 b)) depending on certain *characteristics*.
It seems reasonable to bid secondary control based on its marginal costs. However in an oligopolistic environment higher prices for certain quantities could be beneficial. To be able to learn this, bid characteristics are introduced. A bid characteristic could be e.g. a strategy where the price for the quantity with the highest marginal cost would be further increased. Proposed are ten different characteristics which construct the discrete action set A from which the agents choose their actions $a^i \in \mathrm{A}$.

The question remains on how much each price has to be increased. The following assumptions were made:

- each price s_k is at least the corresponding marginal cost
- each price s_k is at maximum the prices of the quantities with higher marginal costs $s_{k+1}, s_{k+2}, \ldots$

The first assumptions seems clear and the second assumption is due to the bidding of cumulated quantities. This leads to adjusting the *differences* between the marginal prices based on the chosen action a. Fig. 4 a) shows the factors by which each difference is multiplied and added to each marginal cost. On the right side it is shown for the forth factor how the prices

increase.
It should be noted, that similar characteristics are close and that the first characteristic is also similar to the last one.

Whereas one bid will consists only of a few different prices the offered quantity can vary much more because an agent can bid each power quantity per PP individually or bid even a fraction of it. This makes sense since the demanded power is limited so that a bigger quantity is less probable for being accepted. Therefore the corresponding cumulated power quantity for each price (Fig. 2 b)) is split into increasing values up to the full amount. This results in many quantity-price pairs for each different marginal costs type.

C. Bid decision

In each game round the agents have to decide which characteristic to apply on their secondary control bid, the respective actions $a^i \in A$. The actions are first sorted, so that similar actions are close together and are given a number $j \in \mathbb{N}^+$ (see also Fig. 4 a)). The agent i selects his action based on normal distribution with some standard deviation and the mean of the number j_{max} of the action which maximizes its believed reward:

$$\mu = j_{max} = \arg\max_{a \in \mathrm{A}} Q_t^i$$

The result is then rounded to the nearest integer and the respective action is taken.
With this procedure each agent chooses actions which are similar to the one with the best reward. In order to introduce some complete randomness actions are drawn for a certain probability out of a uniform distribution.

D. Profit estimation

Each agent has to estimate the reward r^i for the set of actions for a game round t. From the market operator he gets the accepted quantity-price pair q^{acc}, s^{acc}, if any. The provision of secondary control limits the production capabilities, as depicted in Fig. 3. In order to be able to deliver negative secondary control the PP has to be running at a certain setpoint. The remaining unconstrained power range can be used for producing energy to bid in the pool market.
To model this pool market an estimated hourly priced forward curve (HPFC) is used. The profit from this market depends on the difference between market price and water values wv_{pp} of the respective PP. For the sake of simplicity a number of simplifications are made:

- the water values wv_{pp} remain the same for the whole week.
- water inflows as well as water balance in the basins are neglected.
- same water values are assumed for PP with similar marginal secondary control costs

For bigger basins where weekly production and water inflows does not influence the filling much the first two simplifications are reasonable. However for smaller basins those simplifications would result in large errors.
The third simplification is only valid, if the PPs are technically similar.

With these simplifications the PP-portfolio can be clustered. To estimate the profit with this portfolio in the pool market with given accepted quantity-price pair q^{acc}, s^{acc}, a linear program can be formulated. For each PP in the portfolio there is:

$$\begin{aligned}
&\max\ (\mathrm{HPFC} - \mathrm{wv}_{pp})^T \cdot x \\
&\text{s.t.} \\
&0 \leq x \leq \text{unconstrained power}_{pp}(q^{acc})
\end{aligned}$$

x denotes hourly bidding in the pool market, where it is constrained depending on the accepted secondary control quantity q^{acc} as well as the technical minimum (Fig. 3). The profits of all PP in the portfolio together with the remuneration of the accepted quantity-price pairs $(q^{acc})^T \cdot s^{acc}$ leads to the reward r^i for each agent. Note that alternatively to the linear program a direct computation would be also possible.

E. Memory / learning algorithm

In a Q-learning algorithm the agent i keeps in memory a function $Q^i : A \to \mathbb{R}$, which represent the expected profit previously calculated for action $a^i \in A$. The agent updates his memory after each game round t. This is done as following:

$$Q^i(a_t^i) \leftarrow Q^i(a_{t-1}^i) + \alpha_t^i \left(r^i(a_t^1, \ldots, a_t^n) - Q^i(a_{t-1}^i) \right)$$

$\alpha_t^i \in [0, 1]$ is known as degree of correction specifying how much new knowledge change the memory. r^i denotes expected reward for agent i if actions $a_t^1, \ldots, a_t^n$ are performed with n as the number of agents.
So if $\alpha_t^i = 0$ the agents leaves the memory unchanged, if $\alpha_t^i = 1$ the agent doesn't consider past observations at all. For this paper we choose α_t to be the same for all agents.

F. Market Clearing

The market operator collect the bids from the agents and performs a market clearing. The operator can select at most one of the quantity-price pairs within each bid. The sum of the selected power quantities has to exceed the control demand. This is a typical optimization problem, which can be modeled as a binary linear program:

$$\begin{aligned}
&\min\ s^T \cdot x \cdot q \\
&\text{s.t.} \\
&q^T \cdot x \geq \text{control demand} \\
&\sum_b x \leq 1 \\
&x \in \{0, 1\}, s \in \mathrm{S}_t, q \in \mathrm{Q}_t, b \in \mathrm{B}_t
\end{aligned}$$

The binary variable x specifies which quantity-price pairs get accepted. Within each bid b only one pair can get accepted.

TABLE II
DATA

Agent (Company)	bid power quantity [MW]	marginal cost [Euro/MW/h]	total installed capacity [MW]	tech. minimum [MW]
1: Alpiq	0 100 20 60 20 540	0 30 38 60 69 152	0 359 187 187 132 1260	0 70 43 19 38 126
2: BKW	0 20 40 20 80 100	0 38 44 57 74 154	0 84 260 55 593 253	0 42 59 5 167 25
3: Axpo	340 40 60 120 260 0	28 31 46 135 174 349	967 158 188 506 715 67	193 61 57 234 109 33
4: EWZ	0 0 0 120 80 0	0 0 0 31 47 94	0 0 0 386 330 54	0 0 0 89 77 27
5: ewb	0 0 0 0 20 0	0 0 0 0 94 188	0 0 0 0 204 27	0 0 0 0 59 3
6: iwb	0 0 20 20 40 100	0 0 31 57 67 154	0 0 81 48 256 253	0 0 15 5 73 25

III. CASE STUDY

A. Input Data

The proposed model is now applied to the Swiss system. In Switzerland storage hydro PPs account roughly for one third of total produced electrical energy. These plants are more than enough to provide the needed amount of secondary control power. The three biggest Swiss GENCOs and the three biggest Swiss public utilities own more than 80% of total capacity. So those six entities were chosen for modeling the agents. The following simplifications were made:

- consideration only of hydro storage PPs with more than 50 MW.
- technical minimum: francis turbines: 50%, pelton turbines: 10%
- further technical issues were disregarded
- water values calculated only for six reference PP
- minimum amount of 20 MW for secondary control provision

Water values are highly depending on the ratio of yearly produced energy to installed capacity. That's why the PP are clustered based on this ratio and are allocated a water value out of six reference ones. The six reference water values are estimated based on a stochastic dynamic programming optimization of six different typical PPs in Switzerland for the last week of June. At this time point the storage basins are usually half filled. The taken HPFC is also the estimated price curve for this week done some days beforehand. The resulting data is summarized in Table II.

Demanded are 400 MW of secondary control power. The algorithm is iterated in parallel, but the knowledge is shared repetitively. CPLEX is used for solving the linear and binary program. Ten characteristics are modeled as already shown in Fig. 4. The results are shown for 4000 game rounds which took around 30 seconds on a standard computer, with a quad-core 2.3 GHz Intel Core i7 processor and 8 GB of RAM.

There should be noted, that the shown results depend heavily on the estimated water values and HPFC. That's why we believe it is very important to estimate those carefully.

B. Results

In Fig. 5 a) the development of the market operators costs for the market clearing is shown (brown curve). The orange curve indicates the costs if every agent bid marginal costs. The brown curve was smoothened by a moving average filter of a window of 5 values in order to reduce spikes. Important to

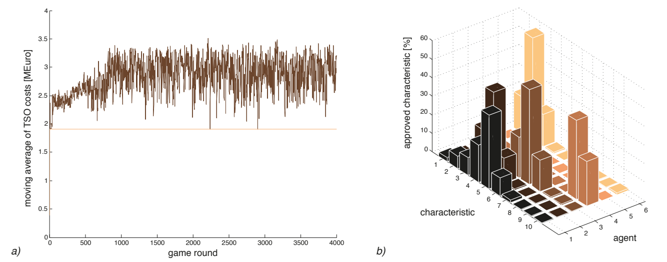

Fig. 5. a) Development of costs of the market operator for the market clearing. b) Approved characteristic per agent in % of total number of game rounds.

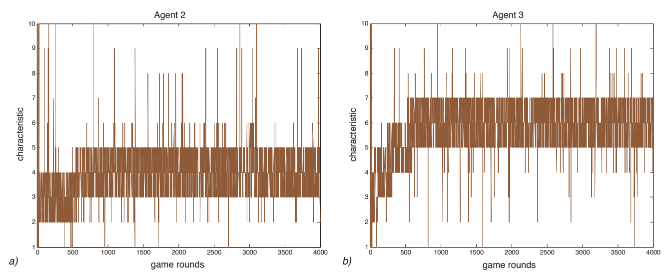

Fig. 6. Bid decision: chosen characteristics for a) agent 2 and b) agent 3.

note is the increase in costs which could be seen as proof that strategic behavior is present.

On the right hand side in Fig. 5 b) for all agents the characteristics are shown, which bids were accepted by the market operator. The bids from agent 5 were seldom accepted which is obvious if the data in Table II is inspected.

The other agents however each want to get their bids

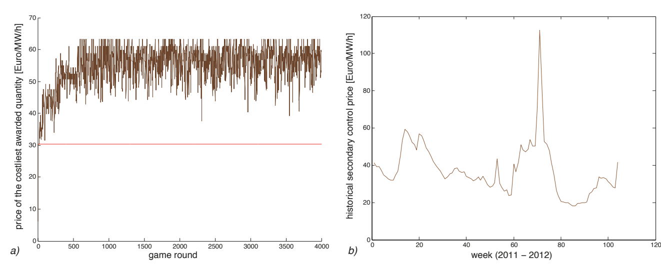

Fig. 7. a) Simulated prices of costliest awarded quantity. b) Historical values of the average price of the costliest awarded 20 MW.

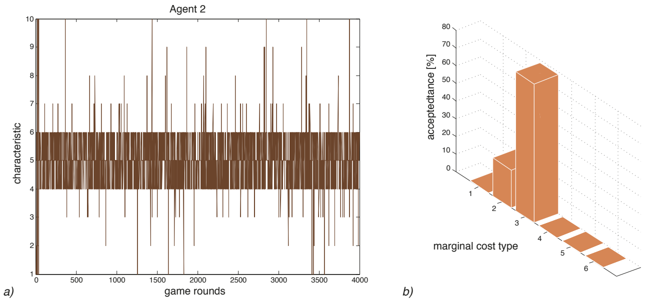

Fig. 8. Market with reduced amount of secondary control quantity and only agent 2 able to act strategically: a) Chosen characteristics. b) Acceptance of marginal cost types for agent 2.

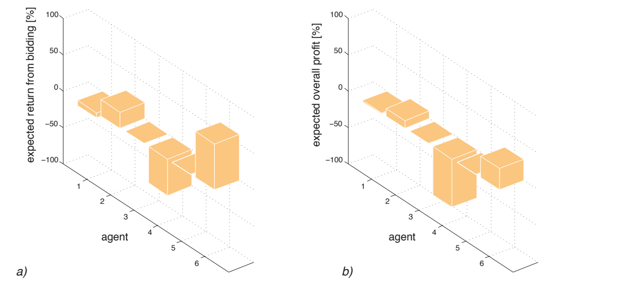

Fig. 9. Comparison of simulations with and without agent 2 acting strategically. a) Expected return from bidding in % of expected return without learning. b) Expected overall profit if pool market is also considered in % of overall profit without learning.

approved and there are certain characteristics where this is more probable. Those characteristics are then also chosen more frequently, which can be seen in Fig. 6. After around 2000 game rounds the results stabilize. Agent 2 chooses characteristic 4 (increased prices for medium and high marginal costs)[2] and agent 3 characteristic 6 (increased prices for low and medium marginal costs)[2]. Note, that the spikes in the figures are due to frequent complete randomly chosen characteristics.

Fig. 7 compares simulated prices for the costliest awarded quantity with historical values, the only public available data about the market clearing in secondary control market in Switzerland. The historical prices are around 20% lower than the simulated ones which indicates either model inadequateness and/or missing strategic behavior in the real market. If the agents would bid their marginal costs, the costliest awarded quantity would be around 30 Euro/MW/h (orange curve in Fig. 7 a)), which would fit historical ones well.

In the view of a self-scheduling problem of one of the agents the question arises how an agent could make use of this agent-based simulation. The results clearly suggest that strategic bidding is only beneficial, if every agent pursues this strategy. So if only one agent is given the opportunity to bid strategically, he will choose to bid marginal costs. So in reality there is a high probability that despite an existing Nash-equilibrium no agent bids strategically.

However there could be special situations, where the secondary control amount offered by the agents is reduced (such situations may have resulted in high prices in Fig. 7 b)). Usually this is known before the actual market clearing is performed and an agent could make use of this knowledge. Fig. 8 shows the results of a simulation, where some PPs are not available and therefore the amount of bided secondary control quantity is reduced. Additionally only agent 2 is learning. In this case agent 2 acts strategically although the other agents bid their marginal costs. Further in comparison with Fig. 6 b) one can see, that first the results stabilize much faster. This was expected since there is only one agent who adaptively changes his bidding strategy. Secondly agent 2 chooses characteristic 5 instead of characteristic 4, which means further increased demanded charges for the first two marginal cost types. Accepted were in this case marginal cost type 3 (see also Fig. 8 b)).

The same simulation was done without agent 2 given the possibility to act strategically, so all agents bid their marginal costs. Fig. 9 shows the comparison between this and the previous simulation. The expected return from bidding increases for agent 2 by more than 20%. If the profit out of energy bidding in the pool market is also considered an increase of more than 10% can be achieved. So it is indeed more beneficial for agent 2 to act strategically in this case. It should be clear that the costs for the market operator increases if agent 2 acts strategically. So most probable the operator would change market rules if he detect this.

[2]See also Fig. 4 a)

IV. Conclusion

This paper presented an agent-based simulator for examination of a secondary control market as decision support for one of the market participants. Used were a Q-learning algorithm for determining possible strategic behavior. Learning was made possible by application of certain characteristics to agents quantity-price pairs bids, which are based on marginal costs. Considered were for each agent its portfolio of different hydro power plants with their water values estimated by a stochastic dynamic programming scheme. The simulator was applied to the Swiss system where it was shown that strategic behavior would be beneficially however would hardly be applied in practice. This result was supported by real data. It was also shown, that in case of special occurrences it would be indeed beneficial for a single agent to act strategically even if the other agents would bid their marginal costs.

References

[1] J. W. Labadie, "Optimal operation of multireservoir systems: State-of-the-art review," *Journal of Water Resources Planning and Management*, vol. 130, no. 2, pp. 93–111, March 2004.

[2] H. Abgottspon and G. Andersson, "Approach of integrating ancillary services into a medium-term hydro optimization," in *XII SEPOPE: Symposium of Specialists in Electric Operational and Expansion Planning*, 2012.

[3] F. Sensfuss, M. Ragwitz, M. Genoese, and D. Möst, "Agent-based simulation of electricity markets - a literature review," *Working Paper Sustainability and Innovation*, vol. S5, 2007.

[4] A. Weidlich and D. Veit, “A critical survey of agent-based wholesale electricity market models,” *Energy Economics*, vol. 30, no. 4, pp. 1728 – 1759, 2008.

[5] E. Guerci, M. Rastegar, and S. Cincotti, “Agent-based modeling and simulation of competitive wholesale electricity markets,” in *Handbook of Power Systems II*, ser. Energy Systems, S. Rebennack, P. M. Pardalos, M. V. F. Pereira, and N. A. Iliadis, Eds. Springer Berlin Heidelberg, 2010, pp. 241–286.

[6] M. Barbati, G. Bruno, and A. Genovese, “Applications of agent-based models for optimization problems: A literature review,” *Expert Systems with Applications*, vol. 39, no. 5, pp. 6020 – 6028, 2012.

[7] A. Morinec and F. Villaseca, “Optimal generator bidding strategies for power and ancillary services using game theory,” in *Power Symposium, 2008. NAPS ’08. 40th North American*, sept. 2008, pp. 1 –8.

[8] A. Ugedo and E. Lobato, “Validation of a strategic bidding model within the spanish sequential electricity market,” in *Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on*, June 2010, pp. 396 –401.

[9] A. Ugedo, E. Lobato, A. Franco, L. Rouco, J. Fernandez-Caro, and J. Chofre, “Strategic bidding in sequential electricity markets,” *Generation, Transmission and Distribution, IEE Proceedings-*, vol. 153, no. 4, pp. 431 – 442, july 2006.

[10] T. Krause and G. Andersson, “Evaluating congestion management schemes in liberalized electricity markets using an agent-based simulator,” in *Power Engineering Society General Meeting, 2006. IEEE*, 0-0 2006, p. 8 pp.

[11] “Swissgrid website.” [Online]. Available: http://www.swissgrid.ch/swissgrid/en/home/experts/topics/ancillary_services.html

[12] L. P. Kaelbling, M. L. Littman, and A. W. Moore, “Reinforcement learning: a survey,” *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.

[13] C. Watkins, “Learning from delayed rewards,” Ph.D. dissertation, King’s College, 1989.

[14] J. Hu and M. P. Wellman, “Nash q-learning for general-sum stochastic games,” *JOURNAL OF MACHINE LEARNING RESEARCH*, vol. 4, pp. 1039–1069, 2003.