# A Multi-Adaptive Agent Model of Generator Bidding in the UK Market in Electricity

**A. J. Bagnall**
School of Information Systems
University of East Anglia
Norwich, NR4 7TJ, England
e-mail : ajb@sys.uea.ac.uk, tel : +44 1603 593794

## Abstract

A model of the UK market in electricity combining key factors influencing generator bidding is proposed and a hierarchical multi-objective adaptive agent architecture using case based reasoning and learning classifier systems is described. Experimentation shows that the adaptive agents learn bidding strategies that have been observed in the real world, and that in some market scenarios the agents appear to be learning the benefits of cooperating to receive increased long term rewards. The potential of the adaptive agent model is illustrated by experimentation with an alternative market structure.

## 1 INTRODUCTION

The goal of this work is to produce a simplified model of the UK market in electricity in which artificial adaptive agents, representing generating companies, compete to generate electricity in a partially observable constrained environment. Agents have been used to model competitors in markets [1] and players in games, particularly the iterated prisoners dilemma [2], using a variety of reinforcement learning techniques [3]. Whilst the agent model presented here is related to this work, there are several key differences concerning both the nature of the market and the restrictions placed on the agents. The UK market is administered by the National Grid Company, UK (NGC). The salient features of the market cycle are as follows: each day, the generating companies submit a bid for each generating unit they own, summarising the price and operating conditions under which they are willing to generate. The NGC collects these bids and produces a generation schedule to meet electricity demand throughout the day, which is split into 48 half hour time slots. The schedule specifies a generation profile for each generating unit, and is formed by attempting to minimize total costs. The NGC also calculates the price that the generators will be paid for generation in each time slot. Further details are given in Section 2, but the key points to note are:

- The market does not operate by matching buyers to sellers. The buyers (the companies that finally supply electricity to the consumer) have no direct influence over the price they have to pay for electricity used by their customers.
- The NGC has a legal obligation to meet demand and to satisfy certain constraints on its network of power lines.
- The method of scheduling and forming a price includes mechanisms to counteract the obvious problems arising from the first two points.
- Some of the information used in the scheduling and price setting are also available to the generating companies.

The last point means that the success or otherwise of a particular agent bid is influenced by external variables in addition to the competing agents' bids, and these variables can be quantified by all agents. Thus the particular payoff achieved for any bid combination will vary, and the agents need to learn strategies to cope with this known variation in environment. The situation where either buyers or sellers are constrained to act in in some way and where an external influence which is both quantifiable and known to all agents affects performance is common to many markets. For example, competing in telecoms markets may require an equal consideration of network constraints and the need to meet demand.

The adaptive agents used in this experiment utilize structures and methods based on learning classifiers

(LCS) [4, 5], case based reasoning (CBR) [6] and simple reinforcement learning algorithms [7]. The structure is complex, and was developed through consideration of the basic requirements of the agents and through experimentation with simpler single adaptive agent models and two adaptive agent models (i.e. models where all but one or two agents follow pure strategies). The issues behind the choice of structure and an overview of the learning mechanisms employed can be found in Section 4, but are covered in more detail in [8, 9].

Ultimately, this type of model could be used to discover previously unknown strategies used by generating companies to increase profits and hence increase the final cost to the electricity consumer. Furthermore, it could provide a tool to experiment with the effect of changes in market structure and regulator influence on patterns of generator behaviour. This paper is concerned with examining the potential usefulness of this type of model as an advisory tool to decision makers in the electricity industry. In Section 5 the following questions are addressed:

- *Can the agents learn bidding strategies that can be identified in real world behaviour (Section 5.1)?* This issue relates to model and agent structure validation. The market has been constructed to allow certain patterns of known behaviour to emerge (such as exploiting the fact that a unit may have to run even if it bids very high), and the goal is to derive discernible strategies that can be related to strategies that generator companies are known to follow.
- *Can the agents learn to cooperate to increase mutual profitability and ultimate cost to the consumer (Section 5.2)?* To assess the learning and action selection mechanisms a behaviour in subset of environments where cooperation can lead to greater profitability is examined.
- *How does altering certain aspects of the market model affect the behaviour of the agents (Section 5.3)?* The potential for possibly useful qualitative advice is illustrated by examining agent behaviour in an alternative market structure.

Finally, a brief discussion of the results and future directions is presented in Section 6.

## 2 THE ENVIRONMENT

Figure 1 shows the daily cycle for the simplified market model. An overview of operation of the real market can be found in [10] and a more detailed description of the market model is given in [9]. Through consultation with NGC and examination of historical bid data, the following market variables were deemed to have the most quantifiable influence on bidding behaviour: *Constraints*, *Demand* and *Capacity Premium*. These variables are used by the market overseer in the tasks of *scheduling* and *price and payment calculation* and hence influence the profitability of a particular bid. The market information is mapped on to a bit

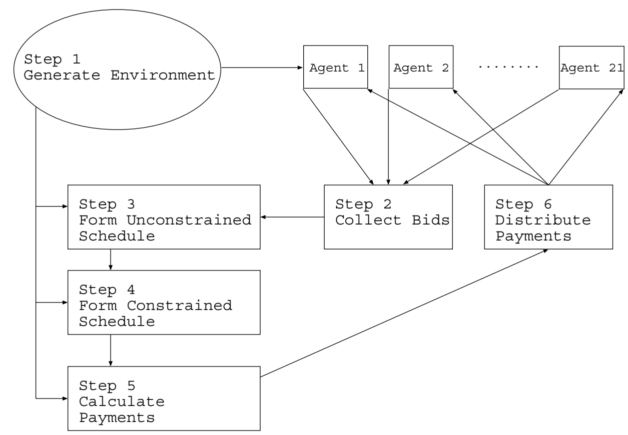


Figure 1: Market structure: Each day an environment is generated probabilistically (with unequal probabilities) and passed to the 21 adaptive agents. A bid is made by each agent, and these are collected and passed to the scheduling unit. This forms the unconstrained schedule using the environmental demand information. The constrained schedule is then formed by taking into account the environmental constraint information. The unconstrained and constrained schedules are passed to the settlement unit which uses them to form SMP. PPP is calculated by adding the capacity premiums for the particular environment. Finally the payment due to each generating unit is calculated, and this plus the schedule information is passed back to the agents. The agents use this information to calculate their profit.

string of 10 bits. An agent's detector maps the constraint information onto the first six bits of the environment message. Bits one and two represent the constrained-on group (00 if no group is constrained-on) bits three and four represent the constrained-off group and bits five and six represent the 4 possible levels of constraint allowed in this model. Forecast demand consists of 48 half hourly demand figures. The Forecast Demand data is highly seasonal and the profile of the demand curve and the level of demand differs considerably between weekday and weekend. We therefore categorize the demand curve as either typical to weekend/weekday winter level (bit 7 of the environmental message), or weekend/weekday summer level (bit 8). We assume that the generators know, or at least are able to estimate, the capacity premiums prior to making a bid. As with demand we need to characterize the graphs to limit the quantity of information passed to the agent, and through observation of the real world data we identified four distinct types of capacity premium curves determined by bits 9-10 of the message. Each agent receives this bitstring and has to choose one of 32 possible actions. An action is an integer between 0 and 31. The effector copies the agent action to the agent bid, a value £0 per MWh to £31 per MWh. Note that the agent has a large action space relative to many reinforcement learning problems. This is to reflect the real world situation, where a single bid consists of up to eight price parameters and a host of technical parameters. For the experiments presented in this paper, the salient points of interest concerning the operation of the market are:

*System Marginal Price* (SMP). The SMP consists of 48 prices in pounds per MWh (a price for each half hour time slot), and is set as the bid of the marginal generator for each period (i.e. the most expensive unit scheduled to generate). Hence if a unit is not the marginal generator but is scheduled to run it still receives the same payment as the marginal unit for power it generates. This system was introduced to encourage units to bid to reflect fundamental generation costs rather than making bids because of market conditions such as high demand. It allows generating companies to behave passively by bidding zero and still receive reasonable payment. in Section 5.3 we examine whether using SMP as opposed to paying generating companies the price they bid has the effect of reducing the agents average bids in the simplified model.

*Constraints*. When a unit is *constrained-on*, it may be required to run and when it is constrained-off it may be forced not to run. If a unit is constrained-on it is paid at its bid price for required generation and if it is *constrained-off* it is paid at *(SMP- bid price)* for the power it would have generated. Units are constrained in groups, thus if a group is constrained-on or off there is a minimum/maximum level of generation required in that group, and the decision on which units to run is based on the bids of agents in the group. Constraints essentially define an alternative game amongst a subset of agents where bids by units outside the group have a lesser effect on the payments than the bids of those in the group. Our interest in the environments focuses on whether the agents can learn to distinguish between these constrained environments and form alternative strategies for acting in these situations.

*Capacity Premiums*. The capacity premium is a function of what is called the *loss of load probability* and is an additional payment added to the SMP to form the *Pool Purchase Price* (PPP). The capacity premium for the simplified model is supposed to encourage units to bid to run (i.e. bid low) in times of high demand. PPP is the amount that unconstrained units are paid for the power they generate.

## 3 TASK FACING THE AGENT

There are 21 agents in the model, and each agent owns a single generating unit. With the environmental information encoded into 10 bits there are 1024 potential distinct environments, but only 732 of these have a non zero probability of occurrence. Essentially, each environment can be considered as a different 21 player non-cooperative non-zero sum game. The task facing each agent is two fold: firstly they have to learn how to play these games to attempt to meet their objectives, and secondly they have to learn how to group games with similar payoffs together in order to form strategies. This second requirement makes the problem much harder, but is introduced to avoid the necessity of having to store a complete environment/action space matrix. The model is a massive simplification of the real world situation, thus maintaining the potential for scalability is important. We also wish to be able to relate bidding strategies back to the real world problem, hence our adoption of a rule based system driven by learning classifiers. A strategy such as 'if demand is very low, bid 20' can easily be expressed in the message syntax as the rule (******00**/20).

We are interested in agents that can handle multiple objectives, as this more accurately reflects the real world. If generating companies were driven solely by the need to maximize profits electricity prices would be much higher. The fact that the companies do not always use their power over the market to fix prices must be due to the need to meet other objectives, such as avoiding regulator pressure and maintaining market share. Thus the agents have two objectives: objective 1 is to discover strategies that will ensure the agent does not make a loss, and objective 2 is to find rules that maximize profit. Note that these objectives are related, but the agent architecture briefly described in Section 4 is easily adaptable to different, more conflicting objectives.

Each agent has three cost parameters which are used to calculate profit from the payment and the generation profile. These are fixed costs, unit generation cost and start up cost. Fixed cost is a daily charge incurred independent of the generation level. Unit generation cost is the cost of generating a single megawatt for an hour (MWh), and start up cost is the cost of restarting the generator after it has been taken off line. These costs are preset to reflect *station type*. We classify each generation unit as one of four types based on the real world.

• *Nuclear units* are expensive to take off line and have little to gain by doing so. They tend to bid zero or close to zero in order to ensure generation. There are 5 nuclear units in our model, which are set with low unit generation costs and high fixed cost and start up cost. If the unit runs all day (i.e. no start up costs) at full capacity the nuclear units need to receive an average payment of £3 per MWh (i.e. PPP averaged over the 48 time slots needs to be 3 or more) to make a profit.

• *Gas units* tend to bid low in order to get in the schedule and infrequently set SMP. Gas units in our model have low generation costs. There are 6 gas units which can make a profit at full generation if average payment is £6 per MWh.

• *Coal units* tend to bid higher than gas stations and set SMP more often. Our coal units have higher generation costs than Gas units. There are 8 coal units in the model which require payment at 8 per MWh for profitability.

• *Oil/Gas turbine units* have higher generation costs but much lower start up costs, hence they tend to bid to be on during peaks and set SMP. To reflect this these units have a high unit generation cost and low fixed and start up costs.

Each agent has a generation capability of 3000 MW except for the oil/GT units which have a capacity of 1000 MW. Hence the total capacity in the market is fixed at 57000 MW, a level which always exceeds demand. On receiving payment from the overseer each agent calculates its profit. It then quantifies its success in meeting its objectives with two reward functions $R_1$ and $R_2$.

The agents cannot calculate the rewards that would have been achieved by alternative bids, hence the learning is unsupervised. The learning problem is further complicated by the fact that profit levels for a particular environment vary considerably dependent on all the agents' actions, and the best achievable profits from environment to environment also show a wide variance (making generalisation difficult).

# 4 AGENT STRUCTURE

The agents follow a hierarchical structure similar to that described in [11]. The *controller* sends the current environment and the previous days rewards to two learning classifier systems, LCS1 and LCS2. Each classifier concentrates on finding rules to meet one of the objectives. LCS1 receives the reward for objective 1 (not make a loss) and LCS2 gets the reward for objective 2 (maximize profit). From each the controller receives back a *prediction array*, the two classifier systems; estimates of the reward that will be received on the current time step. The controller also maintains two *case lists*, the bad cases list (BCR) and the good cases list (BCR). Each of these can store at most 20 environments and a prediction array for these environments. The controller consults the case lists, forms a final prediction array then makes the final action decision.

## 4.1 The Classifier Systems

Both classifiers are similar to XCS [5]. LCS1 has a maximum of 200 rules and LCS2 a maximum of 400. The major difference between LCS2 and XCS is that the genetic algorithm does not operate in the match or action set, rather it acts panmictically, weighting prediction with error. The reason for this is that it is unreasonable to expect the system to maintain a full coverage of environment × action space. Instead, LCS2 is designed to concentrate on certain areas of the environment that have historically been the most profitable.

## 4.2 The Case Lists

The case lists are a long term memory storage facility used by the controller to focus its resources on certain environments that seem, from past experience, to have an important role in meeting the objectives. There are two case lists, the bad case list (BCL) and the good case list (GCL), which are aids for meeting objective 1 and objective 2 respectively. Each case list consists of up to 20 cases. A case consists of a message string to identify the particular environment the case relates to, a prediction array and an array to count the occurrences of each bid. When a particular case occurs the prediction for the action the agent finally chose is updated using the Widrow-Hoff delta rule. Cases are added to the BCL by the controller when an environment that seemingly occurs often yields a loss, and the output from LCS1 has no suitable alternative actions. Cases are added to the GCL when a particularly large profit is achieved. Once the list is full the controller may decide to replace one case with another.

### 4.3 The Controller

The controllers primary tasks are:

- *Formulate an estimate of the expected reward for each action in relation to its competing objectives.* It does this by combining the prediction array from the classifier, and the prediction array of the closest environment on the appropriate case list (by Hamming distance), with the combination being weighted so that the further away the case the less effect it will have on the prediction array.

- *Decide on which objective it is primarily interested in meeting and hence decide on an action.* The controller bases its choice of current primary objective on long term performance in meeting the objectives and the quality of the prediction arrays for the current environment. The need for exploration and exploitation is balanced by using a Boltzmann weighting to form a probability distribution over the action space. The probability of selecting action $a$ is given by

$$P(a) = \frac{e^{\frac{f(a)}{t}}}{\sum_{a' \in A} e^{\frac{f(a')}{t}}}$$

where $t$ is the temperature, $f$ is the normalized final prediction array and $A$ is the action space. Temperature is determined dynamically by the agent by assessing how well it is meeting its current primary objective.

- *Oversee alterations to strategy by exercising some control over the rule discovery processes of the learning modules.* If the controller deems the prediction from either classifier system to be unacceptable it can send a rule creation cover signal to classifiers and receive an altered prediction array. The controller also handles adding cases to the lists and removing cases from the list. When a case is removed rules may be added to the classifier to ensure the information from the outgoing case is not lost.

## 5 RESULTS

The results presented here address the three issues identified in the introduction. In Section 5.1 long term averaged bidding behaviour is examined to see whether the agents can be considered to be evolving towards an accurate simulation. In Section 5.2 a closer look at certain interactions is taken and issues relating to the evolution of cooperation considered and Section 5.3 describes experimentation with an alternative market structure to illustrate the benefits of a simulated market.

### 5.1 Simulation of the Real World Problem

The first question to address is can the agents learn bidding strategies that can be identified in real world behaviour? Specifically, interest lies in whether the agents approach the unconstrained strategies determined by station type described in section 2, and how they perform in constrained environments (i.e. do they learn to bid higher when constrained-on and lower when constrained-off, and do they adapt to differing constraint levels). The first experiment consists of a run of 100000 days. Environments occur with unequal probabilities, with unconstrained environments being more likely than constrained. Figure 2 illustrates the overall strategies adopted by the four station types during unconstrained days (environment matches 0000******). The data series were formed by taking the average bid by station type for each day then forming a 365 day moving average time series. This graph clearly illustrates that nuclear units are bidding at or close to zero, gas and coal units are bidding close to the level required for profitability and that oil/GT units are bidding high in order to capture peak generation, which is broadly equivalent to the behaviour observed in the real world. Further analysis shows that the nuclear units are fairly unresponsive to demand whereas the coal and gas units tend to increase their bids in times of high demand unless the capacity premium (the incentive to bid low) is at maximum, in which case they revert to lower bids. This behaviour is also broadly consistent with real world strategies. Some of the variation displayed in Figure 2 is due to different seasonal demands, but it does seem there is some interaction between the agents of different type in overall strategy. In the initial 10000 unconstrained days the coal and gas agents learn to bid on average at just above the level needed for profit. They then compete in order to remain fully in the schedule. First the average gas bid drops as the gas agents learn to rely on the coal units to set non-peak SMP. The coal units then suffer from not generating at these times and hence become more defensive, dropping their bids to get in schedule. However, the result of this is lower SMP, hence everyone suffers from reduced payments. The gas units respond by raising their bids. Gradually, the coal units seem to discover more stable strategies at around bids of 9 and 10, with occasional dips being met with a compensatory rise by the gas units (or vice versa).

We are also interested in an overview of bidding behaviour evolving in constrained environments. When a group is constrained-on, the generation required can be at 4 levels, and broadly speaking it would be expected that the agents to bid higher when the con-

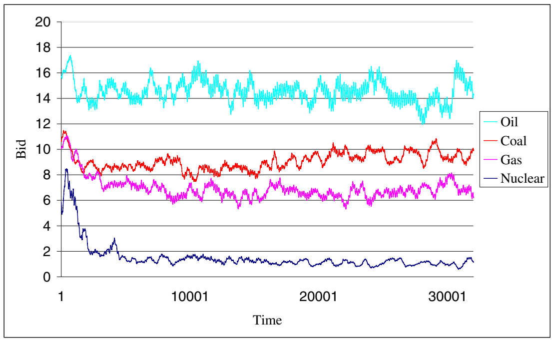


Figure 2: 365 day moving average time series of bids by unit type for unconstrained environments. A data point is the average bid for units of that type for the day in question.

straints are high (when more units are constrained-on). The only exception to this is when the capacity premium is maximum it may be worth bidding low in order to get the bonus payment. Figure 3 illustrates the overall strategies adopted by showing the difference between SMP and the average bid of the constrained units. Figure 3 indicates that the agents

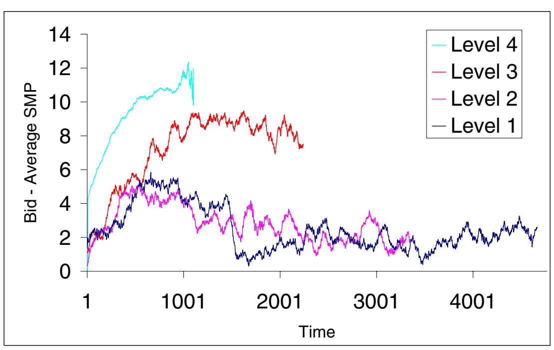


Figure 3: For each constraint level the difference between the average bid of the constrained units and the average SMP for that day was calculated. The data shown is the 100 day moving average of this difference. Constraint levels occur with different probabilities, level 4 being the rarest and level 1 the most common

are learning to bid higher when they are maximally constrained-on (level 4), with a steady increase in the difference between SMP and average bid. When constrained at level 3 the agents progressively learn to bid higher, but tend not to bid higher than 8 or 9 pounds above SMP. With level 1 and 2 constraints the agents seem to initially start cooperating to receive higher rewards, but, after exposure to approximately 1000 environments progress is halted and bidding becomes closer to SMP. The possible reasons for this behaviour are discussed in Section 5.2, but the overall strategies are as expected. This can be confirmed by examining part of the final rule set of one of the agents. Table 1 gives a representative sample of rules matching 01******** for unit 17 (a coal unit; each group consists of a balance of station types). It clearly illustrates that the agent has learnt the value of bidding at maximum when the constraint level is maximum (environments matching 01**11****). There is a good spread of rules covering both level 3 and 4 constraint levels (i.e. matching 01**1****), illustrating the formation of a default hierarchy for these environments. There are fewer rules covering levels 1 and 2 as resources have been concentrated more on the higher constraint levels.

Table 1: Rules for agent 17 matching 01****** at the end of a run of 100000 days

| Condition | Bid | Pred | Condition | Bid | Pred |
|---|---|---|---|---|---|
| 01*11101** | 31 | 374.62 | 01*111**1* | 22 | 309.43 |
| 01**11**1* | 31 | 396.66 | 01**1001** | 22 | 231.76 |
| 011*11**0* | 31 | 384.20 | 01001***0* | 22 | 288.51 |
| 01*011**0* | 29 | 384.32 | 01001*1*** | 22 | 218.39 |
| 01*011**** | 29 | 384.51 | 01001**1** | 22 | 239.69 |
| 01*01101** | 30 | 390.31 | 01001****0 | 22 | 210.07 |
| 01001***0* | 26 | 221.1 | 01*01001** | 22 | 234.24 |
| 01001**1** | 26 | 205.57 | 01001**1*0 | 22 | 238.68 |
| 01001***10 | 26 | 175.78 | 01*01****0 | 22 | 218.22 |
| 01*11101** | 25 | 340.62 | 01**1*1*0* | 22 | 216.31 |
| 01**10**0* | 25 | 210.48 | 01001***0* | 21 | 272.75 |
| 01**10**0* | 25 | 207.88 | 01001***0* | 20 | 230.61 |
| 01**10*10* | 24 | 204.69 | 01111***1* | 17 | 226.08 |
| 01001**1*0 | 23 | 202.02 | *10*10**11 | 6 | 203.41 |
| 01**1000*1 | 22 | 283.85 | *1***0**11 | 7 | 165.19 |
| 01001*1*0* | 22 | 296.81 | *1***0**11 | 11 | 210.63 |

## 5.2 Cooperation in Constraint Bidding

Whilst in terms of the model the results above are encouraging, they do not give much insight into how the agents are actually interacting. The observation that agents are behaving in predictable ways in unconstrained environments essentially means that the agents are not cooperating in order to maximize their mutual payoffs: Simplistically, if every agent bid the maximum bid every day, SMP would be maximum, they would all be partially scheduled and hence make larger profits. More complex forms of cooperation can also increase profits: if a certain proportion of agents bid high the others get an immediate payoff advantage, thus if the agents take turns in making these large bids, they can all increase their long term profit. Such complex cooperation is unlikely to be achieved without some coordination. We are interested in whether the agents have the capacity to learn to cooperate, and whether there is any evidence of this cooperation emerging. To illustrate that the agents can learn the benefit of cooperative strategies, the strategies of agents in a subset of environments where behaviour is easier to analyze and cooperation has a larger benefit is examined.

Consider environments where group 1 is constrained-on at level 2 (environments matching 01**10****). In these environments, units in group 1 are constrained-on in order of price until 13000 MW capacity is reached. The total capacity of the group is 19000 MW, thus each agent wants to bid as high as it can, but not to bid higher than all the other agents since the highest bidding agent will not generate. However, if two agents make the same bid they are both partially constrained-on and thus share rewards. These environments have similarities with the prisoners' dilemma, where cooperation corresponds to two or more agents in the group making the same large bid. Analysis of the bid data used to create Figure 4 shows that of the last 1000 environments matching 01**10****, 221 met this cooperation criteria, although generally at a bid less than the optimal cooperative bid of 31 (20-25 is most common). Although this seems discouraging, further examination reveals that in a further 351 environments from the last 1000 the cooperation criteria would have met (i.e. several agents made the same high bid) except for the bid of a single agent, which attempted a higher bid in the hope of greater profit (a hope that is not realised). It seems the agents are attempting to reach a higher equilibrium cooperative bid by alternatively attempting higher bids, but the dynamics of rule creation and the limited size of the rule set stop the agents learning to fully exploit the higher potential cooperative bids. (Higher bids have an associated higher risk in non constrained environments which the rule may mistakenly also cover, hence the rules suggesting high actions will, on average, have higher deletion probability). However, this very restriction may well aid the emergence of the cooperative equilibrium at a lower level. The fact that the agent cannot store a complete environment/action space representation means the agent will generally only pick from a subset of the possible actions. Table 1 illustrates that there are a preponderance of rules with action 22, a common cooperative bid. This in itself does not make the action more likely to be selected, but the high associated fitness and prediction mean the action will maintain a strong presence thanks to the operation of the genetic algorithm.

### 5.3 The Effects of Alterations of the Environment

The final experiment is designed to test the effect of changing market conditions on agent behaviour. The aspect of the market looked at is the payment calculation. SMP is the price of the last unit loaded onto the unconstrained schedule, and this price is paid to all units in the schedule. We are interested in seeing how altering this system so that agents are actually paid at their bid price rather than SMP effects the system. The main question arising is: does the removal of the safety net option of bidding low to accept SMP make the agents likely to learn cooperative behaviour patterns which result in higher average bids. A further run of 100000 days was conducted with the new payment calculation method for unconstrained generation. Figure 4 shows the bids by station type. It is apparent that each type of agent is, on average, bidding higher than in the previous experiment (illustrated to figure 2). This experiment is interesting because it lends anecdotal evidence to the decision to use SMP to calculate payments. In terms of the model, the ability to avoid complex game playing means that the agents have less incentive to learn cooperative strategies

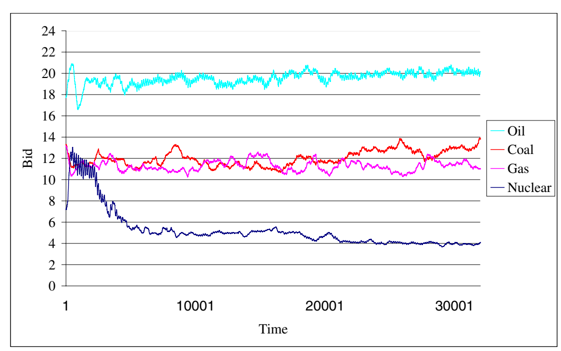


Figure 4: 365 day moving average time series of bids by unit type for unconstrained environments when agents are paid at their bid price rather than SMP

## 6 CONCLUSIONS AND FUTURE DIRECTIONS

In order to convince people of the long term potential of an evolutionary approach to economic modelling it is important to be able to define a system where the agents behave in ways the are comprehensible in terms of the real world scenario. The model described in this paper contains aspects of the actual market known to be important in generator bidding, and agents acting in the simulated environment have been shown to evolve strategies analogous to real world bidding strategies. The agents, with relatively limited resources, learn to group similar environments together through the use of classifier system rules while maintaining a selective focusing on certain influential environments through the case lists. The hierarchical structure allows the balancing of objectives and by forming a default defensive rule set with LCS1 the agents can further concentrate their memory on grouping and modelling the more profitable environments. Certain aspects of cooperation have been seen to emerge, although the large number of available actions, the exploration/exploitation policy and the potential incorrect generalisation over environments make it difficult to maintain long term mutually beneficial strategies. The possible future application of this type of model as an experimental advisory tool was illustrated by showing the effect of altering the market structure and observing behaviour, although obviously the application is far from being of real practical use. However, the agent architecture was designed in such a way as to allow for the potential scaling up of the environment size and for the inclusion of more realistic agent objectives. The model could be made more realistic in many ways. These include allowing a dynamic alteration of generation capacity by allowing agents to stop trading and new agents to enter the market, giving the agents alternative objectives such as maintaining market share, letting one agent control more than one generation and more accurately modelling the demand curve. All these features would allow the potential for more complex and subtle strategies to emerge. It is probably of more interest initially to examine the potential for the evolution of cooperation and the effect of altering the distribution of game types on this potential cooperation. In terms of iterated games with unknown reward functions, the model introduces a realistic feature rarely considered when looking at evolutionary agents interacting: namely, the fact that interactions in a game environment are rarely under identical circumstances, i.e. with identical reward functions. Agents may play games against each other where known factors, variables other than the other agents' bids, effect the reward matrix. The use of classifier rules and the hybridization with case based reasoning give the agents the potential to transfer knowledge from one game to a potentially similar, but less tried, game.

**Acknowledgements**

I would like to thank the National Grid Company for their support and advice on this EPSRC CASE sponsored project.

## References

[1] R.G. Palmer, W. Brian Arthur, J. H. Holland, B. LeBaron, and P. Tayler. Artificial economic life: A simple model of a stockmarket. *Physica D*, 75:264–274, 1994.

[2] R. Axelrod. *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration.* Princetown University Press, 1997.

[3] T. W. Sandholm and R. H. Crites. Multiagent reinforcement learning in the iterated prisoner's dilemma. *BioSystems*, 37:147–166, 1996.

[4] J. H. Holland. Genetic algorithms and classifier systems: Foundations and future directions. In J. J. Grefenstette, editor, *Proceedings of the Second International Conference on Genetic Algorithms.* Lawrence Earlbaum Associates, 1987.

[5] S. W. Wilson. Classifier fitness based on accuracy. *Evolutionary Computation*, 3(2), 1995.

[6] I. Watson. *Applying Case Based Reasoning.* Morgan Kaufmann, 1997.

[7] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction.* MIT Press, 1998.

[8] A. J. Bagnall and G. D. Smith. Using an adaptive agent to bid in a simplified model of the uk market in electricity. In *Proceedings of Genetic and Evolutionary Computation Conference*, 1999.

[9] A. J. Bagnall and G. D. Smith. An adaptive agent model for generator company bidding in the uk power pool. In *Proceedings of Evolution Artificielle*, 1999.

[10] W. Fairney. Power generation in the 1990s. *Power Engineering Journal*, pages 239–246, December 1993.

[11] M. Dorigo and U. Schnepf. Genetics-based machine learning and behaviour based robotics: a new synthesis. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-23(1), 1993.