An Intelligent Learning Mechanism for Trading Strategies for Local Energy Distribution

Muhammad Yasir, Martin Purvis, Maryam Purvis,
Bastin Tony Roy Savarimuthu

Department of Information Science
University of Otago
Dunedin, New Zealand
{muhammad.yasir, martin.purvis, maryam.purvis, tony.savarimuthu}@otago.ac.nz

Abstract. Although most electric power is presently generated using fossil fuels, two abundant renewable and clean energy sources, solar and wind, are increasingly cost-competitive and offer the potential of decentralized (and hence more robust) sourcing. However, the intermittent nature of solar and wind power presents difficulties in connection with integrating them into national power grids. One approach to addressing these challenges is through an agent-based architecture for coordinating locally-connected energy micro-grids, each of which manages its own local energy production, distribution, and storage. By integrating these micro-grids into a larger network structure, there is the opportunity for them to be more responsive to local needs and hence more cost effective overall. In such an arrangement, the micro-grids have agents that can choose to resell their excess energy in an open, regional market in alignment with respect to their specific goals (which could be to reduce carbon emissions or to maximize their financial outcomes). In this study, we have investigated how agents operating in such an open environment can learn to optimize their individual trading strategies by employing Markov-Decision-Process-based reasoning and reinforcement learning. We empirically show that our learning trading strategies improve net profit loss by up to 29% and can reduce carbon emissions by 78% when compared to the original (non-learning) trading strategies.

Keywords: Renewable Energy, Multi-agent Systems, Power trading, Micro-grids.

1 Introduction

A micro-grid refers to a local energy system that can generate and store its own renewable energy and also be connected to a main electrical energy supply grid [5]. The idea of a micro-grid is to utilize the distributed local renewable energy resources and satisfies power needs locally. It can also sell or buy power from an energy utility company. However, renewable energy sources (wind, sun) are

intermittent in nature and vary hour to hour, even minute to minute, depending upon local conditions [5], which can compromise electric system reliability. Different energy management strategies are used to mitigate or eliminate the impact of supply variations, such as storage devices (batteries, fly wheel, capacitors, etc.), forecasting techniques, demand load management, and backup generators. One of the approaches to address this issue is the interconnection of nearby micro-grids which reduce the impact of non-steadiness of renewable energy sources, as communities having micro-grid can trade power with each other to satisfy their demands [5]. An agent-based architecture for local energy distribution among micro-grids is presented in Yasir et al [16]. In this architecture, each micro-grid represents a community which has its own power generator based on the renewable energy sources and also has its own electric energy demand which varies from hour to hour. Every community has a coordinator agent which, when it has a power surplus or deficit, is responsible for power trading to other interconnected communities or to the utility grid. Each community may employ a different strategy for power trading depending upon either of two goals i.e. profit maximization, and reduction in carbon emission.

In this work, we study the trading activities of coordinator agents that can learn to adapt their strategies in an electric market. The learning trading strategies of the coordinator agents that improve trading performance by employing Markov Decision Process (MDP) and reinforcement learning. Our experimental results show that the learned trading strategies respond to changing conditions and lead to superior out-comes (with respect to financial rewards and reduced carbon emissions), when com-pared with fixed trading strategies.

The rest of the paper is organized as follows: In Section 2 we describe the electric market and the trading strategies, Section 3 covers the MDP learning framework. Section 4, shows empirical comparison of our learning trading strategies. In Section 5, we review related work in this area. Finally, Section 6 discuses some future prospects and also provides a conclusion.

2 Electric Market and Trading Strategies

In our work, we design an hour-ahead market and assume that on the hour of delivery, suppliers always provide the amount of power they had committed in the market. Also, there is one market for all the communities. The market mechanism in our work employs a double auction algorithm [8] to facilitate market clearing (determining the price at which a unit of energy is sold by matching bids and offers). All the sellers and buyers submit their offers and bids in certain time interval. Once the time is over, market clearing algorithm starts making match. In this algorithm, the buyer of the highest bid will be matched by the seller with the lowest bid. The clearing price (or unit price per Kwh) is set as the mean of the (bid and offer) prices. Every match (i.e. a pair of seller and buyer)has its own clearing price. There is no one clearing price for all buyers and sellers. A coordinator agent typically buys from the market; and if no power

is available in the local market or the price is high in the market, it buys from the utility gird.

As discussed above, a community employs a trading strategy for power trading. Some communities may be environmentally conscious and more interested in minimizing environmentally harmful emission, while others may be concerned about the maximization of financial benefits.

2.1 Trading Strategies

In this section, we describe the two main trading strategies that a community can use for trading described by Yasir et al.[16].

Altruistic Strategy (AS) The goal of a community that employs this strategy is the minimization of carbon emissions by using more green power (and thus not buy from the main utility grid, some of whose power is produced by fossil fuels). This strategy also encourages other communities to use renewable energy by selling its own (green) power in the market at the lowest price. So, the agent using this strategy:

- is willing to buy green power at high prices, if green power is available in the market, and
- sells green power in the market at a low price (lower than that charged by the utility), so that more communities can use its power.

Greedy Strategy (GS) A community using this strategy always wants to maximize its financial benefit by buying power at a low price and selling power at a high price in the market. GS always tries to

- Buy at a low price from any place (i.e. market or utility grid)
- Sells energy surpluses at a high price to any place

Fixed Strategy (FS) A community that employs this strategy always offers and bids power in the market at a fixed price. For this paper, the fixed price is 21.5 cents.

Note that the generation cost of electricity is 7 cents per Kwh. So no strategy offers below 7 cents.

In [16] the authors empirically compared AS and GS strategies in term of two variables:

- Net profit/loss by community
- Carbon emission produced as a result of the strategy

There are several limitations in this work. Although AS and GS adjust their current trading price based on the history of how they traded 24 hours ago, they do not learn from past experience. As a result, these communities do not put

optimal trading bids and offers in the competitive electric market and do not get the maximum profit or minimization of the carbon emissions. Similarly, above mentioned strategies make trading bids randomly by adding or subtracting small random numbers from their last bids. They do not know how much exactly they add or subtract from their last bids. Also, while making the bids they do not consider the current position of the market. By using the MDP model, we overcome all these limitations.

In this paper, we propose two types of learning strategies for trading that a community can use. One learning strategy is used to optimize the financial gain and the other leaning strategy is used to optimally reduce the carbon emissions to the atmosphere. The strategy that maximizes profits is called the greedy learner strategy while the strategy that minimizes the carbon emissions is called the green learner strategy. Both strategies learn an MDP policy using Q-learning. The learning algorithm uses a typical exploration-exploitation trade-off that explores more often in early episodes of simulation and progressively less in the later episodes of the simulation.

3 An MDP-based Learning Framework

Let T_s be the Learning Trading agent for which we develop an action policy using the framework of MDPs and reinforcement learning. The MDP for T_s is defined as:

$M^{T_s} = (S, A, \delta, r)$

Where:

- S = si : i = 1....I is a set of states,
- A = aj : j = 1.....J is a set of actions,
- δ(s,a) = $s^{'}$ is a transition function, and
- r (s,a) is a reward function

The state space should capture the two sets of features that are important to how Ts would set its trading prices:

- Last hour average clearing price (ACP) in the market
- Level of electric demand (DL) available in the market for trading for current hour.

The average market clearing price is difficult to represent because prices in the real world are virtually continuous. We restrict the range of prices from 10 cents to 36 cents per Kwh of electricity in New Zealand dollars [4] and discretizing the prices in 2 cents increments to get 13 possible values for the clearing price. Similarly we categorized the level of demand in to three bands: low, medium, and high. Information about ACP and DL is provided by the market to all communities participating in the market.

In the final representation, the state space S is the set defined by all the values of the elements in the following:

$$S = \{ACP_{t-1}, DL_t\}$$

Next, we define the set of MDP actions A. Each learning agent will generate its bid-ding prices by using the information of ACP_{t-1} and DL_t. The learning agent will either increase or decrease its trading price with reference to the ACP_{t-1}. After observing few real world electric markets[3][9], we found that price for the next hour electric power is either increased or decreased up to 2 cents per Kwh as compared to the last hour. The action that an agent can do as a buyer or a seller is to increment or decrement certain amount from the last average clearing price. The range of increments and increments are between -2.5 and 2.5, with discrete options chosen by agents in 0.5 units, resulting in 11 possible actions. So the set of actions is the increment or decrement of last average clearing price from the range of price from 0 to 2.5 cents, and we discretize the prices in 0.5 units cents to get 11 possible values for the actions.

The transition function δ is defined by numerous stochastic interactions within the simulator. The reward function, r, is unknown to the MDP, and it is calculated by the learning agent depending upon its learning objective, i.e. maximize profit or minimize the carbon emissions.

Let T_{sP} be the learning trading agent that wants to maximize the profit. Then the reward function for the seller agent is:

$$r_{sp} = \begin{cases} Amt_{st}/Amt_{max-s} * (p_t - \Psi_{bt}), & \text{if traded inside market} \\ 0, & \text{if no buyer in market} \\ Amt_{st}/Amt_{max-s} * (\Psi_{bt} - ACP_t), & \text{if traded outside market} \end{cases}$$

The reward for the buying agent that wants to maximize its profit is:

$$r_{bp} = \begin{cases} Amt_{bt}/Amt_{max-b} * (\Psi_{st} - p_t), & \text{if traded inside market} \\ 0, & \text{if no seller in market} \\ Amt_{bt}/Amt_{max-b} * (ACP_t - \Psi_{st}), & \text{if traded outside market} \end{cases}$$

where: r_{sp} and r_{bp} are the rewards for selling and buying agents that want to maximize the profit, respectively,
Amt_{st} and Amt_{bt} are the total quantities sold and bought from the market at time t by community
Amt_{max-s} and Amt_{max-b} are the maximum quantities sold and bought from the market by community at any time. We use the denominator values to normalize the value of reward in certain range.
P_t is the price at which the quantity sold is/bought from the market at time t.
Ψ_{bt} and Ψst is the price at which utility grid buys and sells at time t.
Similarly, let T_{sg} be the learning agent that aims to reduce the carbon emission as much as possible.The reduction in carbon emission is only possible if community buys (when it needs) power from the market. Being the green community, it also prefer to sell power inside the market. So the reward function for the selling

agent is:

$$r_{sg} = \begin{cases} Amt_{st}/Amt_{available}, & \text{if traded inside market} \\ 0, & \text{if no buyer in market} \\ -1, & \text{if traded outside market} \end{cases}$$

The reward for learning buying agent for carbon emission minimization is:

$$r_{bg} = \begin{cases} Amt_{bt}/Amt_{needed}, & \text{if traded inside market} \\ 0, & \text{if no seller in market} \\ -1, & \text{if traded outside market} \end{cases}$$

where: $Amt_{available}$ is the total surplus power for selling and Amt_{needed} is the total power quantity required to meet its demand. The value of reward equals to 1 when all the power (available or needed) is sold or bought, from the market.

Since this is a non-deterministic MDP formulation with unknown reward and transition functions, we use the Watkins & Dayans [14] Q-learning formula:

$$Q_t(s,a) = (1-\alpha)Q_t(s,a) + \alpha[r_t + \gamma max_{a'} Q_{t+1}(s^{'}, a^{'})]$$

4 Experiments

This section presents the results of four types of comparative experiments we have conducted using learning trading strategies.

4.1 Experimental Setup

In order to compare the above mentioned strategies, we set up forty micro-grid communities (C1 to C40). The communities have an average hourly consumption of 1150 kWh and a wind turbine of 2000 kW generation capacity. However, these values for an individual community will vary, since the communities are dispersed geographically, and hence have different wind speeds and patterns in their regions. Thus the power produced by each community is also different. Power generated by the wind turbine is calculated by using the formula [6]:

$$P = 1/2\rho AV^3Cp$$

Where P is the power in watts (W), ρ is the power density in kilograms per cubic meter (kg / m^3), A is the swept rotor area in square meters (m^2), V is the wind speed in meters per second (m/s), and Cp is the power co-efficient. We obtained the synthetic wind speed (V) data of New Zealand from Electricity Authority New Zealand [7]. We also obtained hourly power consumption data of nine different places from the Property Services office of the University of Otago [11].

The assumptions made while running our experiments are as follows. All communities are situated at the sea level. So the value of is 1.23 kg/m^3. The blade length of the wind turbines is 45 meter (m). The cut-in and cut-out wind

speeds of the turbines is 3 and 25 meters per second (m/s), respectively. Theoretically the maximum value of Cp is 59%, which is known as Betz limit [6]. However, in practice the value of Cp is in between 25%-45% [6] depending upon the height and size of the turbine. The value of the power co-efficient (Cp) is 0.4 (i.e. 40%). We also assume that the utility grid is always ready to buy power and sell power to the micro-grids at the rates of 18 cents per kWh and 25 cents per kWh, respectively. To trade into the market, a community uses the market-based trading mechanism discussed in section 2.1. The value of learning rate (α) in our Q-learning scheme is 0.5, and the discount factor (γ) is 0.1. When exploiting the learned policy, we randomly select one of the possible actions that is within 20% of the highest Q-value.

We computed the value of net profit/loss[1] and carbon emission[2] by varying the learning and non-learning strategies during simulation. We calculate the amount of carbon emission produced by using electricity emission factor of 0.137 (kg Co2-equivalent per Kwh) for New Zealand [7].

4.2 Results

We ran all experiments for 35,000 simulated hours. Due to space constraints, we have presented the results of only two representative communities: one that experiences net power deficits, and one that has power surpluses. One community (C1) has overall surplus power generation during 35,000 simulated hours, and the other community (C2) has deficit in power generation for the same period.

Experiment 1 In this experiment, first all communities in the simulation employ the fixed strategy (as described in section 2.1) for the baseline scenario. To make comparison of learning strategies with baseline scenario all communities again employ the fixed strategy except communities C1 and C2, which use the two learning strategies (successively, one by one). Table 1 shows the results. The dollar sign ($) represents the net profit/loss and Kg shows the total carbon emission produced. The results clearly show that communities using learning strategies do better as compared to the communities that use only the fixed strategy. There is an improvement of around 10% in net profit loss, and about 78% reduction in the carbon emission by the employing learning strategies. Also the greedy and green learners perform better than fixed agent.

Experiment 2 This experiment is similar to Experiment 1; with the difference being that all the other communities employ the GS strategy (in contrast to fixed strategy). Communities C1 and C2 used the learning strategies (again, one by one). Table 2 shows the results for the two communities using learning strategies in the presence of greedy strategies used by the other communities.

[1] Net Profit/Loss = ((cash in - generation cost) - cash out)

[2] Carbon emission stores the amount of carbon di oxide emitted during electricity production, transmission and distribution.

Table 1. Fixed vs. Learning Strategies

Community	Fixed Strategy		Greedy learner		Green learner	
	$	Kg	$	Kg	$	kg
C1	1,430,788	495,151	1,558,207	137,067	1,531,480	108,986
C2	-1,503,315	870,403	-1,358,975	2,360,820	-1,449,140	185,688

Table 2. Greedy Strategy vs. Learning Strategies

Community	Greedy strategy		Greedy learner		Green learner	
	$	Kg	$	Kg	$	kg
C1	1,429,156	505,229	1,434,999	367,941	1,329,129	80,893
C2	-1,528,295	877,395	-1,503,213	616,847	-1,612,990	127,700

The results clearly show that communities using the learning strategies are better off as compared to communities using non-learning strategies (GS, AS, FS) in terms of net profit loss and carbon emissions.

Experiment 3 This experiment is also similar to Experiment 1 & 2, with the difference being that all the other communities employ the AS strategy. Communities C1 and C2 used the learning strategies (one by one). Table 3 shows the result for the two communities using learning strategies in the presence of altruistic strategy used by the other thirty-nine communities. By employing learning strategies in this setup, community improved up to 29% and carbon emission reduced up to 31%. It clearly shows that learning strategies obtaining their objectives even in the presence of the Altruistic strategy.

Table 3. Altruistic Strategy vs. Learning Strategies

Community	Altruistic strategy		Greedy learner		Green learner	
	$	Kg	$	Kg	$	kg
C1	1,531,710	495,913	1,924,035	714,683	1,500,108	338,360
C2	-1,503,551	879,686	-1,060,448	1,321,194	-1,519,907	685,113

Experiment 4 This experiment consists of three parts: in each part all communities employ one of the strategies described in section 2.1 (i.e. Fixed, Greedy, and Green) except C1 and C2 which use the greedy and green learning strategies. The results are shown in Tables 4, 5, 6, 7, 8, and 9. Tables 4, 6, and 8 show that

both learning strategies perform better when compared to non-learner strategies (tables 5, 7, and 9) even when they learn together in the same environment at the same time.

Table 4. Two Learning strategies in Greedy strategy Background

Community	Greedy strategy	
	$	Kg
C1 (Greedy Learner)	1,422,662	391,738
C2 (Green Learner)	-1,611,041	131,858
C1 (Green Learner)	1,317,871	76,995
C2 (Greedy Learner)	-1,509,336	643,874

Table 5. Non-learner strategies in Greedy strategy Background

Community	Greedy strategy	
	$	Kg
C1 (Greedy Strategy)	1399980	517,146
C2 (Green Strategy)	-1,575,271	642,525
C1 (Green Strategy)	1,350,825	364,804
C2 (Greedy Strategy)	-1,533,322	902,284

Table 6. Two Learning strategies in Altruistic strategy Background

Community	Altruistic strategy	
	$	Kg
C1 (Greedy Learner)	1,912,356	730,615
C2 (Green Learner)	-1,545,476	640,341
C1 (Green Learner)	1,480,572	327,909
C2 (Greedy Learner)	-1,067,943	1,337,009

Table 7. Non-learner strategies in Altruistic strategy Background

Community	Altruistic strategy	
	$	Kg
C1 (Greedy Strategy)	1,838,603	804,407
C2 (Green Strategy)	-1,504,681	873,256
C1 (Green Strategy)	1,522,661	485,900
C2 (Greedy Strategy)	-1,140,483	1,467,335

Table 8. Two Learning strategies in Fixed strategy Background

Community	Fixed strategy	
	$	Kg
C1 (Greedy Learner)	1,599,149	132,305
C2 (Green Learner)	-1,437,841	181,317
C1 (Green Learner)	1,541,371	101,318
C2 (Greedy Learner)	-1,351,420	236,285

Table 9. Non-learner strategies in Fixed strategy Background

Community	Fixed strategy	
	$	Kg
C1 (Greedy Strategy)	1,395,544	474,323
C2 (Green Strategy)	-1,535,957	851,424
C1 (Green Strategy)	1,396,106	474,270
C2 (Greedy Strategy)	-1,540,306	872,806

Experiment 5 In Experiment 5, a community instead of using either the green or greedy learning strategy, uses a hybrid strategy of half-green and half-greedy learning strategies (HGGL) at the same time. A community using this approach splits its surplus or deficit into two chunks. It generates two different trading prices for these chunks and then enters into the market for trading. This way,

a community learns to improve its profit and also learns to reduce its carbon emissions at the same time. Tables 10, 11, and 12 show the results of experiment 5.

Table 10. Half greedy, Half green learner in Greedy strategy Background

Community	Greedy strategy	
	$	Kg
C1 (Green Learner)	1,329,129	80,893
C1 (HGGL)	1,384,942	222,020
C1 (Greedy Learner)	1,434,999	367,941

Community	Greedy strategy	
	$	Kg
C2 (Green Learner)	-1,612,990	127,700
C2 (HGGL)	-1,553,823	400,220
C2 (Greedy Strategy)	-1,503,213	616,847

Table 11. Half greedy, Half green learner in Altruistic strategy Background

Community	Greedy strategy	
	$	Kg
C1 (Green Learner)	1,500,108	338,360
C1 (HGGL)	1,647,174	411,278
C1 (Greedy Learner)	1,924,035	714,683

Community	Greedy strategy	
	$	Kg
C2 (Green Learner)	-1,519,907	685,113
C2 (HGGL)	-1,427,069	753,328
C2 (Greedy Strategy)	-1,060,448	1,321,194

Table 12. Half greedy, Half green learner in Fixed strategy Background

Community	Fixed strategy	
	$	Kg
C1 (Green Learner)	1,531,480	108,986
C1 (HGGL)	1,551,825	113,224
C1 (Greedy Learner)	1,558,207	137,067

Community	Fixed strategy	
	$	Kg
C2 (Green Learner)	-1,449,140	185,688
C2 (HGGL)	-1,398,512	202,703
C2 (Greedy Strategy)	-1,358,975	2,360,820

These show that the community that uses use the half-green and half-greedy learning approach makes a compromised trade-off between the net profit/loss and the carbon emissions. It can be observed from the results that the resultant net profits and carbon emission produced by HGGL is between the Green and Greedy learner strategies.

5 Related Work

Recently there has been increasing interest in the application of multi-agent systems to the power trading, management and control of distributed energy resources on micro-girds or smart grids. For example, Reddy & Veloso [13] presented a learning strategy for the broker agent in the smart-grid market. They used Markov Decision Process (MDPs) and reinforcement learning to train their

broker agent about market tariffs. Peters et al. [10] also proposed a learning approach for the broker agent to learn the bidding price in the smart grid market. Similarly, Rahimi-Kian et al. [12] and Xiong et al. [15] proposed learning mechanisms for the electric suppliers to bid in the electric market. Alam et al. [1] introduced an agent-based model for energy exchanges among communities. In their proposed model, agents use a game-theory approach to form a coalition to exchange power. Fatimah Ishowo-Oloko et al [4] presented a model for a dynamic storage-pricing mechanism that uses storage information from the renewable energy providers to generate daily, real-time electricity prices which are communicated to the customers. In addition Cossentino et al. proposed a multi-agent system for the management of micro-grids [2]. The prime responsibility of their proposed system is to provide an electronic market to the consumers and generators within a micro-grid. In case of any mismatches between supply and demand, their agent-based system will disconnect the loads or feeders depending upon the priority.

With respect to the above-mentioned proposed systems, none of the models considers the environmental concerns during energy trading or exchange. All models are concerned exclusively with profit maximization. In the model of Reddy & Velso [13], the goal of the broker is to maximize profit and avoid the balancing fee imposed by the utility grid in case of mismatch between production and consumption. In this model, the broker is an individual entity who does not represent any community and does not have its own power generating unit. The broker in this situation makes a match between the consumer and producers. Similarly in [10], the authors presented a learning strategy for the broker who makes a match between the producers of the wholesale market and the consumers of in the retailer markets. In two of the models [12] [15], the bidding strategy for the electric suppliers (who trade in the wholesale market) is presented. Buyers haven't been considered. In Alam et al. model [1], they do not consider the profitability of the individual households as well as s the amount of carbon emission mitigated through coalition formation. In [4], there has not been much attention given to the consideration of robust energy distribution across locally-connected communities. Cossentino et al [2] uses a trading mechanism among the internal agents of the micro-grid to balance supply and demand inside the micro-grids. In contrast, our research focuses on learning trading strategies for a community that not only supports inter-micro-grid power trading but also improves the financial benefits and reduce the carbon emissions of the participants in the proportions that the communities choose.

6 Conclusion

Interconnected micro-grids, with renewable energy sources and energy storage devices have already been shown to be effective respect to financial advantage, local autonomy, and more energy distribution [16]. We have presented the two learning trading strategies and have shown that a community can improve its financial benefit and reduce its carbon emission using an agent-based architec-

ture. This has been accomplished by means of a multi-agent simulation that employs synthetic wind data, real electric demand, and current energy pricing data.

Based on the studies conducted, we found that if a community wants to maximize its financial gain then it should adopt the greedy learning strategy; and if a community is environmentally conscious, then it should employ the green strategy to reduce carbon emissions. A community can also maintain a balanced position between these two parameters by using, for example, a half green and half greedy learning strategy.

Agent-based system coordination and collaboration are inherently scalable. So in the future we intend to extend our analysis by conducting more elaborate tests with our agent-based modeling approach. We will then explore the following:

- Dynamic adoption of learning strategies (e.g. a community will choose by itself to become greedy learner, or green learner, or hybrid learner with varying mix of greedy-green strategies, by looking at market conditions).
- Use of reinforcement learning for battery storing strategies.
- Extension of our model by considering power generating units at the consumer level (e.g. solar PVs).
- Reinforcement learning at the demand response management.

References

1. M. Alam, S. Ramchurn, and A. Rogers. Cooperative Energy Exchange for the Efficient Use of Energy and Resources in Remote Communities. In *12th Autonomous Agents and Multiagent Systems (AAMAS) Conference*, pages 731–738, Saint Paul Minnesota, USA, 2013.
2. M. Cossentino, C. Lodato, M. Pucci, and G. Vitale. A Multi-Agent Architecture for Simulating and Managing Microgrids. In *Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 619–622, Szczecin, Poland, 2011.
3. IESO. Independent Electricity System Operator Canada. "`http://www.ieso.ca/`", [Date Accessed: 10-02-2014].
4. F. Ishowo-oloko, P. Vytelingum, N. Jennings, and I. Rahwan. A Storage Pricing Mechanism for Learning Agents in Masdar City Smart Grid. In *11th International Conference on Autonomous Agents and Multiagent Systems*, pages 1167–1168, Valencia, Spain, 2012.
5. M. Z. Jacobson and M. a. Delucchi. Providing all Global Energy with Wind, Water, and Solar Power, Part I: Technologies, Energy Resources, Quantities and Areas of Infrastructure, and Materials. *Energy Policy*, 39(3):1154–1169, Mar. 2011.
6. A. Miller, E. Muljadi, and D. S. Zinger. A Variable Speed Wind Turbine Power Control. *Energy Conversion*, 12(2):181–186, 1997.
7. M. f. t. E. New zealand. Guidance for Voluntary Corporate Greenhouse Gas Reporting Data and Methods for the 2010 Calendar Year. Technical report, Wellington, New Zealand, 2010.
8. J. Nicolaisen, V. Petrov, and L. Tesfatsion. Market Power and Efficiency in a Computational Electricity Market with Discriminatory Double Auction Pricing. *IEEE Transactions on Evolutionary Computation*, 5(5):504–523, 2001.

9. G. Nordic Energy. Electricity Market Group. "http://www.nordicenergy.org/", [Date Accessed: 08-02-2014].
10. M. Peters, W. Ketter, M. Saar-Tsechansky, and J. Collins. A reinforcement learning approach to autonomous decision-making in smart electricity markets. *Machine Learning*, 92(1):5–39, Apr. 2013.
11. H. Pietsch. Property service Division. "http://www.propserv.otago.ac.nz/", [Date Accessed: 1-02-2014].
12. A. Rahimi-kian, B. Sadeghi, S. Member, and R. J. Thomas. Q-Learning Based Supplier-Agents for Electricity Markets. In *Power Engineering Society Summer Meeting, 2002 IEEE*, pages 1–8, Chicago, IL, USA, 2002.
13. P. P. Reddy and M. M. Veloso. Strategy Learning for Autonomous Agents in Smart Grid Markets. 2013.
14. C. J. C. H. Watkins and P. Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, May 1992.
15. G. Xiong, T. Hashiyama, and S. Okuma. An electricity supplier bidding strategy through Q-Learning. In *Power Engineering Society Summer Meeting,*, Chicago, IL, USA, 2002.
16. M. Yasir, M. K. Purvis, M. Purvis, and B. T. R. Savarimuthu. Agent-based community coordination of local energy distribution. *Ai & Society*, Dec. 2013.