University of Strathclyde

Department of Electronic and Electrical Engineering

Learning to Trade Power

by

Richard W. Lincoln

A thesis presented in fulfilment of the requirements for the degree of

Doctor of Philosophy

2010

Signed: Date: August 10, 2010

Acknowledgements

I wish to thank Prof. Jim McDonald for giving me the opportunity for postgraduate study and for permitting me the freedom to pursue my own research interests. I am also very grateful to my supervisors, Prof. Graeme Burt and Dr Stuart Galloway, for their guidance and scholarship. I wish to offer special thanks to my parents, my brother and my sister for their support throughout my PhD.

This research made extensive use of software projects by researchers from other institutions. I wish to thank Dr Ray Zimmerman from Cornell University for his work on the optimal power flow formulations in MATPOWER, the researchers from the Dalle Molle Institute for Artificial Intelligence (IDSIA) and the Technical University of Munich involved in the PyBrain project for their work on reinforcement learning algorithms and artificial neural networks and Charles Gieseler from Iowa State University for the Java Reinforcement Learning Module (JReLM) from which an implementation of the Roth-Erev learning method was translated.

This research was funded by the United Kingdom Engineering and Physical Sciences Research Council through the Supergen Highly Distributed Power Systems consortium under grant GR/T28836/01.

Abstract

In electric power Engineering, learning algorithms can be used to model the strategies of electricity market participants. The objective of this work is to establish if *policy gradient* reinforcement learning methods can provide superior models of market participants than previously applied *value function based* methods.

Supply of electricity involves technology, money, people, natural resources and the environment. All of these aspects are constantly changing and electricity markets must be suitably researched to ensure that they are fit for purpose. In this thesis electricity markets are modelled using optimal power flow: a non-linear optimisation problem solved using primal-dual interior point methods. Policy gradient reinforcement learning algorithms are used to adjust the parameters of multi-layer perceptrons that approximate market participant's policies for selecting quantities and prices bid into the marketplace.

Traditional reinforcement learning methods that learn a value function have been previously applied in simulated electricity trade, but are restricted to discrete representations of the market environment. Policy gradient methods have been proven to offer convergence guarantees in continuous environments, such as in robotic control applications, and avoid many of the problems that mar value function based methods.

The benefits of using policy gradient methods in electricity market simulation are explored and the results demonstrate their superior trading ability when operating in large constrained networks. By advancing the use of learning methods in electricity market simulation, this work provides the opportunity to revisit previous research in the field and creates the possibility for policy gradient methods to be used in decision support and automated energy trading applications.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis examines learning algorithms in the domain of electric power trade. The present chapter explains the motivation for research into electricity trade. It defines the problem under consideration, states the principle research contributions made and outlines the remainder of the thesis.

1.1 Research Motivation

The average total demand for electricity in the United Kingdom (UK) is approximately 45GW and the cost of buying 1MW for one hour is around £40 (Department of Energy and Climate Change, 2009). This equates to yearly transaction values of £16 billion. The value of electricity to society is especially apparent when supply fails. The New York black-out in August 2003 involved a loss of 61.8GW of power supply to approximately 50 million consumers. The majority of supplies were restored within two days, but the event is estimated to have cost more than $6 billion (Minkel, 2008; ICF Consulting, 2003).

Quality of life for a person has been shown to be directly proportional to that person's electricity usage (Alam, Bala, Huo, & Matin, 1991). The world population is currently 6.7 billion and forecast to exceed 9 billion by the year 2050 (United Nations, 2003). Electricity production currently demands over 1/3 of the annual primary energy extracted. As people endeavour to improve their quality of life, finite primary energy fuel resources are becoming increasingly scarce and market mechanisms (e.g. auctions), where the final allocation is based upon the claimants' willingness to pay for the goods, provide a device for efficient allocation of resources in short supply.

Commercialisation of large electricity supply industries began two decades ago in the UK. The inability to store electricity, once generated, in a commercially viable quantity prevents trade as a conventional commodity. Trading mechanisms must allow shortfalls in electric energy to be purchased at short notice from quickly dispatchable generators. Market structures that facilitate this have been implemented in countries and states around the world. Designed correctly, a competitive electricity market promotes efficiency and drives down costs to the consumer, while design errors can lead to market power abuse and elevated market prices.

The value of electricity to society means that it is not feasible to experiment with radical changes to trading arrangements on real systems. A practical alternative is to study abstract mathematical models (with sets of simplifying approximations and assumptions) and find analytical solutions, where possible, using computer programs. Competition is fundamental to markets, but accurate participant strategy models are difficult to obtain. Reinforcement learning methods can be used to represent adaptive behaviour in competing players and have been shown to be capable of learning complex strategies (Tesauro, 1994).

1.2 Problem Statement

Individuals participating in an electricity market (be they representing a generating company, load serving entity or a firm of traders) utilise multi-dimensional data, mostly continuous in nature. Certain data, such as demand forecasts, exhibits a degree of uncertainty and other market information, such as competitor's bids, is hidden.

Traditional reinforcement learning methods accociate a value with each available action in a given state. When used with look-up tables they are restricted by Bellman's Curse of Dimensionality (Bellman, 1961) and can not be applied to complex problems with high-dimensional state and actions space. When used with function approximation techniques (e.g. artificial neural networks) to allow operation in continuous environments, these methods have been shown to have poor convergence properties, even in simple problems (Gordon, 1995; Baird, 1995; Tsitsiklis & Roy, 1994).

Policy gradient reinforcement learning methods do not attempt to approximate a value function, but use function approximation techiniques to represent a policy for seleting actions and search directly in the space of its parameters. They do not suffer from many of the problems that mar value-function based methods in high-dimensional domains. They have strong convergence properties, do not require that all states be continuously visited and work with state and action spaces that are continuous, discrete or mixed. Policy performance may be degraded by uncertainty in state data, but the learning methods need not be altered. Policy gradient methods have been successfully applied in many operational settings (Sutton, McAllester, Singh, & Mansour, 2000; Peters & Schaal, 2006; Moody & Saffell, 2001; Peshkin & Savova, 2002).

It is proposed in this thesis that agents which learn using policy gradient methods may outperform those using value function based methods in simulated competitive electricity trade. It is further proposed that policy gradient methods may operate better under dynamic electric power system conditions, achieving greater profit by exploiting constraints to their financial benefit. This thesis will use electricity market simulation techniques to compare value function based and policy gradient learning methods and explore these proposals.

1.3 Research Contributions

The research presented in this thesis pertains to the academic fields of Electric Power Engineering, Artificial Intelligence and Economics. The principle contributions made by this thesis in these fields are:

- The first application of policy gradient reinforcement learning methods in simulated energy trade.
- The first application of a non-linear optimal power flow formulation in agent based electricity market simulation.
- A new stateful formulation of the Roth-Erev reinforcement learning method.
- Results that show how policy gradient reinforcement learning methods converge more slowly than value function based methods when learning simple power trade policies.
- Results that show how agents using policy gradient reinforcement learning methods achieve greater profitability than those using value function methods when when competing to supply electric power on equal terms.
- An implementation of a multi-agent system for electricity market simulation with discrete and continuous sensor and action space representations.
- The idea of applying Neuro Fitted Q Iteration and GQ(λ) in simulations of competitive energy trade.

- A formulation of optimal power flow as a reinforcement learning problem.
- Capital investment planning formulated as a reinforcement learning problem.
- The idea of using data from the National Grid seven year statement to simulate the UK electricity market.

1.4 Thesis Outline

The presentation of this thesis is organised into 9 chapters. Chapter 2 provides background information on electricity supply, wholesale electricity markets and reinforcement learning. It describes how optimal power flow formulations can be used to model electricity markets and defines the reinforcement learning methods under examination.

In Chapter 3 the research in this thesis is described in the context of previous work, related in terms of application field and methodology. Publications on agent based electricity market simulation are reviewed with emphasis on the reinforcement learning methods used. Previous applications of policy gradient learning methods in other market settings are also discussed.

Chapter 4 describes the electric power exchange auction market model and the multi-agent system used to simulated electricity trade. It defines association of learning agents with a portfolio of generators and/or dispatchable loads, the submission of offers and/or bids and the calculation of rewards. The chapter also shows how artificial neural networks, used for policy function approximation, are structured to operate with environment state information and the design of look-up tables used with value function based methods.

An initial experiment that aims to establish the relative learning capabilities of the different methods is provided in Chapter 5. Data for the six bus test case is given and results are presented using sparkline diagrams.

Chapter 6 examines the ability of agents to exploit constraints in the power system model using their learning algorithms. A larger 30 bus test case is defined and the choice of generator cost data is explained. By tightening constraints, imposing line outages and profiling demand, the learning methods compete under dynamic conditions.

In Chapter 7.1 ideas for futher work are outlined. Alternative reinforcement learning methods that show promise for application in a similar study are listed. An optimal power flow formulation as a reinforcement learning problem is defined.

A model of the UK transmission system, constructed from data in the National Grid seven year statement, is described and it is explained how the model could be used, along with the advances made in this thesis, to study issues pertinent to the nation. The chapter goes on to explain how the option to use AC optimal power flow, provides opportunities for more accurate and revealing research. The need to model interelationships between multiple markets, including gas and emissions, is recognised and a method for achieving this using the existing model is described.

The primary conclusions of this thesis are summarised in Chapter 7.2. Shortcommings are noted and the broaded implications are addressed.

Chapter 2

Background

This chapter provides background information on the electricity market and electric power system simulation. A brief introduction to national electricity supply is given along with a history of UK wholesale electricity market designs. Approaches to market simulation that account for transmission system constraints are introduced and definitions of the learning algorithms later used to model market participant behaviour are provided.

2.1 Electric Power Supply

Generation and bulk movement of electricity in the UK takes place in a three-phase alternating current (AC) power system. The *phases* are high voltage, sinusoidal electrical waveforms, offset in time from each other by 120 degrees and oscillating at a frequency of approximately 50Hz. Synchronous generators (sometimes known as Alternators), typically rotating at 3000 or 1500 revolutions per minute, generate apparent power S at a line voltage V_l typically between 11kV and 25kV. One of the principal reasons that AC, and not direct current (DC), systems are common in electricity supply is that they allow power to be transformed between voltages with very high efficiency. The output from a power station is typically stepped-up to 275kV or 400kV for transmission around the country. The apparent power conducted by a three-phase transmission line l is the product of the line current I_l and the line voltage

$$S = \sqrt{3} V_l I_l. \tag{2.1}$$

Therefore the line current is inversely proportional to the voltage at which the power is transmitted. Ohmic heating losses are directly proportional to the square of the line current

$$P_r = 3I_l^2R \tag{2.2}$$

where R is the resistance of the transmission line. Hence, any reduction in line current dramatically reduces the amount of energy wasted through heating losses. One consequence of high voltages is the larger extent and integrity of the insulation required between conductors, neutral and earth. This is the reason that transmission towers are large and undergrounding systems is expensive.

The UK transmission system operates at 400kV and 275kV (and 132kV in Scotland), but systems with voltages upto and beyond 1000kV are used in larger countries such as Canada and China. For transmission over very long distances or undersea, high voltage DC (HVDC) systems have become economically viable in recent years. The reactance of a line is proportional to frequency, therefore the reactive power component in an HVDC system is nil and more active power flow can be transmitted in a line/cable of a certain diameter.

The ability to transform power between voltages and transmit large volumes over long distances allows electricity generation to take place at high capacity power stations, which offer economies of scale and lower operating costs. It allows electricity to be transmitted across country borders and from renewable energy plant such as hydro power stations located in remote areas. Figure **??** shows the UK's existing HVDC interconnectors and how larger power stations are located away from load centres and close to sources of fuel, such as the coal fields in northern England and gas supply terminals near Cardiff and London.

For delivery to most consumers, electric energy is transferred at a substation from the transmission system to the grid supply point of a distribution system. Distribution networks are also three-phase AC power systems, but typically operate at lower voltages and differ in their general structure (or topology) from transmission networks. Transmission networks are typically highly interconnected, providing multiple paths for power flow. Whereas distribution networks, in rural areas, typically consist of long radial feeders (usually overhead lines) or, in urban areas, consist of many ring circuits (usually cables). Three-phase transformers, that step the voltage down to levels more convenient for general use (typically from 11kV or 33kV to 400V), are spaced along the feeders/rings. All three-phases at 400V may be provided for industrial and commercial loads or individual phases at 230V supply typical domestic and other commercial loads. Splitting of phases is usually planned so that each is loaded equally. This produces a balanced, symmetrical system that may be analysed, as explained in Section 2.3.2, as a *single* phase circuit. Figure **??** illustrates the basic structure of a typical national electric power system (U.S.-Canada Power System Outage Task Force, 2004).

2.2 Electricity Markets

The UK was the first large country to privatise its electricity supply industry when it did so in the early 1990s. The approach has been used as a model by other countries and the market structures that have since been implemented in the UK have utilised some of the main concepts in national electricity market design.

The England and Wales Electricity Pool was created in 1990 to break up the vertically integrated Central Electricity Generating Board (CEGB) and to gradually introduce competition in generation and retail supply. Early adoption of electricity markets by the UK has lead to the country hosting many of the main European power and gas exchanges and the UK boasts a high degree of consumer switching compared to other European countries. The Pool has since been replaced by trading arrangements in which market outcomes are not centrally determined, but arise largely from bilateral agreements between producers and suppliers.

2.2.1 The England and Wales Electricity Pool

The Electric Lighting Act 1882 initiated the development of the UK's electricity supply industry by permitting persons, companies and local authorities to set up supply systems, principally at the time for the purposes of street lighting and trams. The Central Electricity Board started operating the first grid of interconnected regional networks (synchronised at 132kV, 50Hz) in 1933. This began operation as a national system five years later and was nationalised in 1947. Over 600 electricity companies were merged in the process and the British Electricity Authority was created. It was later dissolved and replaced with the CEGB and the Electricity Council under The Electricity Act 1957. The CEGB was responsible for planning the network and generating sufficient electricity until the beginning of privatisation.

The UK electricity supply industry was privatised, and The England and Wales Electricity Pool created, in March 1990. Control of the transmission system was transferred from the CEGB to the National Grid Company, which was originally owned by twelve regional electricity companies and has since become publicly listed. The Pool was a multilateral contractual arrangement between generators and suppliers and did not itself buy or sell electricity. Competition in generation was introduced gradually, by first entitling customers with consumption greater than or equal to 1MW (approximately 45% of the non-domestic market (Department of Energy and Climate Change, 2009)) to purchase electricity form any listed supplier. This limit was lowered in April 1994 to included customers with peak loads of 100kW or more. Finally, between September 1998 and March 1999 the market was opened to all customers.

Scheduling of generation was on a merit order basis (cheapest first) at a day ahead stage and set a wholesale electricity price for each half-hour period of the schedule day. Forecasts of total demand in MW, based on historic data and adjusted for factors such as the weather, for each settlement period were used by generating companies and organisations with interconnects to the England and Wales grid to formulate bids that had to be submitted to the grid operator by 10AM on the day before the schedule day.

Figure **??** illustrates four of the five price parameters that made up a bid. A start-up price would also be stated, representing the cost of turning on the generator from cold. The no-load price c_{noload} represents the cost in pounds of keeping the generator running regardless of output. Three incremental prices c_1, c_2 and c_3 specify the cost in £/MWh of generation between set-points p_1, p_2 and p_3.

A settlement algorithm would determine an unconstrained schedule (with no account being taken for the physical limitations of the transmission system), meeting the forecast demand and requirements for reserve while minimising cost. Cheapest bids up to the marginal point would be accepted first and the bid price from the marginal generator would generally determine the system marginal price for each settlement period. The system marginal price would form the basis of the prices paid by consumers and paid to generators, which would be adjusted such that that the costs of transmission are covered by the market and that the availability of capacity is encouraged at certain times.

Variations in demand and changes in plant availability were adjusted for by the grid operator between day close and physical delivery, producing a constrained schedule. Generators having submitted bids would be instructed to increase or reduce production appropriately. Alternatively, the grid operator could instruct large customers with contracts to curtail their demand to do so or instruct generators contracted to provide ancillary services to adjust production.

2.2.2 British Electricity Transmission and Trading Arrangements

Concerns over exploitation of market power in The England and Wales Electricity Pool and its effectiveness in reducing consumer electricity prices prompted the introduction of New Electricity Trading Arrangements (NETA) in March 2001 (D. Bunn & Martoccia, 2005). The aim was to improve efficiency and provide greater choice to participants. Control of the Scottish transmission system was included with the introduction of the nationwide British Electricity Transmission and Trading Arrangements (BETTA) in April 2005 under The Energy Act 2004. While The Pool operated a single daily auction and dispatched plant centrally, under the new arrangements participants became self-dispatching and market positions became determined through continuous bilateral trading between generators, suppliers, traders and consumers.

The majority of power is traded under the BETTA through long-term contracts that are customised to the requirements of each party (Kirschen & Strbac, 2004). These instruments suit participants responsible for large power plant or those purchasing large volumes of power for many customers. Sizeable amounts of time and effort are required for these long-term contracts to be formed and this results in a high associated transaction cost. However, they reduce risk for large players and often include a degree of flexibility.

Electric power is also traded directly between participants through over-the-counter contracts that are usually of a standardised form. Such contracts typically concern smaller volumes of power and have much lower associated transaction costs. Often they are used by participants to refine their market position ahead of delivery time (Kirschen & Strbac, 2004).

Trading facilities, such as power exchanges, provide a means for participants to fine-tune their positions further, through short-term transactions for often relatively small quantities of energy. Modern exchanges are computerised and accept anonymous offers and bids submitted electronically. A submitted offer/bid will be paired with any outstanding bids/offers in the system with compatible price and quantity values. The details are then displayed for traders to observe and to use in subsequent trading.

All bilateral trading must be completed before “gate-closure”, a point in time before delivery that gives the system operator an opportunity to balance supply and demand and mitigate potential breaches of system limits. In keeping with the UK’s free market philosophy, a competitive spot market (Schweppe, Caramanis, Tabors, & Bohn, 1988) is used in the balancing process. A generator that is not fully loaded may offer a price at which it is willing to increase its output by a specified quantity, stating the rate at which it is capable of doing so. Certain loads may also offer demand reductions at a price which can typically be implemented very quickly. Longer-term contracts for balancing services are also struck between the system operator and generators/suppliers in order to avoid the price volatility often associated with spot markets.

2.3 Electricity Market Simulation

Previous sections have identified the importance of electricity to modern societies and explained how the majority of electricity supply in the UK is trusted to unadministered bilateral trade. Electricity supply involves technology, money, people, natural resources and the environment. All of these aspects are constantly changing and electricity markets must be suitably researched to ensure that they are fit for purpose. The value of electricity to society means that it is not feasible to experiment with radical changes to trading arrangements on real systems. A practical alternative is to study abstract mathematical models (with sets of simplifying approximations and assumptions) and find analytical solutions, where possible, by simulating them using computer programs.

Game theory is the branch of applied mathematics in which behaviour in strategic situations is captured mathematically. A common approach to doing this is to model the system and players as a mathematical optimisation problem. Optimal Power Flow is a classical optimisation problem in the field of electric power engineering and variants of it are widely used in the research electricity markets. In this thesis, optimal power flow is used in an *agent-based* simulation, an alternative approach to the mathematics of games.

2.3.1 Agent-Based Simulation

Social systems, such as electricity markets, are inherently complex and involve interactions between different types of individuals and between individuals and collective entities, such as organisations or groups, the behaviour of which is itself the product of individual interactions. This complexity drives classical monolithic equilibrium models to their limits. The models are often highly stylised and limited to small numbers of players with strong constraining assumptions made on their behaviour.

Agent-based simulation involves modelling the simultaneous operations of and interactions between adaptive agent then assessing their effect on the system as a whole. Macro-level system properties arise from agent interactions, even those with simple behavioural rules, that could not be deduced by simply aggregating the agent's properties.

Following Tesfatsion and Judd (2006), the objectives of agent-based modelling research fall roughly into four strands: empirical, normative, heuristic and methodological. The *empirical* objectives are to understand how and why macro-level regularities have evolved from micro-level interactions when little or no top-down control is present. Research with *normative* goals aims to relate agent-based models to an ideal standard or optimal design. The objective being to evaluate proposed designs for social policy, institutions or processes in their ability to produce socially desirable system performance. The *heuristic* strand aims to generate theories on the fundamental causal mechanisms in social systems that can be observed, even in simple systems, when there are alternative initial conditions. This thesis aims to provide *methodological* advancement. Improvements in the tools and methods available aids research with the former objectives.

2.3.2 Optimal Power Flow

Nationalised electricity supply industries were for many years planned, operated and controlled centrally. A system operator would determine which generators must operate and the required output of the operating units such that demand and reserve requirements were met and the overall cost of production was minimised. In Electric Power Engineering, this is termed the *unit commitment* and *economic dispatch* problem.

In 1962 a unit commitment formulation was published that incorporated electric power system constraints (Carpentier, 1962). Optimal power flow is this combination of the economic and the power flow aspects of power systems into a mathematical optimisation problem. The ability to use optimal power flow to solve centralised power system operation problems and determine prices in power pool markets has led to it being one of the most widely studied subjects in the electric power systems community.

Power Flow Formulation

Optimal power flow derives its name from the power flow (or load flow) steady-state power system analysis technique. Given sets of generator data, load data and a nodal admittance matrix, a power flow study determines the complex voltage

$$V_i = |V_i|\angle\delta_i = |V_i|(\cos\delta_i + j\sin\delta_i) \tag{2.3}$$

at each node i in the power system, from which branch flows may be calculated (Grainger & Stevenson, 1994).

Nodal Admittance Matrix The nodal admittance matrix describes the electrical network and its formulation is dependant upon the transmission line, transformer and shunt models employed. A branch in a nodal representation of a power system is typically modelled as a medium length transmission line in series with a regulating transformer at the "from" end (Zimmerman, 2010, p.11). A nominal-π model with total series admittance $y_s = 1/(r_s + jx_s)$ and total shunt capacitance b_c is used to represent the transmission line. The transformer is assumed to be ideal, phase-shifting and tap-changing, with the ratio between primary winding voltage v_f and secondary winding voltage $N = \tau e^{j\theta_{ph}}$ where τ is the tap ratio and θ_{ph} is the phase shift angle. Figure **??** diagrams this conventional branch model. From Kirchhoff's Current Law the current in the series impedance is

$$i_s = \frac{b_c}{2}v_t - i_t \tag{2.4}$$

and from Kirchhoff's Voltage Law the voltage across the secondary winding of the transformer is

$$\frac{v_f}{N} = v_t + \frac{i_s}{y_s} \tag{2.5}$$

Substituting i_s from equation (2.4), gives

$$\frac{v_f}{N} = v_t - \frac{i_t}{y_s} + v_t\frac{b_c}{2y_s} \tag{2.6}$$

and rearranging in terms if i_t, gives

$$i_t = v_s\left(\frac{-y_s}{\tau e^{\theta_{ph}}}\right) + v_r\left(y_s + \frac{b_c}{2}\right) \tag{2.7}$$

The current through the secondary winding of the transformer is

$$N^* i_f = i_s + \frac{b_c}{2}\frac{v_f}{N} \tag{2.8}$$

Substituting i_s from equation (2.4) again, gives

$$N^* i_f = \frac{b_c}{2}v_t - i_t + \frac{b_c}{2}\frac{v_f}{N} \tag{2.9}$$

and substituting $\frac{v_f}{N}$ from equation (2.6) and rearranging in terms if i_s, gives

$$i_s = v_s \left(\frac{1}{\tau^2} \left(y_s + \frac{b_c}{2} \right) \right) + v_r \left(\frac{y_s}{\tau e^{-j\theta}} \right) \tag{2.10}$$

Combining equations (2.7) and (2.10), the *from* and *to* end complex current injections for branch l are

$$\begin{bmatrix} i_f^l \\ i_t^l \end{bmatrix} = \begin{bmatrix} y_{ff}^l & y_{ft}^l \\ y_{tf}^l & y_{tt}^l \end{bmatrix} \begin{bmatrix} v_f^l \\ v_t^l \end{bmatrix} \tag{2.11}$$

where

$$y_{ff}^l = \frac{1}{\tau^2} \left(y_s + \frac{b_c}{2} \right) \tag{2.12}$$

$$y_{ft}^l = \frac{y_s}{\tau e^{-j\theta_{ph}}} \tag{2.13}$$

$$y_{tf}^l = \frac{-y_s}{\tau e^{j\theta_{ph}}} \tag{2.14}$$

$$y_{tt}^l = y_s + \frac{b_c}{2} \tag{2.15}$$

Let Y_{ff}, Y_{ft}, Y_{tf} and Y_{tt} be $n_l \times 1$ vectors where the l^{th} element of each corresponds to y_{ff}^l, y_{ft}^l, y_{tf}^l and y_{tt}^l, respectively. Furthermore, let C_f and C_t be the $n_l \times n_b$ branch-bus connection matrices, where $C_{fij} = 1$ and $C_{tik} = 1$ if branch i connects from bus j to bus k. The $n_l \times n_b$ branch admittance matrices are

$$Y_f = \mathbf{diag}(Y_{ff})C_f + \mathbf{diag}(Y_{ft})C_t \tag{2.16}$$

$$Y_t = \mathbf{diag}(Y_{tf})C_f + \mathbf{diag}(Y_{tt})C_t \tag{2.17}$$

and the $n_b \times n_b$ nodal admittance matrix is

$$Y_{bus} = C_f^\mathsf{T} Y_f + C_t^\mathsf{T} Y_t. \tag{2.18}$$

Power Balance For a network of n_b nodes, the current injected at node i is

$$I_i = \sum_{j=1}^{n_b} Y_{ij} V_j \tag{2.19}$$

where $Y_{ij} = |Y_{ij}|\angle\theta_{ij}$ is the $(i,j)^{th}$ element if the Y_{bus} matrix. Hence, the apparent power entering the network at bus i is

$$S_i = P_i + Q_i = V_i I_i^* = \sum_{n=1}^{n_b} |Y_{ij} V_i V_j| \angle(\delta_i - \delta_j - \theta_{ij}) \tag{2.20}$$

Converting to polar coordinates and separating the real and imaginary parts, the active power

$$P_i = \sum_{n=1}^{n_b} |Y_{ij} V_i V_j| \cos(\delta_i - \delta_j - \theta_{ij}) \tag{2.21}$$

and the reactive power

$$Q_i = \sum_{n=1}^{n_b} |Y_{ij} V_i V_j| \sin(\delta_i - \delta_j - \theta_{ij}) \tag{2.22}$$

entering the network at bus i are non-linear functions of V_i, as indicated by the presence of the sine and cosine terms. Kirchoff's Current Law requires that the net complex power injection (generation - load) at each bus equals the sum of complex power flows on each branch connected to the bus. The power balance equations

$$P_g^i - P_d^i = P^i \tag{2.23}$$

and

$$Q_g^i - Q_d^i = Q^i, \tag{2.24}$$

where the subscripts g and d indicate generation and demand respectively, form the principal non-linear constraints in the optimal power flow problem.

Optimal Power Flow Formulation

Optimal power flow is a mathematical optimisation problem constrained by the complex power balance equations (2.23) and (2.24). Mathematical optimisation problems have the general form

$$\min_x f(x) \tag{2.25}$$

subject to

$$g(x) = 0 \tag{2.26}$$
$$h(x) \leq 0 \tag{2.27}$$

where x is the vector of optimisation variables, f is the objective function and equations (2.26) and (2.27) are sets of equality and inequality constraints on x, respectively.

In optimal power flow, typical inequality constraints are bus voltage magnitude contingency state limits, generator output limits and branch power or current flow limits. The vector of optimisation variables x may consist of generator set-points, bus voltages, transformer tap settings etc. If x is empty then the formulation reduces to the general power flow problem described above.

A common objective in the optimal power flow problem is total system cost minimisation. For a network of n_g generators the objective function is

$$\min_{\theta,V_m,P_g} \sum_{k=1}^{n_g} C_P^k(P_g^k) + C_Q^k(Q_g^k) \tag{2.28}$$

where C_P^k and C_Q^k are cost functions (typically quadratic) of the set-points P_g^k and Q_g^k for generator k, respectively. Alternative objectives may be to minimise losses, maximise the voltage stability margin or minimise deviation of an optimisation variable from a particular schedule (Kallrath, Pardalos, Rebennack, & Scheidt, 2009, §18).

Nodal Marginal Prices

Many solution methods for optimal power flow have been developed since the problem was introduced by Carpentier (1962) and a review of the main techniques can be found in Momoh, Adapa, and El-Hawary (1999); Momoh, El-Hawary, and Adapa (1999).

One of the most robust strategies is to solve the Lagrangian function

$$\mathcal{L}(x) = f(x) + \lambda^\mathsf{T} g(x) + \mu^\mathsf{T} h(x), \tag{2.29}$$

where λ and μ are vectors of Lagrangian multipliers, using an Interior Point Method. When solved, the Lagrangian multiplier for a constraint gives the rate of change of the objective function value with respect to the constraint variable. If the objective function is equation (2.28), the Lagrangian multipliers λ_P^i and λ_Q^i for the power balance constraint at each bus i, given by equations (2.23) and (2.24), are the nodal marginal prices and can be interpreted as the increase in the total system cost for and additional injection at i of 1MW or 1MVAr, respectively. For a case in which none of the inequality constraints $h(x)$ (such as branch power flow or bus voltage limits) are binding, the nodal marginal prices are uniform across all buses and equal the cost of the marginal generating unit. When the constraints *are* binding, the nodal marginal prices are elevated for buses at which adjustments to power injection are required for the constraints to be satisfied. Nodal marginal prices are commonly used in agent-based electricity market simulation to determine the revenue for generating units as they reflect the increased value of production in constrained areas of the power system.

2.4 Reinforcement Learning

Reinforcement learning is learning from reward by mapping situations to actions when interacting with an uncertain environment (Sutton & Barto, 1998). An agent learns *what* to do in order to achieve a task through trial-and-error using a numerical reward or a penalty signal without being instructed *how* to achieve it. Some actions may not yield immediate reward or may affect the next situation and all subsequent rewards. Always, a compromise must be made between the exploitation of past experiences and the exploration of the environment through new action choices. In reinforcement learning an agent must be able to:

- Sense aspects of its environment,
- Take actions that influence its environment and,
- Have an explicit goal or set of goals relating to the state of its environment.

In the classical model of agent-environment interaction, at each time step t in a sequence of discrete time steps $t = 1, 2, 3 \ldots$ an agent receives as input some form of the environment's state $s_t \in \mathcal{S}$, where $\mathcal{S}$ is the set of possible states. From a set of actions $\mathcal{A}(s_t)$ available to the agent in state s_t and the agent selects an action a_t and performs it upon its environment. The environment enters a new state s_{t+1} in the next time step and the agent receives a scalar numerical reward $r_{t+1} \in \mathbb{R}$ in part as a result of its action. The agent then learns from the state representation, the chosen action a_t and the reinforcement signal r_{t+1} before beginning its next interaction. Figure **??** diagrams the classical agent-environment interaction event sequence in reinforcement learning.

For a finite number of states $\mathcal{S}$, if all states are Markov, the agent interacts with a finite Markov decision process (MDP). Informally, for a state to be Markov it must retain all relevant information about the complete sequence of positions leading up to the state, such that all future states and expected rewards can be predicted as well as would be possible given a complete history. A particular

MDP is defined for a discrete set of time steps by a state set $\mathcal{S}$, an action set $\mathcal{A}$, a set of state transition probabilities $\mathcal{P}$ and a set of expected reward values $\mathcal{R}$. In practice not all state signals are Markov, but should provide a good basis for predicting subsequent states, future rewards and selecting actions.

If the state transition probabilities and expected reward values are not known, only the states and actions, then samples from the MDP must be taken and a value function approximated iteratively based on new experiences generated by performing actions.

2.4.1 Value Function Methods

Any method that can optimise control of a MDP may be considered a reinforcement learning method. All search for an optimal policy π^* that maps state $s \in \mathcal{S}$ and action $a \in \mathcal{A}$ to the probability $\pi^*(s, a)$ of taking a in s and maximises the sum of rewards over the agents lifetime.

Each state s under policy π may be associated with a *value* $V^\pi(s)$ equal to the expected return from following policy π from state s. Most reinforcement learning methods are based on estimating the state-value function

$$V^\pi(s) = E\left\{ \sum_{t=0}^{\infty} \gamma^t r_t \,\middle|\, s_0 = s \right\} \tag{2.30}$$

where γ is a discount factor, with $0 \leq \gamma \leq 1$ and E indicates an estimate. Performing certain actions may result in no state change, creating a loop and causing the value of that action to be infinite for certain policies. The discount factor γ prevents values from going unbounded and represents reduced trust in the reward r_t as discrete time t increases. Many reinforcement learning methods estimate the action-value function

$$Q^\pi(s, a) = E\left\{ \sum_{t=0}^{\infty} \gamma^t r_t \,\middle|\, s_0 = s, a_0 = a \right\} \tag{2.31}$$

which defines the value of taking action a in state s under fixed policy π.

Temporal-Difference Learning

Temporal Difference (TD) learning is a fundamental concept in reinforcement learning (Sutton & Barto, 1998). TD methods do not attempt to estimate the state transition probabilities and expected rewards of the finite MDP, but estimate the value function directly. They learn to *predict* the expected value of total reward returned by the state-value function (2.30). For an exploratory policy π and a non-terminal state s, an estimate of $V^\pi(s_t)$ at any given time step t is updated using the estimate at the next time step $V^\pi(s_{t+1})$ and the observed reward r_{t+1}

$$V^\pi(s_t) = V^\pi(s_t) + \alpha\big[r_{t+1} + \gamma V^\pi(s_{t+1}) - V^\pi(s_t)\big] \tag{2.32}$$

where α is the learning rate, with $0 \leq \alpha \leq 1$, which controls how much attention is paid to new data when updating V^π. TD learning evaluates a particular policy and offers strong convergence guarantees, but does not learn better policies.

Sarsa

Sarsa (or modified Q-learning) is an on-policy TD control method that approximates the state-action value function in (2.31). Recall that the state-action value function for an agent returns the total expected reward for following a particular policy for selecting actions as a function of future states. The function is updated according to the rule

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\big[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)\big]. \tag{2.33}$$

This update also uses the action from the next time step a_{t+1} and the requirement to transition through state-action-reward-state-action for each time step is from where the algorithm derives its name. Sarsa is referred to as an on-policy method since it learns the same policy that it follows.

Q-Learning

Q-learning is an off-policy TD method that does not estimate the finite MDP directly, but iteratively approximates a state-action value function which returns the value of taking action a in state s and following an *optimal* policy thereafter. The same theorems used in defining the TD error also apply for state-action values that are updated according to

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\big[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)\big]. \tag{2.34}$$

The method is off-policy since the update function is independent of the policy being followed and only requires that all state-action pairs be continually updated.

Eligibility Traces

With the TD methods described above, only the value for the immediately preceding state or state-action pair is updated at each time step. However, the prediction $V(s_{t+1})$ also provides information concerning earlier predictions and TD methods can be extended to update a set of values at each step. An eligibility trace $e(s)$ represents how eligible the state s is to receive credit or blame for the TD error:

$$\delta = r_{t+1} + \gamma V^{\pi}(s_{t+1}) - V^{\pi}(s_t) \tag{2.35}$$

When extended with eligibility traces TD methods update values for all states

$$\Delta V_t(s) = \alpha \delta_t e_t(s) \tag{2.36}$$

For the current state $e(s) \leftarrow e(s) + 1$ and for all states the $e(s) \leftarrow \gamma\lambda e(s)$ where λ is the eligibility trace attenuated factor from which the extended TD methods TD(λ), Q(λ) and Sarsa(λ) derive their names. For $\lambda = 0$ only the preceding value is updated, as in the unextended definitions, and for $\lambda = 1$ all preceding state-values or state-action values are updated equally.

Action Selection

The ϵ-greedy approach to action selection is defined by a randomness parameter ϵ and a decay parameter d. A random number x_r where $0 \leq x_r \leq 1$ is drawn for each selection. If $x_r < \epsilon$ then a random action is selected, otherwise the perceived optimal action is chosen. After each selection the randomness is attenuated by d.

Action selection may also be accomplished using a form of the *softmax* method (Sutton & Barto, 1998, §2) using the Gibbs (or Boltzmann) distribution to select action k for the $(t+1)^{th}$ interaction with probability

$$p_{jk}(t+1) = \frac{e^{q_{jk}(t+1)/\tau}}{\sum_{l=0}^{K} e^{q_{jl}(t+1)/\tau}} \tag{2.37}$$

where τ is the *temperature* parameter. This parameter may be lowered in value over the course of an experiment since high values give all actions similar probability and encourage exploration of the action space, while low values promote exploitation of past experience.

2.4.2 Policy Gradient Methods

Value function based methods have been successfully applied with discrete lookup table parameterisation to many problems. However, the number of discrete states required increases exponentially as the dimensions of the state space increase and if all possibly relevant situations are to be covered then these methods become subject to Bellman's Curse of Dimensionality (Bellman, 1961). Value function based methods can be used in conjunction with function approximation techniques (artificial neural networks typically) in order to operate with continuous state and action spaces. However, value function approximation has been shown to cause these methods to exhibit poor convergence or divergence characteristics, even in simple systems (Peters & Schaal, 2008).

These convergence problems have motivated research into alternative methods that can operate with continuous environments, such as policy gradient methods. These algorithms make small incremental changes to the parameter vector θ of a policy function approximator. Using artificial neural networks the parameters are the weights of the network connections (see below). Policy gradient methods update θ in the direction of the gradient of some policy performance measure Y with respect to the parameters

$$\theta_{i+1} = \theta_i + \alpha \frac{\partial Y}{\partial \theta_i} \tag{2.38}$$

where α is a positive definite step size learning rate.

As well as working with continuous state and actions space, policy gradient methods offer strong convergence guarantees, do not require all states to be continually updated and although uncertainty in state data can degrade policy performance, the techniques need not be altered.

Policy gradient methods are differentiated largely by the techniques used to obtain an estimate of the policy gradient $\partial Y / \partial \theta$. The most successful real-world robotics results have been yielded using Williams' Reinforce likelihood ratio methods (Williams, 1992) and natural policy gradient methods such as Natural Actor-Critic (Peters & Schaal, 2008).

Artificial Neural Networks

This subsection provides a very brief introduction to the theory of artificial neural networks, concentrating on the aspects utilised in reinforcement learning. These mathematical models mimic aspects of biological neural networks, such as the human brain, and are widely used in supervised learning applications. A wealth of literature is available that covers the field in greater depth (Bishop, 1996; Fausett, 1994).

In reinforcement learning, the most widely used type of artificial neural network is the multi-layer feedforward network (or multi-layer perceptron). This model consists of an input layer and an output layer of artificial neurons, plus any number of optional hidden layers. Weighted connections link the neurons, but unlike architectures such as the recurrent neural network, only neurons from adjacent layers connect. Most commonly, a fully connected scheme is used in which all neurons from one layer are connected to all neurons in the next. Figure diagrams a fully connected three layer feedforward network.

McCulloch and Pitts (1943) conceived of an artificial neuron j that computes a function g as a weighted sum of all n inputs

$$y_j(x) = g\left(\sum_{i=0}^{n} w_i x_i\right) \tag{2.39}$$

where $(w_0 \ldots w_n)$ are weights applied to the inputs $(x_0 \ldots x_n)$. In an multi-layer neural network the output y_j forms part of the input to the neurons in any following layer. The activation function g is typically either:

- Linear, where the output is simply the weighted sum of inputs,
- A threshold function, with an output of either 0 or 1,
- Sigmoidal, such that the output is between 0 and 1,
- A hyperbolic tangent, where the output ranges between -1 and 1.

The paramemters of the activation functions could be adjusted along with the connection weights to tune the transfer function between input and output that the network provides. To simplify this process a *bias* node that always outputs 1 may be added to a layer and connected to all neurons in the following layer. This can be shown to allow the activation function parameters to be removed and for network adjustment to occur using only connection weights.

The output is obtained during the network's *execution* phase by presenting an input to the input layer that propagates through. It can be shown that a suitably configured feedforward network with one hidden layer can approximate any non-linear function [ref].

2.4.3 Roth-Erev Method

The reinforcement learning method formulated by Alvin E. Roth and Ido Erev is based on empirical results obtained from observing how humans learn decision making strategies in games against multiple strategic players (Roth et al., 1995; Erev & Roth, 1998). It learns a stateless policy in which each action a is associated with a value q for the propensity of its selection. In time period t, if agent j performs action a' and receives a reward $r_{ja'}(t)$ then the propensity value for action a at time $t+1$ is

$$q_{ja}(t+1) = \begin{cases} (1-\phi)q_{ia}(t) + r_{ja'}(t)(1-\epsilon), & a = a' \\ (1-\phi)q_{ia}(t) + r_{ja'}(t)(\frac{\epsilon}{A-1}), & a \neq a' \end{cases} \tag{2.40}$$

where A is the total number of feasible actions, ϕ is the *recency* parameter and ϵ is the *experimentation* parameter. The recency (forgetting) parameter degrades the propensities for all actions and prevents propensity values from going unbounded. It is intended to represent the tendency for players to forget older action choices and to prioritise more recent experience. The experimentation parameter prevents the probability of choosing an action from going to zero and encourages exploration of the action space.

Erev and Roth proposed action selection according to a discrete probability distribution function, where action k is selected for interaction $t+1$ with probability

$$p_{jk}(t+1) = \frac{q_{jk}(t+1)}{\sum_{l=0}^{K} q_{jl}(t+1)} \tag{2.41}$$

Since $\sum_{l=0}^{K} q_{jl}(t+1)$ increases with t, a reward $r_{jk}(t)$ for performing action k will have a greater effect on the probability $p_{jk}(t+1)$ during early interactions while t is small. This is intended to represent Psychology's Power Law of Practice in which it is qualitatively stated that, with practice, learning occurs at a decaying exponential rate and that a learning curve will eventually flatten out.

Modified Roth-Erev Method

Two shortcomings of the basic Roth-Erev algorithm have been identified and a modified formulation proposed by Nicolaisen, Petrov, and Tesfatsion (2002). The two issues are that

- the values by which propensities are updated can be zero or very small for certain combinations of the experimentation parameter ϵ and the total number of feasible actions A and

- all propensity values are decreased by the same amount when the reward, $r_{jk'}(t)$ is zero.

Under the variant algorithm, the propensity for agent j to select action a for interaction $t+1$ is:

$$q_{ja}(t+1) = \begin{cases} (1-\phi)q_{ia}(t) + r_{ja'}(t)(1-\epsilon), & a = a' \\ (1-\phi)q_{ia}(t) + q_{ja}(t)(\frac{\epsilon}{A-1}), & a \neq a' \end{cases} \tag{2.42}$$

As with the original Roth-Erev algorithm, the propensity for selection of the action that the reward is associated with is adjusted by the experimentation parameter. All other action propensities are adjusted by a small proportion of their current value.

2.5 Summary

The combination of electricity markets and electric power systems presents a complex dynamic environment to participants. Network power flows are non-linear functions of the bus voltages and thus one party's generation or consumption decisions effect all other parties. Substantial modifications to the design of the UK's electricity trading arrangements have been required since the their introduction two decades ago and further changes will likely be necessary if ambitious greenhouse gas emission reduction commitments are to be met.

The main electricity trading mechanisms can be modelled using well established mathematical optimisation formulations. Robust techniques exist for computing solutions to these problems, which also provide price information that reflects the network topology and conditions.

The combination of non-linear optimisation problems and participant behavioural models is beyond the capabilities of conventional equilibrium approaches to market simulation when analysing large systems. An alternative is to take a "bottom-up" approach to modelling them and examine the system dynamics that result from interactions between goal driven individuals.

Reinforcement learning is an unsupervised machine learning technique that can be used to model the dynamic behaviour of these individuals. Traditional methods associated a *value* with each state and the available actions, but are limited to realtively small discrete problems. Policy gradient methods that search directly in the space of the parameters of an action selection policy can operate in continuous environements and have good convergence properties.

Chapter 3

Related Work

This chapter describes the research in this thesis in the context of similar work. It reviews previously published research with particular focus made on the learning methods and simulation models used. For a similar review with greater emphasis on criticism of simulation results and the conclusions drawn from them, the interested reader is referred to Weidlich and Veit (2008).

3.1 Custom Learning Methods

The earliest agent-based electricity market simulations in the literature do not utilise traditional learning methods from Artificial Intelligence, but rely upon custom heuristic methods. They are typically formulated using the author's intuition and represent basic trading rules, but do not encapsulate many of the key concepts from reinforcement learning theory.

3.1.1 Market Power

Under Professor Derek Bunn, researchers from the London Business School performed some the first and most reputable agent-based electricity market simulations. Their research was initially motivated by proposals in 1999 to transform the structure of The England and Wales Electricity Pool, with the aim of combating the perceived generator market power that was widely believed to be resulting in elevated market prices.

In Bower and Bunn (2001) a detailed model of electricity trading in England and Wales is used to compare day-ahead and bilateral contract markets under uniform price and discriminatory settlement. Twenty generating companies operating in the Pool during 1998 are modelled as agents endowed with portfolios of generating plant. Plant capacities, costs and expected availabilities are synthesised from public and private data sources and the author's own estimates. In simulations of the day-ahead market, each agent submits a single price for the following simulated trading day, for each item of plant in its portfolio. Whereas, under the bilateral contract model, 24 bids are submitted for each generator, corresponding to each hour of the following simulated day. Revenues are calculated at the end of each trading day and are determined either by the bid price of the marginal unit or the generator's own bid price. Each generating plant is characterised in part by an estimated target utilisation rate that represents its desire for forward contract cover. The agents learn to achieve this utilisation rate and then improve profitability.

If the utilisation rate is not achieved, a random percentage from a uniform distribution with a range of $\pm 10\%$ and 0% mean is subtracted from the bid price of all generators in the agent's portfolio. Agents with more than one generator transfer successful bidding strategies between plant by setting the bid price for a generator to the level of the next highest submitted bid price if the generator sold at a price lower than that of other generators in the same portfolio. If an agent's total profit does not increase, a random percentage from the same distribution as above is added or subtracted from the bid price from the previous day for each of its generators. A cap on bid prices is imposed at £1000 in each period. Demand follows a 24-hour profile based on the 1997-1998 peak winter load pattern. The response of the load schedule to high prices is modelled as a reduction of 25MW for every £1/MWh that the system marginal price rises above £75/MWh.

In total, 750 trading days are simulated for each of the four combinations of a day-ahead market and the bilateral trading model under uniform pricing and discriminatory settlement. Prices were found to generally be higher under pay-as-bid pricing for both market models. Agents with larger portfolios are shown to have a significant advantage over smaller generators due to their greater ability to gather scarce market price information and distribute it among generators.

The research question addressed is a common one in agent-based electricity market simulation and the paper demonstrates the use of a relatively simple learning method. It is an example of how such simulations need not be restricted to simple models, but can be scaled to study systems at a national level.

In Bower, Bunn, and Wattendrup (2001) a more sophisticated custom learning method, resembling the Roth-Erev method, is applied to a more detailed model of the New Electricity Trading Arrangements. The balancing mechanism is modelled as a one-shot market, that follows the contracts market, to which increment and decrement bids are submitted. Active demand side participation is modelled and generator dynamic constraints are represented by limiting the number of off/on cycles per day. Again, transmission constraints and regional price variations are ignored.

Supplier and generator agents are assigned an optimal value for exposure to the balancing mechanism that is typically low due to high price and volume uncertainty. The agents learn to maximise profit, but profits are penalised if the objective for balancing mechanism exposure is not achieved. They learn policies for pricing markups on the bids submitted to the power exchange and the increments and decrements submitted to the balancing mechanism. Markups in the power exchange are relative to prices from the previous day and markups on balancing mechanism bids are relative to power exchange bid prices on the same day. Different markup ranges are specified for generators and suppliers in the power exchange and balancing mechanism and each is partitioned into ten discrete intervals.

As with the Roth-Erev method, a probability for the selection of each markup is calculated by the learning method. Daily profits and acceptance rates for bids/offers from previous trading days are extrapolated out to determine expected values and thus the expected reward for each markup. The markups are then sorted according to expected reward in descending order. The perceived utility of each markup j is

$$U_j = \mu \left(\frac{\phi - n}{\phi} \right)^{i_j - 1} \tag{3.1}$$

where i is the index of j in the ordered vector of markups and ϕ is a search parameter. High values of ϕ cause the agent to adopt a more exploratory markup selection policy. For all of the experiments $\mu = 1000$, $\phi = 4$, $n = 3$ and the probability of selecting markup j is

$$Pr_j = \frac{U_j}{\sum_{k=1}^{K} U_k} \tag{3.2}$$

for K possible markups.

A representative model of the England and Wales system with 24 generator agents, associated with a total of 80 generating units, and 13 supplier agents is analysed over 200 simulated trading days. The authors draw conclusions on the importance of accurate forecasts, greater risk for suppliers than generators, the value of flexible plant and the influence of capacity margin on opportunities for collusive behaviour. The same learning method is applied in D. W. Bunn and Oliveira (2003) as part of an inquiry by the Competition Commission into whether two specific companies in the England and Wales electricity market had enough market power to operate against the public interest.

These papers show that previous research by a respected author was perceived to be of sufficient value for the approach the be pursued further, but the need to improve participant and market models was recognised. Despite neglecting transmission system constraints, the work is an ambitious attempt to extrapolate results out to consquences for a national market.

Visudhiphan and Ilic (1999) is another early publication on agent-based simulation of electricity markets in which a custom learning method is used. The simulations comprise only three generators, market power is assumed, and the authors analyse the mechanisms by which the market power is exercised. Two bid formats are modelled. The single-step supply function (SSF) model requires each generator agent to submit a price and a quantity, where the quantity is determined by the generator's marginal cost function. The linear supply function (LSF) model requires each generator agent to submit a value corresponding to the slope of its supply function. The bid price or slope value for generator i after simulation period t is

$$x_i(t+1) = x_i(t) + b_i(p_m(t))u_i(t) \tag{3.3}$$

where $b_i \in \{-1, 0, 1\}$ is the reward as a function of the market clearing price p_m from stage t and u_i is a reward gain or attenuation parameter. The calculation of b_i is defined according to strategies for estimated profit maximisation and competition to be the base load generator. Both elastic and inelastic load models are considered. Using the SSF model, the two strategies are compared in a day-ahead market setting, using a case where there is sufficient capacity to meet demand and a case where there is excessive capacity to the point where demand can be met by just two of the generators. The LSF model is analysed using both day-ahead and hour-ahead markets with inelastic load. The hour-ahead simulation is repeated with elastic demand response.

The number of if-then rules required to define participant strategies in this paper is demonstrates a drawback of implementing custom learning methods that is exacerbated when defining multiple strategies.

A similar custom learning method is compared with two other algorithms in Visudhiphan (2003). The custom method is designed specifically for the power pool model used and employs separate policies for selecting bid quantities and prices according to several if-then rules that attempt to capture capacity withholding behaviour. The method is compared with algorithms developed in Auer, Cesa-Bianchi, Freund, and Schapire (2003) for application to the n-armed bandit problem (Robbins, 1952; Sutton & Barto, 1998, §2.1) and a method based on evaluative feedback with softmax action selection.

In the algorithms from Auer et al. (2003) each action $i = 1, 2, \ldots K$, for K possible actions, is associated with a weight $w_t(i)$ in simulation period $t \in T$, for T simulation periods, that is used in determining the action's probability of selection

$$p_i(t) = (1-\gamma)\frac{w_i(t)}{\sum_{j=1}^{K} w_j(t)} + \frac{\gamma}{K} \tag{3.4}$$

where γ is a tuning parameter, with $0 < \gamma \leq 1$, that is initialised such that

$$\gamma = \min\left\{\frac{3}{5}, 2\sqrt{\frac{3}{5}\frac{K \ln K}{T}}\right\}. \tag{3.5}$$

Using the received reward $x_t(i_t)$, the weight for action j in period $t+1$ is

$$w_{t+1}(j) = w_t(i)\exp\left(\frac{\gamma}{3K}\left(\hat{x}_t(i) + \frac{\alpha}{p_t(i)\sqrt{KT}}\right)\right) \tag{3.6}$$

where

$$\hat{x}_t(i) = \begin{cases} x_t(j)/p_t(i) & \text{if } j = i_t \\ 0 & \text{otherwise} \end{cases} \tag{3.7}$$

and

$$\alpha = 2\sqrt{\ln(KT/\gamma)}. \tag{3.8}$$

In the evaluative feedback method from Sutton and Barto (1998, §2) each action i has a value $Q_t(i)$ in simulation period t equal to the expected average reward if that action is selected. The value of action i in the $(t+1)^{th}$ period is

$$Q_{t+1}(i) = \begin{cases} (1-\alpha)Q_t(i) + \alpha r_t(i) & \text{if } i_{t+1} = i \\ Q_t(i) & \text{otherwise} \end{cases} \tag{3.9}$$

where α is a constant *step-size* parameter with $0 < \alpha \leq 1$.

Extensive simulation results are presented and the choice of learning method is found to have a significant impact on agent performance, but no quantitative comparison measure is provided and no conclusions are drawn as to which method is the superior.

3.1.2 Financial Transmission Rights

In Ernst, Minoia, and Ilic (2004) a custom learning method is defined and used to study generator and supplier profits where financial transmission rights are included in the electricity market. A two node transmission system is defined with one lossless transmission line of limited capacity that is endowed to a transmission operator agent. Generator agents submit bids for their respective generating units and the transmission owner submits a bid representing the cost per MW of transmitting power between the nodes. The market operator clears the bids, minimising costs while balancing supply and demand and not breaching the line capacity. Prices at each node are calculated to provide a signal to the agents that captures both energy and transmission costs.

Each agent selects its bid according to a calculation of the reward that it would expect to receive if all other agents were to bid as they did in the previous stage. If multiple bids are found to have the same value then the least expensive is selected. In the first period, previous bids are assumed to be at marginal cost.

Several case studies are examined with different numbers of generators and line capacities, but few explicit conclusions are drawn. Financial transmission rights are an important issue in electricity markets, but the learning algorithm and network model are perhaps overly simple for practical conclusions to be drawn. Agent-based simulation has the potential to provide further insight into financial transmission rights and the issue is one that perhaps ought to be revisited as advances in the field are made.

3.2 Simulations Applying Q-learning

Recent agent-based simulations of electricity markets has been carried out with participant's behavioral aspects modelled using the Q-learning methods described in Section 2.4.1.

3.2.1 Nash Equilibrium Convergence

The most prominent work in which Q-learning is applied was conducted at the Swiss Federal Institutes of Technology in Zurich and Lausanne. The foundations for this work were laid in Krause et al. (2004) with a comparison of agent-based modelling using reinforcement learning and Nash equilibrium analysis when assessing network constrained power pool market dynamics. Parameter sensitivity of comparison results were later analysed in Krause et al. (2006).

The authors model a mandatory spot market which is cleared using a DC optimal power flow formulation. A five bus power system model is defined with three generators and four inelastic and constant loads. Linear marginal cost functions

$$C_{g,i}(P_{g,i}) = b_{g,i} + s_{g,i}P_{g,i} \tag{3.10}$$

are defined for each generator i where $P_{g,i}$ is the active power output, $s_{g,i}$ is the slope of the cost function and $b_{g,i}$ is the intercept. Suppliers are given the option to markup their bids to the market not by increasing $s_{g,i}$, but increasing $b_{g,i}$ by either 0, 10, 20 or 30%.

Nash equilibrium is computed by clearing the market for all possible markup combinations and determining the actions for which no player is motivated to deviate from, as it would result in a decrease in expected reward. Experiments are conducted in which there is a single Nash equilibrium and where there are two Nash equilibria.

An ϵ-greedy strategy (Sutton & Barto, 1998) is applied for action selection and a *stateless* action value function is updated at each time step t according to

$$Q(a_t) = Q(a_t) + \alpha(r_{t+1} - Q(a_t)) \tag{3.11}$$

where α is the learning rate. Further to Krause et al. (2004), simulations with discrete sets of values for the parameters α and ϵ were carried out in Krause et al. (2006). While parameter variations effected the frequency of equilibrium oscillations, Nash equilibrium was still approached and the oscillatory behaviour observed for almost all of the combinations.

The significance of this research is that is verifies that the agent-based approach settles at the same theoretical optimum as with closed-form equilibrium approaches and that exploratory policies result in the exploitation of multiple equlibria if they exist.

Convergence to a Nash equilibrium is also shown in Naghibi-Sistani, Akbarzadeh-Tootoonchi, Javidi-D.B., and Rajabi-Mashhadi (2006). Boltzman (soft-max) exploration is used for action selection with the temperature parameter adjusted during the simulations. A modified version of the IEEE 30 bus test system is used with the number of generators reduced from nine to six. No optimal power flow formulation or details of the reward signal used are provided. Generators are given a three step action space where the slope of a linear supply function may be less than, equal to or above marginal cost. The experimental results show that with temperature parameter adjustment Nash equilibrium is achieved and the oscillations associated with ϵ-greedy action selection are avoided.

Dynamic modification of the softmax temperature parameter is a technique that is employed in several other such publications, but as noted in Weidlich and Veit (2008, pp. 1746), the approach taken in this paper conflicts with the need to balance exploration and exploitation.

3.2.2 Congestion Management Techniques

Having validated the suitability of an agent-based, bottom-up, approach to assessing the evolution of market characteristics, the authors applied the same technique to compare congestion management schemes in Krause and Andersson (2006). The first scheme considered is locational marginal pricing (or nodal pricing) where congestion is managed by optimising the output of generators with respect to maximum social welfare. The "market splitting" scheme they considered is similar to locational marginal pricing, but the system is subdivided into zones, within which the nodal prices are uniform. The final "flow based market coupling" scheme also features uniform zonal pricing, but uses a simplified representation of the network. Power flows within the zones are not represented and all lines between zones are aggregated into one equivalent interconnector.

As an alternative to the conventional DC optimal power flow formulation, line power flows computation is done using a power transfer distribution factor (PTDF) matrix. The $(i, j)^{th}$ element of the PTDF matrix corresponds to the change in active power flow on line j given an additional injection of 1MW at the slack bus and corresponding withdrawal of 1MW at node i (Grainger & Stevenson, 1994).

The congestion management schemes get evaluated under perfect competition, where suppliers bid at marginal cost, and under oligopolistic competition, in which markups of 5% and 10% can be added to marginal cost. The benefits obtained between reward at marginal cost and a maximum markup are used to assess market power. The experimental results show that market power allocations are different under each of the three constraint management schemes.

This is a compelling example of how optimal power flow can be used with traditional reinforcement learning methods to address an important research question. The decision not to define environment states is unusual for a Q-learning application and the impact of this deserves investigation.

3.2.3 Gas-Electricity Market Integration

The Q-learning method from Krause et al. (2004, 2006) is used to analyse strategic behaviour in integrated electricity and gas markets in Kienzle, Krause, Egli, Geidl, and Andersson (2007). Again, power flows are computed using a PTDF matrix. Pipeline losses in the gas network are approximated using using a cubic function of flow and three combined gas and electricity models are compared.

In the first model, operators of gas-fired power plant submit separate bid functions for gas and electricity. Bids are then cleared as a single optimisation problem. In model two, operators submit one offer for their capacity to convert gas to electricity. In the third model, bids are submitted only to the electricity market, after which gas is purchased regardless of price. Gas supply offers are modelled as a linear function with no strategic involvement. The models are compared in terms of social welfare, using a three bus power system model with three non-gas-fired power plants and one gas-fired plant.

The experimental results show little difference between electricity prices and social welfare prices between the models. However, this research illustrates the interest in and complexity associated with modelling relationships between multiple markets. The authors recognise the need for further and more detailed simulation in order to improve evaluation of market coupling models.

While this work is of a preliminary nature, it is an important step towards achieving greater understanding the interrelationships between gas and electricity markets using agent-based simulation. Further neglect of state information in the Q-learning method possibly alludes to the difficulty of creating discrete representations of largely continuous environments.

3.2.4 Electricity-Emissions Market Interactions

Researchers at the Argonne National Laboratory have published results from a preliminary study of interactions between *emissions* and electricity markets (J. Wang, Koritarov, & Kim, 2009). A cap-and-trade system for emissions is modelled where generator companies are allocated with CO_2 allowances that may subsequently be traded. Generator companies are assumed to have negligible influence on market clearing prices in the emissions market and allowance prices from the European Energy Exchange were used. In the electricity market, an oligopoly structure is assumed and bids are cleared using a DC optimal power flow formulation.

To improve selection of the ϵ parameter for exploratory action selection, a simulated annealing (SA) Q-learning method based on the Metropolis criterion (Guo, Liu, & Malec, 2004) is used. Under this method ϵ is changed at each simulation step to allow solutions to escape from local optima. A two bus system is used to study cases in which allowance trading is not used, allowances can be exchanged in the emissions market and with variations in the allowance allocations. A one year, hourly load profile with a summer peak is used to model changes in demand. The electricity market is cleared for each simulated hour and the emissions market gets cleared at the end of each simulated week.

The agents learn, when they have a deficit of allowances, to borrow future allowances in the summer when load and allowance prices are high. Conversely, when having a surplus, they learn to sell at this time. In the third case, the authors show the sensitivity of profits to initial allocations and conclude that the experimental results can not be generalised. The authors cite further model validation and agent learning method improvements as necessary future work.

The complexity of the combined electricity and emissions market model illustrates how the search spaces for learning methods grows dramatically as models are expanded. A problem that policy gradient learning methods seek to address.

3.2.5 Tacit Collusion

The SA-Q-learning method was previously used in Tellidou and Bakirtzis (2007) by researchers from the University of Thessaloniki to study capacity withholding and tacit collusion among electricity market participants. A mandatory spot market is implemented, where bid quantities may be less than net capacity and bid prices may be marked up upon marginal cost by increasing the slope of a linear cost function. Again the market is cleared using a DC optimal power flow formulation and locational marginal prices are used to calculate profits that are used as the reinforcement signal in the learning process. Demand is assumed to be inelastic and transmission system parameters constant between simulation periods.

A simple two node power system model containing two generators is applied in three test cases. In a reference case, each generator bids full capacity at marginal cost. In the second case, generators bid quantities in steps of 10MW and price markups in steps of €2/MWh. In the third case, the same generation capacity is split among eight identical generators to increase the level of competition. The experimental results show that generators learn to withhold capacity and develop tacit collusion strategies to capture congestion profits.

This work is similar to earlier research from other institutions and makes only a minor contribution. It suggests that there is potential to accelerate advancement in this field through increased collaboration and sharing of software source code.

3.3 Simulations Applying Roth-Erev

Roth and Erev's reinforcement learning method (defined in Section 2.4.3) has received considerable attention from the agent-based electricity market simulation community.

3.3.1 Market Power

In Nicolaisen et al. (2002) an agent-based model of a wholesale electricity market with both supply and demand side participation is constructed. It is used to study market power and short-run market efficiency under discriminatory pricing through systematic variation of concentration and capacity conditions.

To model the power system, each trader is assigned values of available transmission capability (ATC) with respect to each of the other traders. Offers from buyers and sellers are matched on a merit order basis, with quantities restricted by ATC values. Two issues with the original Roth-Erev method are observed and the modified version defined in Section 2.4.3 is proposed.

A maximum markup (markdown) of $40/MWh is specified for each seller (buyer). Traders are not permitted to make negative profits and the feasible price range is divided into 30 offer prices for 1000 auction rounds cases and 100 offer prices for 10000 auction round cases. The parameters of the Roth-Erev method are calibrated using direct search within reasonable ranges. Nine combinations of buyer and seller numbers and total trading capacities are tested using the calibrated parameter values and *best-fit* values determined empirically in Erev and Roth (1998).

The experimental results show that good market efficiency is achieved under all configurations and sensitivity to method parameter changes is low. Levels of market power are found to be strongly predictive and little difference is found between cases in which opportunistic price offers are permitted and when traders are forced to bid at marginal cost. The results are compared with those from Nicolaisen, Smith, Petrov, and Tesfatsion (2000), in which genetic algorithms are used. The authors conclude that the reinforcement learning approach leads to higher market efficiency due their adaption according to *individual* profits.

Genetic algorithms were a popular alternative to reinforcement learning methods in early agent-based electricity market research. This paper compares the two and illustrates some of the reasons that they have now been largely abandoned in this field. The modified Roth-Erev method proposed in this paper is later used in several other publications.

Further research from Iowa State University, involving the modified Roth-Erev method, has used the AMES wholesale electricity market test bed. A detailed description of AMES is provided in Appendix A.7 below. In Li and Tesfatsion (2009b) it is used to investigate strategic capacity withholding in a wholesale electricity market design proposed by the U.S. Federal Energy Regulatory Commission in April 2003. A five bus power system model with five generators and three dispatchable loads is defined and capacity withholding is represented by permitting traders to bid lower than true operating capacity and higher than true marginal costs.

Comparing results from a benchmark case (in which true production costs are reported, but higher than marginal cost functions may be reported) and cases in which reported production limits may be less than the true values, the authors find that with sufficient capacity reserve there is no evidence to suggest potential for inducing higher net earnings through capacity withholding in the market design.

AMES was the first agent-based electricity market simulation program to be released as open source, but while there are serveral publications on the project, papers involving its application are scarce. This shows how niche the field is and the challenge faced if such projects are to benefits from the collaboration of communities that often leads to successful open source software projects.

3.3.2 Italian Wholesale Electricity Market

Rastegar, Guerci, and Cincotti (2009) from the University of Genoa used the modified Roth-Erev method to study strategic behaviour in the Italian wholesale electricity market. An accurate model of the actual clearing procedure is implemented and the model of the Italian transmission system, including an interconnector to Sicily and zonal subdivision, illustrated in Figure **??** is defined. Within each of the 11 zones, thermal plant is combined according to technology (coal, oil, combined cycle gas, turbo gas and repower) and associated with one of 16 generation companies according to the size of the companies share. The resulting 53 agents are assumed to bid full capacity and may markup bid prices in steps of 5%, with a maximum markup of 300%.

Bids are cleared using a DC optimal power flow formulation with generation capacity constraints and zone interconnector flow limits. Agents are rewarded according to a uniform national price, computed as a weighted average of zonal prices with respect to zonal load. Using real hourly load data it is shown that in experiments in which agents learn their optimal strategy, historical trends can be replicated in all but certain hours of peak load. The authors state a desire to test different learning methods and perform further empirical validation.

3.3.3 Vertically Related Firms and Crossholding

In Micola, Banal-Estañol, and Bunn (2008) a multi-tier model of wholesale natural gas, wholesale electricity and retail electricity markets is studied using another variant of the Roth-Erev method. Coordination between strategic business units (SBU) within the same firm, but participating in different markets, is varied systematically and profit differences are analysed.

A two-tier model involves firms with two associated agents whose rewards r_1 and r_2 are initially independent. A "reward independence" parameter α is used to control the fraction of profit from one market that is used in rewarding the agent in the other market. The total rewards are

$$R_1(t) = (1-\alpha)r_1(t) + \alpha r_2(t) \tag{3.12}$$

and

$$R_2(t) = (1-\alpha)r_2(t) + \alpha r_1(t). \tag{3.13}$$

Each action a is a single price bid between zero and the clearing price from the preceding market. The Roth-Erev method is modified such that similar actions, $a-1$ and $a+1$, are reinforced also. For each agent i, the action selection propensities in auction round t are

$$p_a^i(t) = \begin{cases} (1-\gamma)p_a^i(t-1) + R_i(t) & \text{if } s = k \\ (1-\gamma)p_a^i(t-1) + (1-\delta)R_i(t) & \text{if } s = k-1 \text{ or } s = k+1 \\ (1-\gamma)p_a^i(t-1) & \text{if } s \neq k-1,\, s \neq k \text{ or } s \neq k+1 \end{cases} \tag{3.14}$$

where δ, with $0 \leq \delta \leq 1$, is the local experimentation parameter, γ is the discount parameter and $i \in \{1, 2\}$. Actions whose probability of selection fall below a specified value are removed from the action space.

The initial simulation consists of two wholesalers and three retailers and α is varied from 0 to 0.5 in 51 discrete steps. The experiment is repeated using a three tier model in which two natural gas shippers supply three electricity generators who, in turn, sell to four electricity retailers. The results show a rise in market prices as reward interdependence is increased and greater profits for integrated firms.

The same alternative formulation of the Roth-Erev method is also used in Micola and Bunn (2008) to analyse the effect on market prices of different degrees of producer crossholding[1] under private and public bidding information. Crossholding is represented with the introduction of a factor to each agent's reward function that controls the fraction of profit from the crossowned rival that the agent receives. Public information availability is modelled using a vector of probabilities for selection of each possible action that is the average of each agent's private probability and is available to all agents.

The degree to which the public probabilities influence the agent's action selection probability from equation (2.41) is varied systematically in a series of experiments, along with crossholding levels and buyer numbers. The results are illustrated using three-dimensional plots and show a direct relationship between crossholding and market price. The conclusions drawn on market concentration by the authors are dependant upon the ability to model both the demand and supply side participation in the market and the authors state that this shows, to a certain extent, the value of the agent-based simulation approach.

3.3.4 Two-Settlement Markets

In Weidlich and Veit (2006) the modified Roth-Erev method is used to study interrelationships between contracts markets and balancing markets. Bids on the day-ahead contracts market consist of a price and a volume, which are assumed to be the same for each hour of the day. Demand is assumed to be fixed and inelastic. Bids on the balancing market consist of a reserve price, a *work* price and an offered quantity. The reserve price is that which must be paid for the quantity to be kept on standby and the work price must be paid if that quantity is called upon for transmission system stabilisation. No optimal power flow formulation or power system model is defined.

At the day-ahead stage, contract market and balancing market (according to reserve price) bids are cleared by stacking in order of ascending price until the forecast demand is met. On the following day, accepted balancing bids are cleared according to work price such that requirements for reserve dispatch are met.

Bid prices on the contracts market are stratified into 21 discrete values between 0 and 100 and bid quantities into six discrete values between 0 and maxi-

[1] Crossholdings occur when one publicly traded firm owns stock in another such firm.

mum capacity, giving 126 possible actions. Bid quantities on the balancing market equal the capacity remaining after contract market participation. 21 discrete capacity prices between 0 and 500 and 5 work prices between 0 and 100 are permitted, giving 105 possible actions in the balancing market. Separate instances of the modified Roth-Erev method are used to learn bidding strategies for each agent in each of the markets.

Interrelationships between the markets are studied using four scenarios in which the order of market execution and the balancing market pricing mechanism (discriminatory or pay-as-bid) are changed. Clearing prices in the market executed first are shown to have a marked effect on prices in the following market. The authors find agent-based simulation to be a suitable tool for reproducing realistic market outcomes and recognise a need for more detailed models with larger action domains.

In the same year, the authors collaborated with Jian Yao and Shmuel Oren from the University of California to study the dynamics between two settlement markets using the modified Roth-Erev method (Veit, Weidlich, Yao, & Oren, 2006). The markets are a forward contracts market, in which transmission constraints are ignored, and a spot market that is cleared using a DC optimal power flow formulation with line flows calculated using a PTDF matrix. The authors state that suppliers utility functions are to include aspects of risk aversion in future work. The use of some measure of risk adjusted return to assess performance is commonplace in economics research, but is currently lacking from the agent-based electicity market simulation literature.

Zonal prices are set in the forward market as weighted averages of nodal prices with respect to historical load shares. Profits are determined using the zonal prices and nodal prices from optimisation of the spot market. Demand is assumed inelastic to price, but different contingency states with peak and low demand levels are examined. A stylised 53 bus model of the Belgian electricity system from Yao, Oren, and Adler (2007) and Yao, Adler, and Oren (2008) is used to validate the results against those obtained using equilibrium methods. The nineteen generators are divided among two firms which learn strategies for bid price and quantity selection using the modified Roth-Erev method with a set of fixed parameter values taken from Erev and Roth (1998). The results show that the presence of a forward contracts market produces lower overall electricity prices and lower price volatility. The authors note that risk aversion is to be included in suppliers utility functions in future work.

3.4 Policy Gradient Reinforcement Learning

Policy gradient reinforcement learning methods, defined in Section 2.4.2, have been successfully applied in both laboratory and operational settings (Sutton et al., 2000; Peters & Schaal, 2006; Peshkin & Savova, 2002). This section reviews the *market* related applications of these methods.

3.4.1 Financial Decision Making

Conventionally, *supervised* learning techniques are used in financial decision making problems to minimise errors in price forecasts and are trained on sample data. In Moody, Wu, Liao, and Saffell (1998) a recurrent reinforcement learning method is used to optimise investment performance without price forecasting. The method is "recurrent" in that it uses information from past decisions as input to the decision process. The authors compare direct profit and the Sharpe ratio (Sharpe, 1966, 1994) as reward signals. The Sharpe ratio is a measure of risk adjusted return defined as

$$S_t = \frac{\text{Average}(r_t)}{\text{Standard Deviation}(r_t)} \tag{3.15}$$

where r_t is the return for period t.

The parameters θ of the trading system are updated in the direction of the steepest accent of the gradient of some performance function U_t with respect to θ

$$\Delta\theta_t = \rho \frac{dU_t(\theta_t)}{d\theta_t} \tag{3.16}$$

where ρ is the learning rate. Direct profit is the simplest performance function defined, but assumes traders are insensitive to risk. Investors being sensitive to losses are, in general, willing to sacrifice potential gains for reduced risk of loss. To allow on-line learning and parameter updates at each time period, the authors define a *differential* Sharpe ratio. By maintaining an exponential moving average of the Sharpe ratio, the need to compute return averages and standard deviations for the entire trading history at each simulation period is avoided. Alternative performance ratios, including the Information ratio, Appraisal ratio and Sterling ratio, are also mentioned.

Simulations are conducted using artificial price data, equivalent to one year of hourly trade in a 24-hour market, and using 45 years of monthly data from the Standard & Poor (S&P) 500 stock index and 3 month Treasury Bill (T-Bill) data. In a portfolio management simulation, in which trading systems invest portions of their wealth among three different securities, it was shown that trading systems maximising the differential Sharpe ratio, produced more consistent results and achieved higher risk adjusted returns than those trained to simply maximise profit. This result is important as the majority of reinforcement learning applications in electricity market simulation use direct profit for the reward signal and may benefit from using measures of risk adjusted return.

In Moody and Saffell (2001) the recurrent reinforcement learning method from Moody et al. (1998) is contrasted with value function based methods. In addition to the Sharpe ratio, a Downside Deviation ratio is defined. Results from trading systems trained on half-hourly United States Dollar-Great British Pound foreign exchange rate data and, again, learning switching strategies between the S&P 500 index and T-Bills are presented. They show that the recurrent reinforcement learning method outperforms Q-learning in the S&P 500/T-Bill allocation problem. The authors observe also that the recurrent reinforcement learning method has a much simpler functional form, that the output, not being discrete, maps easily to real valued actions and that the algorithm is more robust to noise in the financial data and adapts quickly to non-stationary environments.

3.4.2 Grid Computing

In Vengerov (2008) a marketplace for computational resources in envisioned. The authors propose a market in which grid service suppliers offer to execute jobs submitted by customers for a price per CPU-hour. The problem formulation requires customers to request a quote for computing a job k for a time τ_k on n_k CPUs. The quote returned specifies a price P_k at which k would be charged and a delay time d_k for the job. The service provider's goal is to learn a policy for pricing quotes that maximises long term revenue when competing in a market with other providers. Price differentiation is implemented though provision of a standard service, priced at \$1/CPU-hour and a premium service a \$$P$/CPU-hour, with premium jobs prioritised over standard jobs. The state of the market environment is defined by the current expected delays in the standard and premium service classes and by $n_k\tau_k$: the product of the number of CPUs requested and the job execution time. The reward $r(s, a)$ for action a in state s is the total price paid for the job. The policy gradient method employed is a modified version of Williams' REINFORCE (Williams, 1992) where

$$Q(s_t, a_t) = \sum_{t=1}^{T} r(s_t, a_t) - \overline{r}_t \tag{3.17}$$

and $\overline{r}_t$ is the current average reward.

The authors recognise that their grid market model could be generalised to other multi-seller retail markets. The experimental results show that if all grid service providers simultaneously use the learning algorithm then the process converges to a Nash equilibrium. The results also showed that significant increases in profit were possible by offering both standard and premium services.

While this work applies policy gradient methods in a different domain, it shows how these methods can be used to set prices in a market and the author recognises the potential for the approach to be extended to other domains.

3.5 Summary

Agent-based simulation of electricity markets has been a consistently active field of research for more than a decade. Researchers around the world have sought to tackle important Electric Power Engineering problems including:

- Market power,
- Congestion management,
- Tacit collusions,
- Discriminatory vs. pay-as-bid pricing,
- Financial transmission rights, and
- Day ahead markets vs. bilateral trade.

Improvements in these areas have the potential to provide large financial benefits to societies.

There is a trend in the literature towards the use of more complex learning methods for participant behavioural representation and increasingly accurate electric power system models. Some of the more ambitious studies have used stylised models of national transmission systems for countires including the UK, Italy, Belgium and Germany. There have been previous attempts to compare learning methods for simulated electricity trade, but no concensus exisits as to which are most appropriate for particular applications.

Actions spaces are growing as researchers extend their studies to investigate energy business structures and the relationships between electricity, fuel and emission allowance markets. It seems that policy gradient reinforcement learning methods have not been previously used in electricity market simulation, but have been shown to work well in similar problems.

Chapter 4

Modelling Power Trade

The present chapter defines the model used in this thesis to simulate electric power trade. The first section describes how optimal power flow solutions are used to clear offers and bids submitted to a simulated power exchange auction. The second section defines how market participants are modelled as agents that use reinforcement learning algorithms to adjust their bidding behaviour. It also explains the modular structure of a multi-agent system that coordinates interactions between the auction model and market participants.

4.1 Electricity Market Model

A double-sided power exchange auction market model is used in this thesis to compare the electricity trading abilities of agents that utilise reinforcement learning algorithms. To determine the dispatch of generators, bespoke implementations of the optimal power flow formulations from MATPOWER (Zimmerman, 2010, §5) are used. Both the DC and AC formulations are utilised. The trade-offs between DC and AC models have been examined in Overbye, Cheng, and Sun (2004). DC models were found suitable for most nodal marginal price calculations and are considerably less computationally expensive. The AC optimal power flow formulation is used in experiments that require a more accurate electric power system representation. A class diagram in the Unified Modelling Language (UML) for the object-orientated power system model that is used to compute optimal power flow solutions is shown in Figure **??**.

As in MATPOWER (Zimmerman, 2010, p.26), generator active power, and optionally reactive power, output costs may be defined by convex n-segment piecewise linear cost functions

$$c^{(i)}(x) = m_i p + c_i \tag{4.1}$$

for $p_i \leq p \leq p_{i+1}$ with $i = 1, 2, \ldots n$ where $m_{i+1} \geq m_i$ and $p_{i+1} > p_i$, as diagramed in Figure **??** (Zimmerman, 2010, Figure5-3). Since these costs are non-differentiable, the constrained cost variable approach from (H. Wang, Murillo-Sanchez, Zimmerman, & Thomas, 2007) is used to make the optimisation problem smooth. For each generator i a helper cost variable y_i added to the vector of optimisation variables. The additional inequality constraints

$$y_i \geq m_{i,j}(p - p_j) + c_j, \quad j = 1 \ldots n \tag{4.2}$$

ensure that y_i lies on the epigraph[1] of $c^{(i)}(x)$. The objective of the optimal power flow problem used in the auction process becomes the minimisation of the sum of cost variables for all generators:

$$\min_{\theta, V_m, P_g, Q_g, y} \sum_{i=1}^{n_g} y_i \tag{4.3}$$

The extensions to the optimal power flow formulations defined in MATPOWER for user-defined cost functions and generator P-Q capability curves are not utilised.

4.1.1 Unit De-commitment

The optimal power flow formulations constrain generator set-points between upper and lower power limits. The output of expensive generators can be reduced to the lower limit, but they can not be completely shutdown. The online status of generators could be incorporated into the vector of optimisation variables, but as they are Boolean the problems would become mixed-integer non-linear programs which are typically very difficult to solve.

To compute a least cost commitment and dispatch the unit de-commitment algorithm from Zimmerman (2010, p.57) is used. Algorithm 1 shows how this involves shutting down the most expensive units until the minimum generation capacity is less than the total load capacity and then solving repeated optimal power flow problems with candidate generating units, that are at their minimum active power limit, deactivated. The lowest cost solution is returned when no further improvement can be made and no candidate generators remain.

[1]Informally, the epigraph of a function is a set of points lying on or above its graph.

Algorithm 1 Unit de-commitment

while $\sum P_g^{min} > \sum P_d$ **do**
    shutdown most expensive unit
end while
$f \leftarrow$ initial total system cost
repeat
    $c \leftarrow$ generators at P_{min}
    for g in c **do**
        $d \leftarrow$ true
        shutdown g
        $f' \leftarrow$ new total system cost
        if $f' < f$ **then**
            $f \leftarrow f'$
            $g_c \leftarrow g$
            $d \leftarrow$ false
        end if
        startup g
    end for
    shutdown g_c
until $d =$ true

4.1.2 Power Exchange

To simulate electric power trade a model is used in which agents representing market participants do not provide cost functions for the generators in their portfolio, but submit offers to sell and/or bids to buy blocks of active or reactive power. The offers/bids are submitted to a power exchange auction market model based on SmartMarket from Zimmerman (2010, p.92).

The clearing process begins by withholding offers/bids outwith maximum offer and minimum bid price limits, along with those specifying non-positive quantities. Valid offers/bids for each generator are then sorted into non-decreasing/non-increasing order and are converted into corresponding generator/dispatchable load capacities and piecewise linear cost functions. The newly configured units are used in a unit de-commitment optimal power flow problem, the solution of which holds generator set-points and nodal marginal prices which are used to determine the proportion of each offer/bid block that should be cleared and the cleared price for each.

A basic nodal marginal pricing scheme is used in which the price of each offer/bid is cleared at the value of the Lagrangian multiplier on the power balance constraint for the bus at which the associated generator is connected. Alternatively, a discriminatory pricing scheme may be used in which offer/bids are cleared at the price at which they were submitted (pay-as-bid). Cleared offers/bids are returned to the agents and used to determine revenue values from which each agent's earnings or losses are derived.

4.1.3 Auction Example

4.2 Multi-Agent System

Market participants are modelled with software agents from PyBrain that use reinforcement learning algorithms to adjust their behaviour (Schaul et al., 2010). Their interaction with the market is coordinated in multi-agent experiments, the structure of which is derived from PyBrain's single player design.

This section describes the environment of each agent, their tasks and the modules used for policy function approximation and storing state-action values in tables. The process by which each agent's policy is updated by a learning algorithm is explained and the sequence of interactions between multiple agents and the market is illustrated.

4.2.1 Environment

In each experiment, agents are endowed with a portfolio of generators from the electric power system model (See Figure **??**). As illustrated by the UML class diagram in Figure **??**, generators are contained within an agent's environment. The environment also holds an association to an instance of the auction market that allows the submission of offers/bids. Each environment is responsible for (i) returning a vector representation of its current state and (ii) accepting an action vector which transforms the environment into a new state. To facilitate testing of value function based and policy gradient learning methods, both discrete and continuous representations of an electric power trading environment are defined.

Discrete Environment

For operation with learning methods that use look-up tables to store state-action values, an environment with n_s discrete states and n_a discrete actions is defined. An agent can not observe offers/bids submitted by competitor agents, but is permitted to sense any aspect of the power system model. However, to ensure that the size of the environment state space is kept resonable the agent is limited to observing a demand forecast. Besides the actions of other agents, the total system demand is likely to be the most significant factor effecting the cleared quantity of its offers/bids. The initial demand at each bus P_{d0}, as defined in the original power system model, is assumed to be peak and the state space is divided into discrete steps of size P_d/n_s. As explained further in Chapter 6, the demand at each bus can follow a profile at each step t of the simulation. The environment computes the total system demand P_{dt} and returns an integer represntation of

a_i	m_1	m_2
1	0	0
2	0	10
3	0	20
4	10	0
5	10	10
6	10	20
7	20	0
8	20	10
9	20	20

Table 4.1: Example discrete action domain.

the state

$$s_t = \frac{P_{dt}}{P_{d0}/n_s} + 1. \tag{4.4}$$

To define the action space, a vector of percentage markups on marginal cost m_e and a vector of percentage markdowns on total capacity d_e is defined for each environment e along with a variable $n_o \in \mathbb{Z}^+$ which denotes the number of offers/bids to be submitted by the agent. A set of all unique permutations of markup and markdown for n_o offers/bids of length n_a is formed, from which the agent can select. The action vector that the discrete environment receives holds a single integer value, corresponding to the column index in the agent's action value table. The quantity and price for each offer/bid submitted to the market is taken from the vector of permutations using the a_t as the index. An example of the possible permuatations of 0, 10 and 20% markups for a portfolio of two generators is given in Table 4.1. It should be clear how quickly the number of possible actions can grow as the number of possible markups and the size of the portfolio increases.

Continuous Environment

A "continuous" environment for agent i may be configured for actions that specify price and optionally quantity. If $q_e^i = 0$ then the agent's action involves only price selection and the offer/bid quantity determined by the maximum rated capacity of the generator in question. The environment accepts a vector a_e of action values of length n_a if $q_e^i = 0$, otherwise a_e is of length $2n_a$. If $q_e^i = 0$, the i-th element of a_e is the offered/bid price in \$/MWh, where $i = 1, 2, \ldots n_{in}$. If $q_e^i = 1$, the j-th element of a_e is the offered/bid price in \$/MWh, where $j = 1, 3, 5, \ldots n_{in} - 1$ and the k-th element of a_e is the offered/bid quantity in MW where $j = 2, 4, 6, \ldots n_{in}$.

The action vector is converted into offers/bids and submitted to the market.

4.2.2 Task

To allow different goals (such a profit maximisation or the meeting some target level for plant utilisation) to be associated with a single type of environment, an agent does not interact directly with the environmetn, but is paired with a particular *task*. A task defines the reward returned to the agent and thus defines the agent's purpose. For all experiments in this thesis the goal of each agent is to maximise financial profit and the rewards are thus defined as the sum of earnings from the previous period t as determined by the revenue from the market and any incurred costs. As explained in Section 3.4.1, utilising some measure of risk adjusted return might be of interest in the context of simulated electricty trade and this would simply involve the definition of a new task without any need for modification of the environment.

Sensor data from the environment is filtered according to the task being performed. Agents with value-function learning methods use a table to store state-action values, with one row per environment state. Thus, observations consist of a single value s_v, where $s_v \leq n_s$ and $s_v \in \mathbb{Z}^+$.

Agents with policy-gradient learning methods approximate their policy functions using artificial neural networks that are presented with input vector w of length n_i where $w_i < n_i$ and $w_i \in \mathbb{R}$. To condition the environment state before input to the connectionist system, where possible, each sensor i in the state vector s is associated with a minimum value $s_{i,min}$ and a maximum value $s_{i,max}$. The state vector is normalised to a vector:

$$s_c = 2\left(\frac{s - s_{min}}{s_{max} - s_{min}}\right) - 1 \tag{4.5}$$

such that $-1 \leq s_c^i \leq 1$.

The output from the policy function approximator, a_c, is denormalised using minimum and maximum action limits, a_{min} and a_{max} respectively, giving an action vector

$$a = \left(\frac{a+1}{2}\right)(a_{max} - a_{min}) + a_{min} \tag{4.6}$$

with valid values for price (and optionally quantity) that may be used to form offers/bids.

4.2.3 Agent

Each agent i is defined as an entity capable of producing an action a_i based on previous observations of its environment s_i, where a_i and s_i are vectors of arbitrary length. In PyBrain each agent is associated with a *module*, a *learner*, a *dataset* and an *explorer*. The module is used to determine the agent's policy for action selection and returns an action vector a_m when activated with observation s_t.

When using a value-function method the module is a $n_s \times n_a$ table, where n_s is the total number of states and n_a is the total number of actions.

$$\begin{array}{c c} & \begin{array}{cccc} a_0 & a_1 & \quad & a_n \end{array} \\ \begin{array}{c} s_0 \\ s_1 \\ \\ s_n \end{array} & \begin{bmatrix} v_{1,1} & v_{1,2} & \cdots & v_{1,m} \\ v_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ v_{n,1} & \cdots & \cdots & v_{n,m} \end{bmatrix} \end{array} \tag{4.7}$$

When using a policy gradient method, the module is a multi-layer feedforward artificial neural network.

The learner can be any reinforcement learning algorithm that modifies the values/parameters of the module to increase expected future reward. The dataset stores state-action-reward triples for each interaction between the agent and its environment. The stored history is used by value-function learners when computing updates to the table values. Policy gradient learners search directly in the space of the policy network parameters.

Each learner has an association with an explorer that returns an explorative action a_e when activated with the current state s_t and action a_m from the module.

4.2.4 Simulation Event Sequence

Each experiment consists one or more agent-task pairs. At the beginning of each simulation step (trading period) the market is initialised and all existing offers/bids are removed. From each task-agent tuple (T, A) an observation s_t is retrieved from T and integrated into agent A. When an action is requested from A its module is activated with s_t and the action a_e is returned. Action a_e is performed on the environment associated with task T. This is the process that involves the submission of offer/bids to the market. Figure **??** provides a UML sequence diagram that illustrates the process of performing an action and Figure **??** shows the class associations of an experiment.

When all actions have been performed the offers/bids are cleared by the market using the solution of an optimal power flow problem. Each task is requested to return a reward r_t. The cleared offers/bids associated with the generators in the task's environment are retrieved from the market and r_t is computed from the difference between revenue and cost values.

$$r_t = \text{revenue} - (c_{fixed} + c_{variable}) \tag{4.8}$$

The reward r_t is given to the associated agent and the value is stored, along with the previous state s_t and selected action a_e, under a new sample is the dataset. The reward process is illustrated in a UML sequence diagram in Figure **??**.

Each agent learns from its actions using r_t, at which point the values/parameters of the module associated with the agent is updated according to the output of the learner's algorithm. Each agent is then reset and the history of states, actions and rewards is cleared. The learning process is illustrated by the UML sequence diagram in Figure **??**.

All of this constitutes one step of the simulation and the process is repeated until a set number of steps are complete.

4.3 Summary

The power exchange auction market model defined in this chapter provides a layer of abstraction over the underlying optimal power flow problem and presents agents with a virtual interface for selling and buying power. The modular nature of the simulation framework described allows the type of learning algorithm, the policy function approximator, the exploration technique or the task to be changed easily. The framework can simulate competitive electric power trade using any conventional bus-branch power system model, requiring little configuration, but provides the facility to adjust all of the main aspects of the simulation. The modular framework and its support for easy configuration is intended to allow transparent comparison of learning methods in the domain of electricity trade under a number of different metrics.

Chapter 5

Nash Equilibrium Analysis

This chapter examines the convergence to a Nash equilibrium[1] of agents comptet-ing with portfolios of generating plant. Value function based and policy gradient reinforcement learning algorithms are compared in their ability to converge on an optimal policy using a six bus electric power system model.

5.1 Introduction

To the best of the author's knowledge, this thesis presents the first case of policy gradient reinforcement learning methods being applied to electricity trading problems. As a first step it is necessary to confirm that when using these methods, a system of multiple agents will converge to the same Nash equilibrium that conventional closed-form simulation techniques produce.

This is the same approach by Krause et al. (2006) before performing the study of congestion management techniques that is reviewed in Section 3.2.2. Nash equilibria can be difficult to determine in complex systems so the experiment presented here utilises a model simple enough that it can be determined through exhaustive search.

By observing the actions taken and the reward received by each agent over the initial simulation periods it is possible to compare different configurations of the algorithms in their speed of convergence to an optimal policy. In the following sections the objectives of this experiment are explictly defined, the setup of the simulations is explained and simulation results with discussion and critical analysis are provided.

[1]Informally, a Nash equlibrium is a point at which no player is motivated to deviate from its strategy as it would result in lower gain.

5.2 Aims and Objectives

Some elements of this experiment are very similar to those presented in Krause et al. (2006) and the initial aim is to reproduce those results. The additional objectives are to show that:

- Policy gradient methods converge to the same Nash equilibrium as value function based methods.
- The differences in speed of convergence to an optimal policy between the learning methods.

Meeting these objectives aims to provide a basis for more complicated experiments that are less intuitively tractable.

5.3 Method of Simulation

Learning methods are compared in this experiment by repeating the same simulation with different algorithms used by the agents. An alternative might be to use a combination of methods in the same simulation, but the approach used here is intended to be an extension of the work in Krause et al. (2006).

Each simulation uses the six bus electric power system model adapted from Wood and Wollenberg (1996, pp. 104, 112, 119, 123-124, 549). The six buses are connected by eleven transmission lines at 230kV. It contains three generating units with a total capacity of 440MW and loads at three locations, each of 70MW. The connectivity of the branches and the locations of the generators and loads is shown in Figure **??**. Data for the power system model is provided in Appendix B.1 and is distributed with the software developed for this thesis (See Appendix A.9).

Two sets of generator operating costs are defined to create two different equilibria for investigation. The first set is listed in Table 5.1. It defines two low cost generators that can not offer a price greater than the marginal cost of the most expensive generator when the maximum markup is applied. The second set is listed in Table 5.2 and narrows the cost differences such that offer prices overlap and may exceed the marginal cost of the most expensive generator.

No load profile is defined, the system load is assumed to be peak for all simulation periods, so only one system state is defined for the value function based algorithms. The minimum operating point, P^{min}, for all generators is

Gen	C_{down}	a	b	c
1	100	0.0	4.0	200.0
2	100	0.0	3.0	200.0
3	100	0.0	6.0	200.0

Table 5.1: Generator cost configuration 1 for 6-bus case.

Gen	C_{down}	a	b	c
1	100	0.0	5.1	200.0
2	100	0.0	4.5	200.0
3	100	0.0	6.0	200.0

Table 5.2: Generator cost configuration 2 for 6-bus case.

made to be zero so as to simplify the experiment and avoid the need to use the unit decommitment algorithm defined in Section 4.1.1.

The maximum capacity for the most expensive generator P_3^{max}=220MW such that it may supply almost all of the load if it is dispatched. This generator is associated with a passive agent that always offers a maginal cost. For the other generators P_1^{max}=110MW and P_2^{max}=110MW. These two generators are each associated with am active learning agent whose activity in the market is restricted to one offer of maximum capacity in each period, at a price representing a markup of between 0 and 30% on marginal cost. Value function based methods are restricted to discrete markup steps of 10%, giving possible markup actions of 0, 10, 20 and 30%. The market price cap is set such that it is never reached by any markup and does not complicate the experiment. Discrimnatory pricing (pay-as-bid) is used in order to provide a clearer reward signal to agents with low cost generators.

The learning methods compared are Q-learning, ENAC, REINFORCE and the variant Roth-Erev technique. For Q-learning $\alpha = 0.3$, $\gamma = 0.99$ and ϵ-greedy action selection is used with $\epsilon = 0.9$ and $d = 0.97$. For Roth-Erev learning $\epsilon = 0.55$, $\phi = 0.3$ and Boltzmann action selection is used with $\tau = 100$ and $d = 0.98$. Both REINFORCE and ENAC use a three-layer neural network with one linear input node, two hidden tanh nodes, one output tanh node and bias nodes in the hidden and output layers.

As in Krause et al. (2006), the point of Nash equilibrium is established by computing each agent's reward for all possible combinations of markup. The rewards for Agent 1 and Agent 2 under cost configuration 1 are given in Table 5.3. The Nash equilibrium points are marked with a *. It shows that the optimal

		G_1							
		0.0%		10.0%		20.0%		30.0%	
		r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2
G_2	0.0%	0.0	0.0	40.0	0.0	80.0	0.0	120.0	0.0
	10.0%	0.0	33.0	40.0	33.0	80.0	33.0	120.0	33.0
	20.0%	0.0	66.0	40.0	66.0	80.0	66.0	120.0	66.0
	30.0%	0.0	99.0	40.0	99.0	80.0	99.0	120.0*	99.0*

Table 5.3: Agent rewards under cost configuration 1

		G_1							
		0.0%		10.0%		20.0%		30.0%	
		r_1	r_2	r_1	r_2	r_1	r_2	r_1	r_2
G_2	0.0%	0.0	0.0	51.0	0.0	0.0	0.0	0.0	0.0
	10.0%	0.0	49.5	51.0	49.5	0.0	49.5	0.0	49.5
	20.0%	0.0	92.2	51.0	99.0	0.0	99.0	0.0	99.0
	30.0%	0.0	126.8	54.8*	138.4*	0.0	148.5	0.0	148.5

Table 5.4: Agent rewards under cost configuration 2

policy for each agent is to apply the maximum markup to each offer as this never results in thier generators failing to be dispatched. The rewards under cost configuration 2 are given in Table 5.4. It shows that the optimal point occurs when Agent 2 applies its maximum markup and Agent 1 offers a price just below the marginal cost of the passive agent's generator.

5.4 Simulation Results

Each action taken by an agent and the consequent reward is recorded for each simulation. Values are averaged over the 10 simulation runs and standard deviations calculated using the formula

$$SD = \sum_{i=0}^{N} \frac{(x_i - m)^2}{N - 1} \tag{5.1}$$

where x_i is the action or reward value in simulation i of N simulation runs and m is the mean of the values.

Figure 5.1 plots the average markup on marginal cost and the standard deviation over the 10 simulation runs for Agent 1 under the first price configuration using the variant Roth-Erev, Q-learning, REINFORCE and ENAC learning methods. The second y-axis in each plot gives the value of the exploration parameter for each method. Figure 5.2 plots the same quantities for Agent 2. Plots of reward are not given as generator prices and the market are configured such that an agent's reward is directly proportional to its action.

Figures 5.3 and 5.4 plot the average markup for Agent 1 and Agent 2, respectively, under the second price configuration and again for the variant Roth-Erev, Q-learning, REINFORCE and ENAC learning methods. Figures 5.5 and 5.6 plot the associated average *rewards* for Agent 1 and Agent 2. Again the standard deviation and exploration parameter values are plotted. The plots are vertically aligned and have equal x-axis limits to aid algorithm comparison.

5.5 Discussion and Critical Analysis

Under the first generator price configuration the agents face a simple control task and receive a clear reward signal that is directly proportional to their markup action. The results show that all of the methods consistently converge to the point of Nash equilibrium. A multitude of parameter and nerual network structure variations could be investigated and a sea of similar plots would be produced. The author's experience is that the speed of convergence is largely determined by the rate at which the exploration parameter value is reduced. The policy gradient methods are sensitive to high learning rate parameter values, but make only very small policy adjustments if this parameter is set too low. All of the plots for REINFORCE and ENAC show that the methods only converge to a stable policy if the exploration parameter σ is manually reduced to below approximately -2.

The second pricing configuration provides a more challenging control problem in which there is some interdependence between the agent's rewards and where Agent 1 must learn to undercut the passive agent. The results show that the variant Roth-Erev and Q-learning methods both consistantly learn their optimal policy and converge to the Nash equilibrium. It should be noted that Agent 1 can markup its marginal price by slightly more than 10% and still undercut the passive agent, but these methods are restricted to discrete actions.

When using REINFORCE or ENAC, Agent 2 tends also to learn to maximise its markup, but less consistantly. Agent 1 typically learns to undercut the passive agent, but does not converge to a consistant value. The problem is similar to the cliff-edge walking problems often used as benchmarks in reinforcement learning research and may be difficult to approximate a policy for using a small number of tanh functions. It may be possible to improve the performance of these agents through more educated policy function approximator design, but these methods

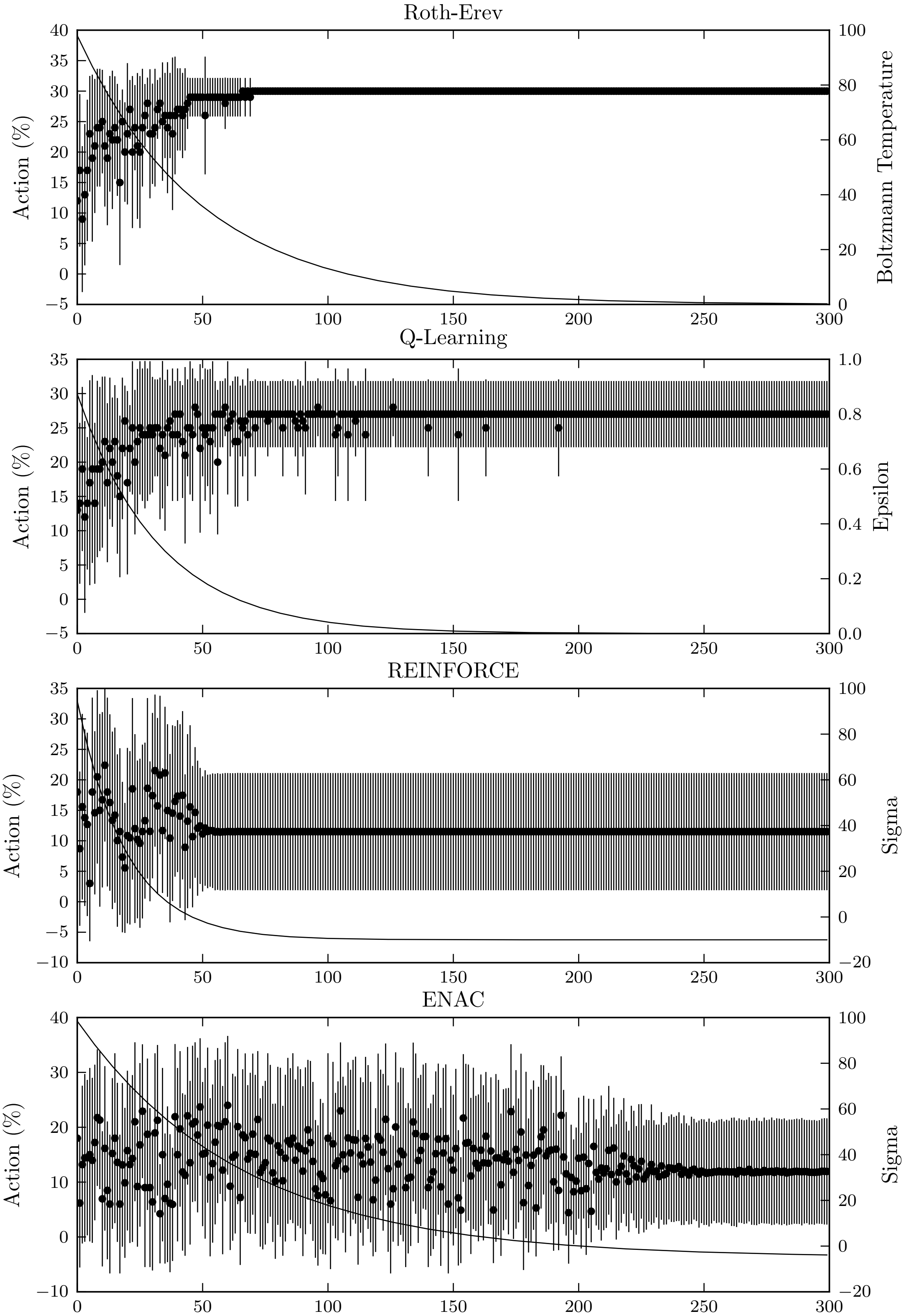

Figure 5.1: Average markup for agent 1 and standard deviation over 10 runs.

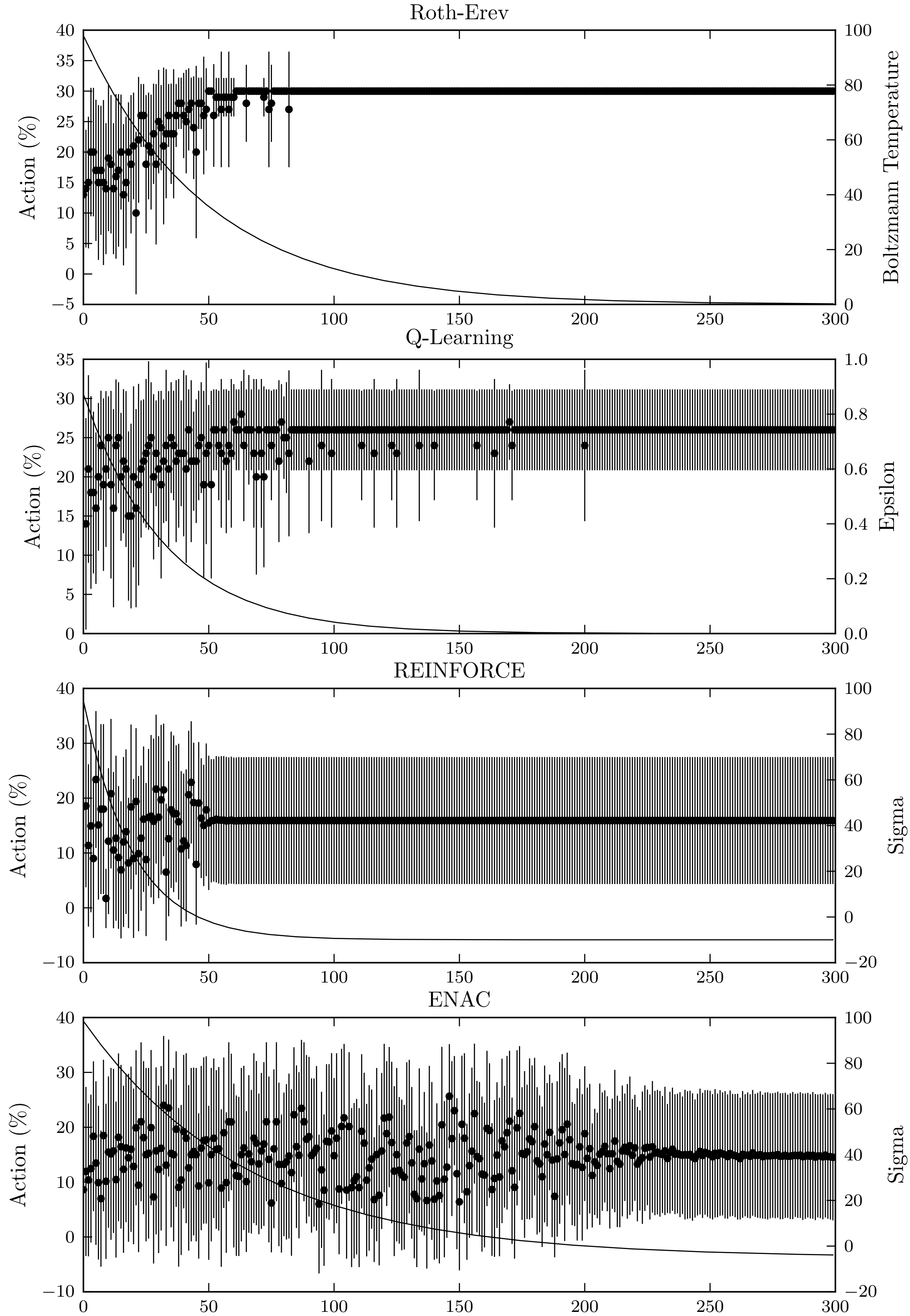

Figure 5.2: Average markup for agent 2 and standard deviation over 10 runs.

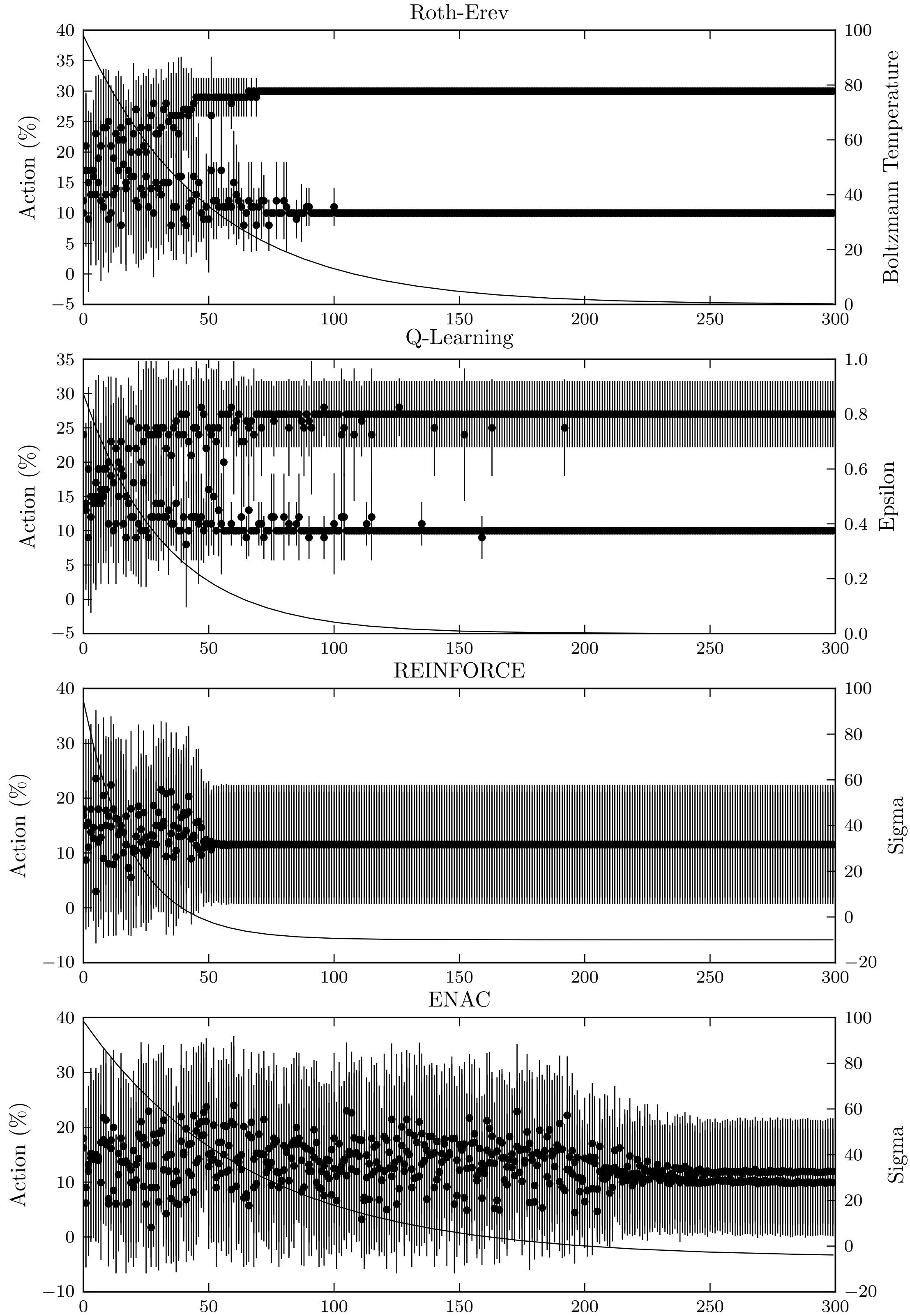

Figure 5.3: Average markup for agent 1 and standard deviation.

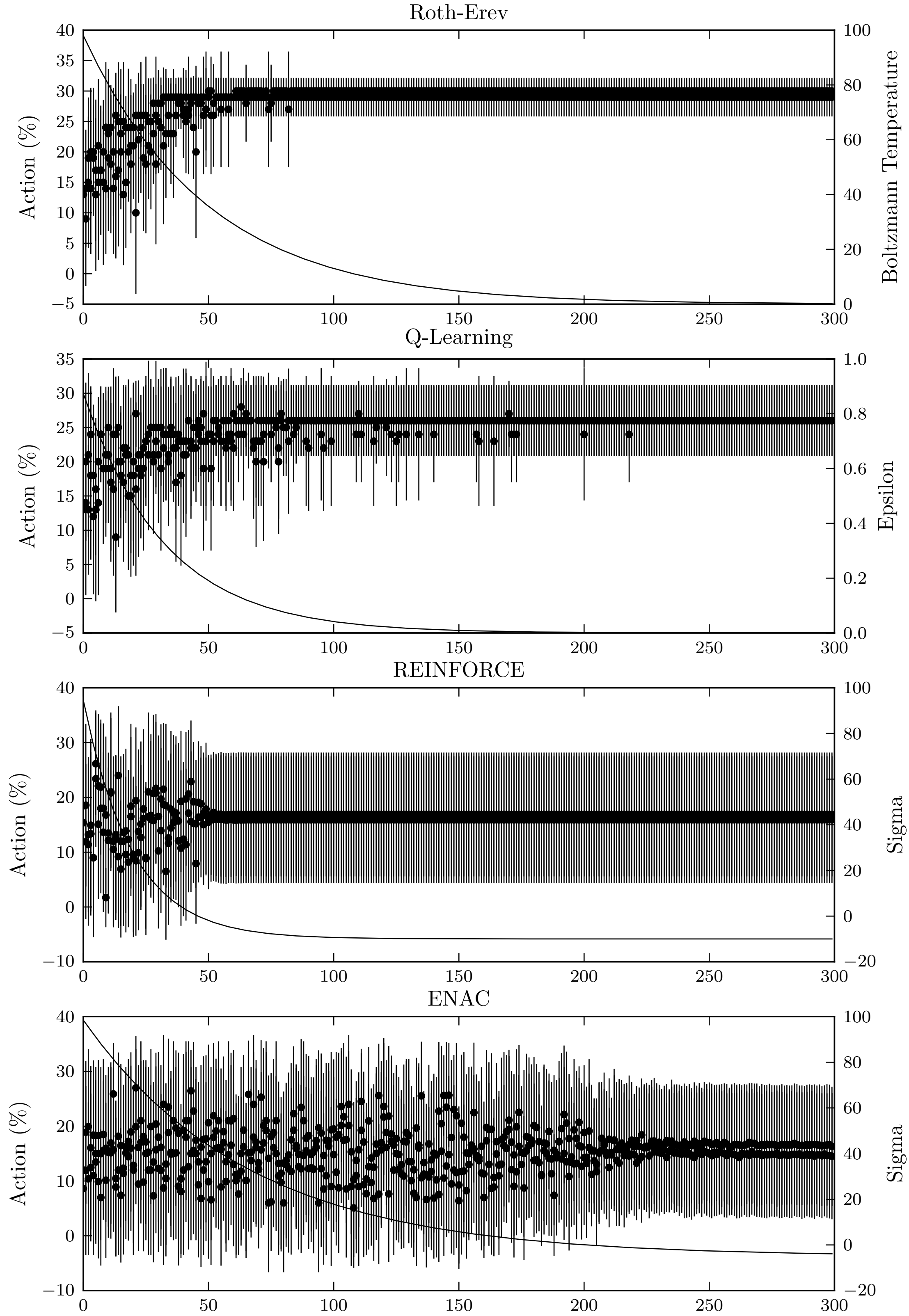

Figure 5.4: Average markup for agent 2 and standard deviation.

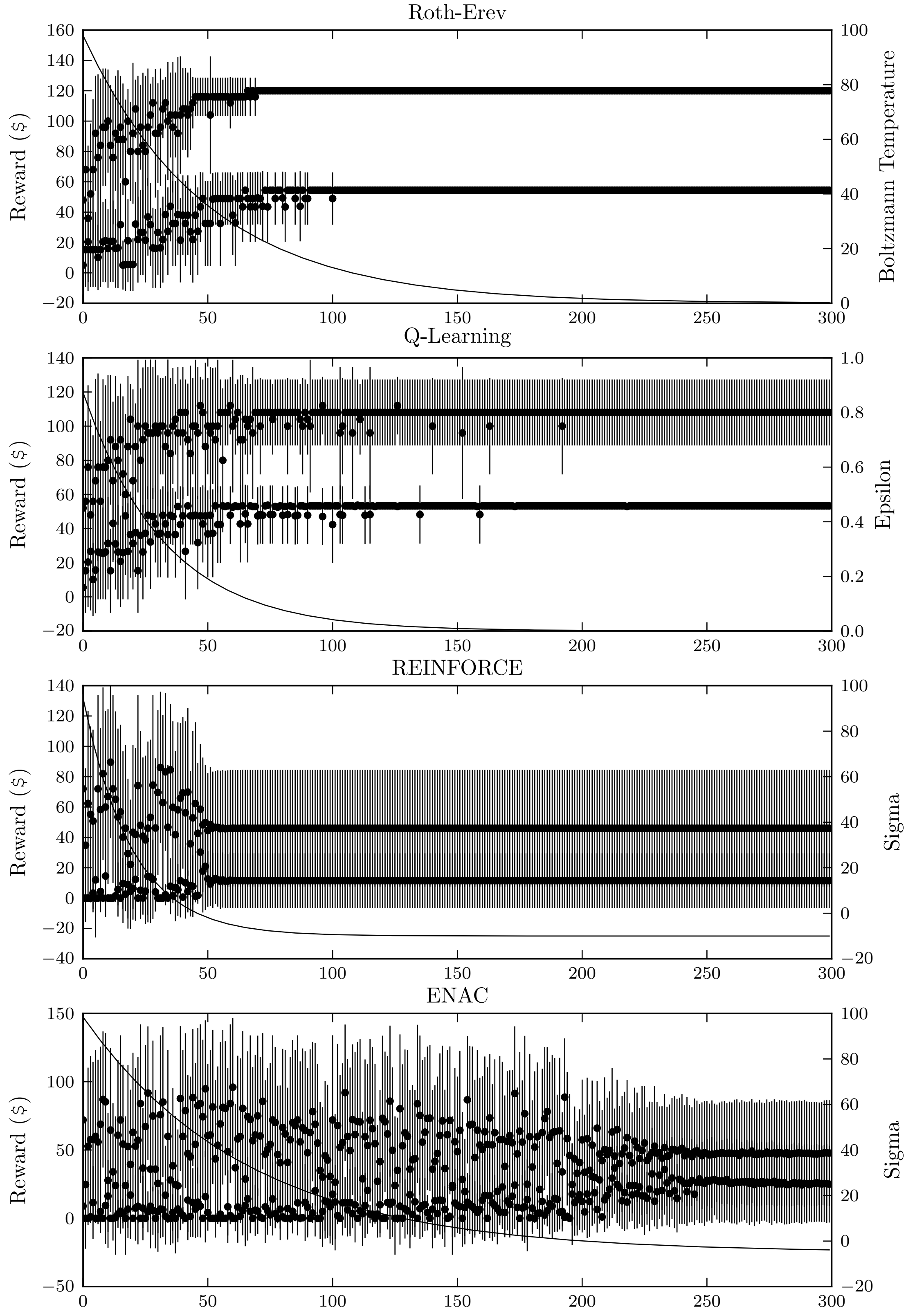

Figure 5.5: Average reward for agent 1 and standard deviation.

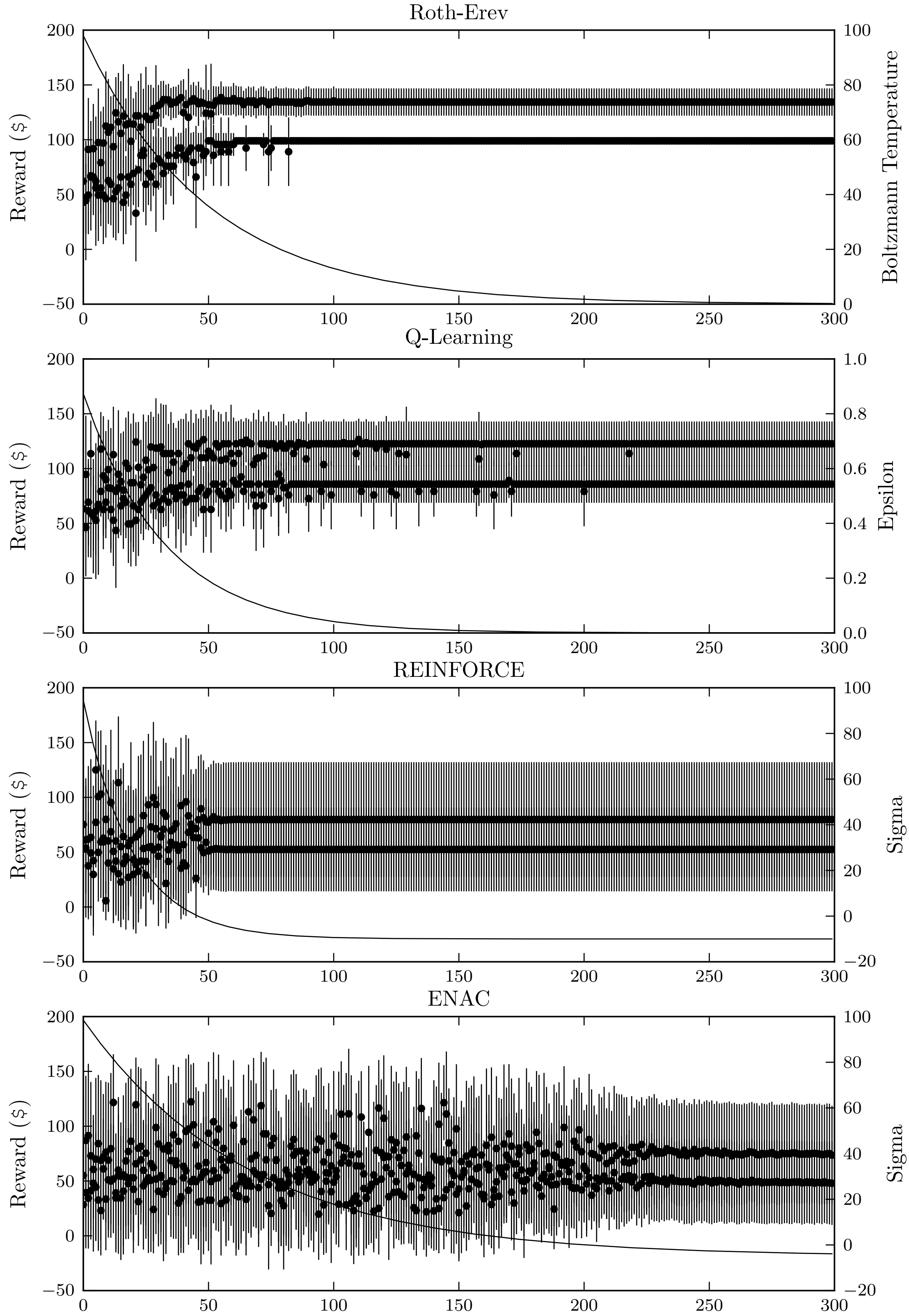

Figure 5.6: Average reward for agent 2 and standard deviation.

are not really intended for operation in such simple environments.

This experiment confirms the convergence to a Nash equilibrium of the Q-learning methods that is published in Krause et al. (2006) and, to a degree, extends the conclusion to policy gradient methods. The results show that while these methods do converge to the same or similar policies as the Q-learning and Roth-Erev methods, they do not exhibit the same level of consistency. Value function based methods or the Roth-Erev method may be the most suitable choice of algorithm in the simple electricity market simulations typically found in the literature.

5.6 Summary

The simulations conducted here do not exploit any of the abilities of policy gradient methods to utilise multi-dimensional continuous state information and their behaviour in more complex environments must be examined.

Chapter 6

System Constraint Exploitation

This chapter explores the exploitation of constraints in electric power system models by agents whose behaviour is determined by reinforcement learning algorithms. Value function based and policy gradient methods are compared using the 24-bus IEEE Reliability Test System with dynamic loads.

6.1 Introduction

Having explored the basic properties of four learning methods in Chapter 5, this experiment examines them under a more complex dynamic scenario. The experiment explores the multi-dimensional, continuous state space handling abilities of policy gradient methods in the context of *learning to trade power*.

Control of a portfolio of generators using continous sensor data from simulations of the IEEE Reliability Test System (Application of Probability Methods Subcommittee, 1979) is investigated. The system is constrained by bus voltage, branch flow and generator capacity limits as the system demand cycles daily over the course of a year. By observing the actions taken and the rewards received by the agents during these periods it is examined if policy gradient methods can successfully observe and exploit the constraints.

6.2 Aims and Objectives

This experiment aims to compare the operation of learning methods in dynamic electric power system environments. Specifically, the objectives are to determine:

- If policy gradient methods can be used to achieve greater profit under dynamic loading conditions.

Code	C_{up}	a	b	c	Type
U12	1500	0.32841	56.564	86.385	Oil
U20	1500	0.0	130.0	400.685	Oil
U50	1500	0.0	0.001	0.001	Hydro
U76	1500	0.01414	16.0811	212.308	Coal
U100	1500	0.05267	43.6615	781.521	Oil
U155	1500	0.00834	12.3883	382.239	Coal
U197	1500	0.00717	48.5804	832.758	Oil
U350	1500	0.00490	11.8495	665.109	Coal
U400	1500	0.00021	4.4231	395.375	Nuclear

Table 6.1: Cost parameters IEEE RTS generator types.

- The value of using AC optimal power flow formulations in agent base electricity market simulation.

Meeting these objectives aims to demonstrate the value of policy gradient methods in electricity market participant modelling and determine if they warrant further research in this domain.

6.3 Method of Simulation

In this experiment learning methods are compared by repeating the simulations of competitive electricity trade with different types of algorithm used by the participating agents. Some simplification of the state and action domains for the value function based methods is required, but the portfolios of generation and load profiles are constant.

The IEEE Reliability Test System provides the power system model and load profiles used in each simulation. The model has 24 bus locations, connected by 32 transmission lines, 4 transformers and 2 underground cables. The transformers tie together two system areas at 230kV and 138kV. The model has 32 generators of 9 different types (See Table 6.1) with a total capacity of 3.45GW and load at 17 locations, totalling 2.85GW. Generator costs are quadratic functions of output, defined by the parameters in Table 6.1. Figure 6.1 plots the cost functions for each type of generator over their production range and illustrates their categorisation by fuel type. Data for the model is provided in Appendix B.2 and the connectivity of branches and the location of generators and loads is illustrated in Figure **??**.

The generating stock is divided into 5 portfolios, as listed in Table 6.2, that are each endowed to a learning agent. The synchronous generator is associated with a passive agent that always offers at marginal cost i.e. \$/MWh 0.0. Markups

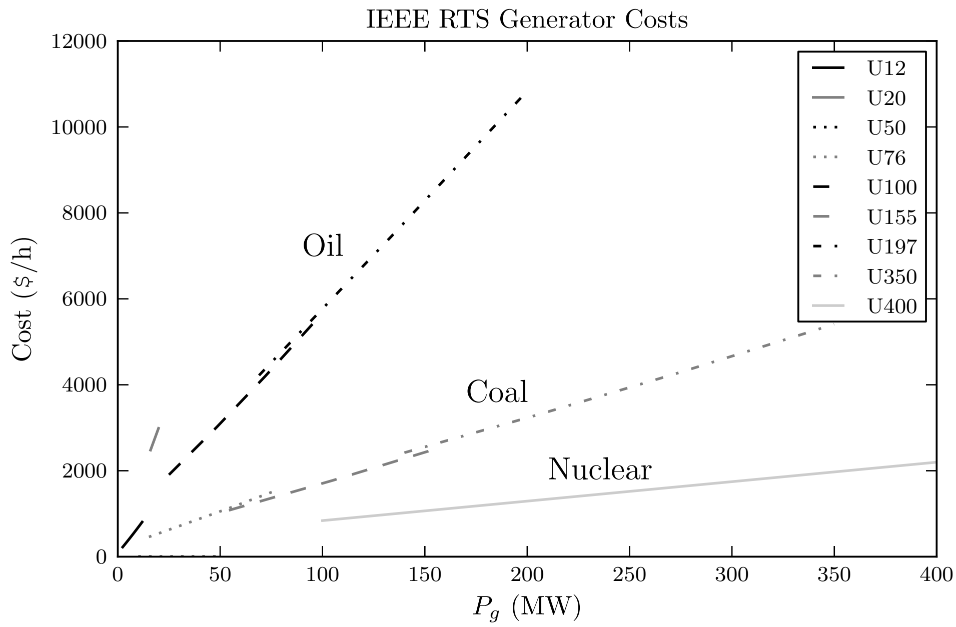

Figure 6.1: Generator cost functions for the IEEE Reliability Test System

of offer price are restricted a maximum of 30% and discrete markup steps of 10% are defined for value function based methods.

6.4 Simulation Results

6.5 Discussion and Critical Analysis

6.6 Summary

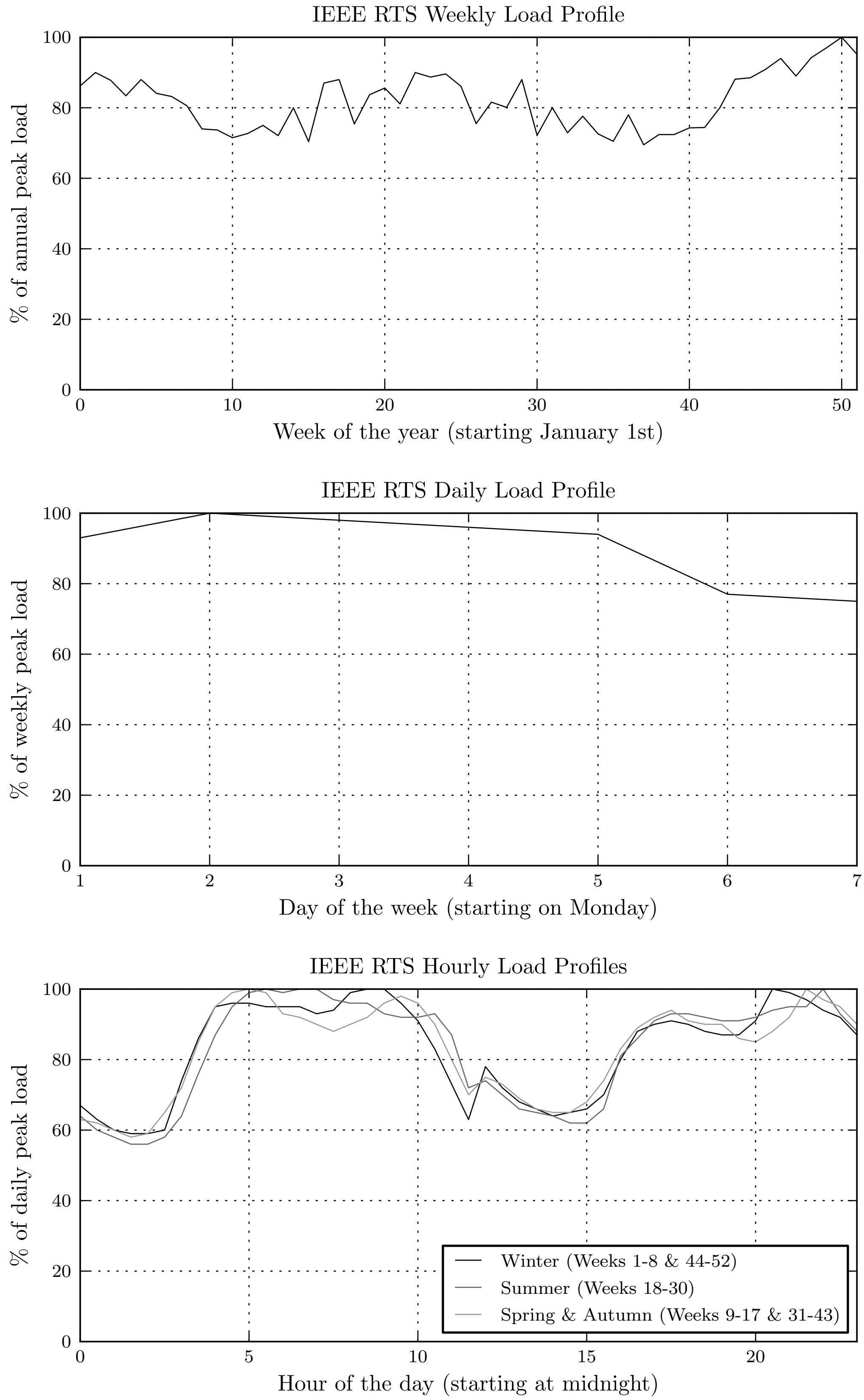

Figure 6.2: Hourly, daily and weekly load profile plots from the IEEE Reliability Test System

Agent	Type	Buses
A_1	U20	1,1,2,2
	U76	1,1,2,2
	U100	7,7,7
A_2	U155	23,23
	U197	13,13,13
	U300	23
A_3	U12	15,15,15,15,15
	U155	15,16
A_4	U50	22,22,22,22,22,22
	U400	18,21
A_5	Sync. Cond.	14

Table 6.2: Agent portfolios by type and location.

Chapter 7

Conclusions and Further Work

This chapter summarizes the conclusions that can be drawn from the results that are presented in this thesis and presents several ideas for further development of the contributions that have been made.

7.1 Further Work

This section introduces some relatively new learning algorithms that have also been developed for operation in continuous multi-dimensional domains and might also be used for simulating electricity trade. Also, two new reinforcement learning problem formulations are defined that concern two highly pertinent issues in electric power Engineering. Finally, it explained is how data from National Grid Ltd. could be used in practical simulations of the UK electricity market and some of the posibilities that AC optimal power flow brings to to electric power market simulation are explored.

7.1.1 Alternative Learning Algorithms

This thesis has concentrated on traditional value function based and policy gradient reinforcement learning methods. However, research in the field of Artificial Intelligence is highly active and there have been some interesting new learning algorithms presented recently that might also be examined in the domain of electric power trade.

Riedmiller (2005) presented Neuro Fitted Q Iteration (NFQ) algorithms that attempt to overcome many of the problems experienced when implementing Q-learning methods with value function approximation using neural networks. They store all transition experiences and perform off-line updates using supervised learning techniques such as RProp (Riedmiller & Braun, 1993). The method has been shown to be robust against parmeterisation and to learn quickly in standard benchmark tests and in real-world applications (Kietzmann & Riedmiller, 2009).

The GQ(λ) algorithm by Maei and Sutton (2010) is another extension of the Q-learning method. Convergence guarantees have been shown and the scaling properties suggest it is suitable for large-scale reinforcement learning applications. Software implementations of GQ(λ) are reportedly in development and due to be made available soon.

Four new Natural Actor-Critic algorithms are presented in Bhatnagar, Sutton, Ghavamzadeh, and Lee (2009). As with the ENAC algorithm by Peters and Schaal (2008), these all utilise linear function approximation techniques, making them suitable for large-scale applications of reinforcement learning. Three of the algorithms are extensions to ENAC, but are fully incremental: the gradient computation is never reset while the policy is updated at every simulation step. The authors state a need to assess the ultimate utility of these algorithms through application in real-world problems.

This thesis provides a framework that would allow implementations of these interesting new algorithms to be assessed and used to examine aspects of electricity markets.

7.1.2 Learning to Optimise Power Flow

Two important problems in electric power Engineering, to which to the application of advanced reinforcement learning algorithms would be of value, are:

- System optimisation, close to real-time, such that sufficient reserve is allocated to ensure acceptable system security while costs are minimised according to the outcome of the electricity markets, and
- Capital investment planning, both by system operators needing to expand transmission capacity and energy companies wanting to develop new generating plant.

As explained in Section 2.3.2, the objective of the classical optimal power flow problem is to find generator set-points that allow all system and plant constraints to be satisfied while the total system cost is minimised. Interior point methods are possibly the most robust technique for finding solutions, but the problem may also be formulated as a continuous reinforcement learning task that could be learned by a system operator agent. To illustrate the concept, this section presents a preliminary formulation of such a task and demonstrates how a system operator agent can use policy gradient methods to learn to optimise power flow.

System Operator Task and Environment

The state of the system operator agent's environment is defined simply as a demand forecast. That is, the environment returns a vector of active power demand at all system buses. The initial demand is assumed to be peak and is used to normalise the values of the sensor vector to be between -1 and $+1$ before input the the mutil-layer perceptron used for policy function approximation. Simulations are divided into episodes (days) over which the demand at each bus follows the profile shown in Figure X.

The agent's action and the output of the policy function approximator is a vector of the active power set-points of all generators, excluding the generator at the system slack bus. The set-points are bound by the generator's minimum and maximum rated capacity. These bounds are used to denormalise the output values from the final Sigmoid layer (that are between -1 and $+1$) to give valid set-point values.

The new generator set-points are used to form an AC power flow problem that is solved using Newton's method (Tinney & Hart, 1967). The power flow solution determines the complex voltage at each bus, the branch power flows and losses, the reactive power output of the generators and the active power output of the slack bus generator.

The reward is defined as the negative of the sum of all generator costs. The negative of the costs must used since the learning methods attempt to maximise reward and the objective is to minimise cost. The power flow solution does not satisfy system constraints, such as voltage limits or generators reactive power limits and penalty costs must be applied to the reward so the agent learns to obey them.

Slack Bus Generator Control

This initial proof of concept attempts only to enforce the set-point limit on the slack bus generator. The reward is updated according to

$$r = \begin{cases} r + \phi(P_{slack} - P_{max}), & P_{slack} > P_{max} \\ r, & \text{otherwise} \end{cases} \tag{7.1}$$

The six bus network model described in Chapter 5 is used with the coefficients of the generator's quadratic cost functions given in Table X. Learning is conducted in batch mode, with 7 episodes conducted before the parameters of the policy are updated. The system operator agent uses the ENAC learning method with RProp gradient descent and an initial value of $\sigma = 50$, which is reduced after each simulated week according to $\sigma_{i+1} = 0.5\sigma_i - 2$.

Figure X shows the average output of each generator over 52 simulated weeks of control. Plotted in Figure Y is the average total system cost for each week along with theoretically optimal values as calcualted by DC and AC optimal power flow solvers. The results show that, with suitable initial experimentation, the agent learns to dispatch the generators in the most economically efficient manner while controlling the slack bus generators to valid output levels. The average total system cost converges to slightly less that of the optimal solution due to the disregard for other constraints, particularly the branch flow constraint between buses 2 and 4 which is binding at times of peak load.

7.1.3 Learning to Plan Investment

7.1.4 UK Transmission System

Some of the more ambitious agent-based electricity market simulations have used stylised models of national transmission systems (Rastegar et al., 2009; Weidlich & Veit, 2006). This work has often been motivated by recent or expected changes to the arrangements in the associated regions. The drop in oil prices around the time of the global economic crash in 2008 was not reflected quickly in energy prices and this amplified concerns over liquiduty levels in the UK electricity markets. Ofgem found competition to be sufficient[ref], but the concerns persists and the market arrangements have been re-examined under Project Discovery[ref].

Several of the the UK's largest power stations are due to be decommissioned around 2015 in accordance with EU Large Combustion Plant Directive[ref]. The ability of the market to sufficiently incentivise new investment in generation that will cover the resulting shortfall is in question. The concern extends to the need for long-term investment in new nuclear power plant that is deemed neccessary for the UK to meet the legally binding obligations, made in the Climate Change Bill, to cut greenhouse gas emissions by 80% by 2050, compared to 1990 levels. Calls have been made for a radical overhaul of the exising arrangments and for a more interventionist strategy from the government.

Ofgem's project discovery makes many assumptions in its model of the UK energy industry. Future examinations could be enhanced by the advanced participant behavioural models and accurate electric power system simulations presented in this thesis. Figure X illustrates the UK transmission system, detailed data for which is providedin National Electricity Transmission System Operator (2010). This data has been converted into PSS/E version 30 raw file format and is distributed with the rest of the code developed for this thesis (See Appendix A.9).

The problem is currently too computationally expensive to be solved repeated in an agent-based simulation, but the profiling output in Table Y shows that much of the execution time is spent constructing matrices in linked list format. Significant improvements in speed should be possible through more efficient construction of the Hessian matrices in the AC optimal power flow solver. Agent-based simulation lends itself to parallelisation and the artificial neural networks could be processed in multiple threads on multi-core processors or on distributed memory architectures.

7.1.5 AC Optimal Power Flow

To the best of the author's knowledge this is the first application of security constrained optimal power flow in agent-based electricity market simulation. AC optimal power flow is more difficult to implement and more computationally expensive due to the non-linear sets of constraints involved. The additional complexity does not always add sufficient value. However, the option to use an AC formulation offers some interesting possibilities for further work.

The inclusion of the cost associated with producing (or absorbing) reactive power in the objective function of an AC optimal power flow problem means that parallel auctions for voltage support may also be included in simulations. This could be open to agents associated with reactive compensation equipment such as that commonly needed for wind farm developments. Reactive power markets have traditionally been largely academic, but as the UK becomes more dependant upon offshore wind power the topic could become of increasing interest.

Bus voltage magnitudes are not all assumed to be 1 per-unit in AC optimal power flow problems, but are part of the vector of optimisation variables. Generator reactive power and bus voltage constraint limits are usually determined by system security constraint requirements. The policy gradient methods used in this thesis can operate with large state spaces that may include information on these constraints. The additional constraints that are part of an AC formulation represent further opportunities for agents to exploit system conditions and receive greater reward. However, bus voltages are typically regulated by tap changing transformers and this is a difficult feature to implement. Tap positions are typically restricted to discrete intervals, making the optimisation a mixed integer problem and thus much more difficult to solve. A formulation extended to includ automatic tap changing might also implement variable phase shifting transformers. These offer a degree of control in directing power flows and the formulation would allow advanced constraint management schemes to be researched.

7.1.6 Multi-Market Simulation

Policy gradient method's superior use of sensory data and their ability to operate in large action domains opens opportunities for more detailed study of inter-market relationships. The global economy is a holistic system of systems and the anaylsis of markets independently must be of limited value. Recent agent-based electricity market simulations studies have investigated the interaction between electricity, gas and emission allowances markets (Kienzle et al., 2007; J. Wang et al., 2009). Non-linear models [ref] have been published for gas flows in pipelines such as those of the UK gas network. The information of the UK gas network provided in National Electricity Transmission System Operator (2010) is relatively limited to that of the electricity transmission system, but suitable models could be used in conjunction to study the the relationsships between gas and electricity markets. As in Kienzle et al. (2007), actions in the gas market would constrain the generators options to sell power in subsequent electricity auctions. In the same way that agents using policy gradient learning methods can better exploit conditions in electricity markets, these methods could be used to learn complex strategies for buying and selling allowances while avioding penalties for exceeding quotas.

7.2 Summary Conclusions

Bibliography

Alam, M. S., Bala, B. K., Huo, A. M. Z., & Matin, M. A. (1991). A model for the quality of life as a function of electrical energy consumption. Energy, 16(4), 739-745.

Amerongen, R. van. (1989, May). A general-purpose version of the fast decoupled load flow. Power Systems, IEEE Transactions on, 4(2), 760-770.

Application of Probability Methods Subcommittee. (1979, November). Ieee reliability test system. Power Apparatus and Systems, IEEE Transactions on, PAS-98(6), 2047-2054.

Auer, P., Cesa-Bianchi, N., Freund, Y., & Schapire, R. E. (2003). The non-stochastic multiarmed bandit problem. SIAM Journal of Computing, 32(1), 48-77.

Baird, L. (1995). Residual algorithms: Reinforcement learning with function approximation. In Proceedings of the twelfth international conference on machine learning (p. 30-37). Morgan Kaufmann.

Bellman, R. E. (1961). Adaptive control processes – A guided tour. Princeton, New Jersey, U.S.A.: Princeton University Press.

Bhatnagar, S., Sutton, R. S., Ghavamzadeh, M., & Lee, M. (2009). Natural actor-critic algorithms. Automatica, 45(11), 2471–2482.

Bishop, C. M. (1996). Neural networks for pattern recognition (1st ed.). Oxford University Press, USA. Paperback.

Bower, J., & Bunn, D. (2001, March). Experimental analysis of the efficiency of uniform-price versus discriminatory auctions in the england and wales electricity market. Journal of Economic Dynamics and Control, 25(3-4), 561-592.

Bower, J., Bunn, D. W., & Wattendrup, C. (2001). A model-based analysis of strategic consolidation in the german electricity industry. Energy Policy, 29(12), 987-1005.

Bunn, D., & Martoccia, M. (2005). Unilateral and collusive market power in the electricity pool of England and Wales. Energy Economics.

Bunn, D. W., & Oliveira, F. S. (2003). Evaluating individual market power in electricity markets via agent-based simulation. Annals of Operations Research, 57-77.

Carpentier, J. (1962, August). Contribution à l'étude du Dispatching Economique. Bulletin de la Society Francaise Electriciens, 3(8), 431-447.

Department of Energy and Climate Change. (2009). Digest of United Kingdom Energy Statistics 2009. In (chap. 5). National Statistics – Crown.

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. The American Economic Review, 88(4), 848-881.

Ernst, D., Minoia, A., & Ilic, M. (2004, June). Market dynamics driven by the decision-making of both power producers and transmission owners. In Power Engineering Society General Meeting, 2004. IEEE (p. 255-260).

Fausett, L. (Ed.). (1994). Fundamentals of neural networks: architectures, algorithms, and applications. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Gieseler, C. (2005). A Java reinforcement learning module for the Repast toolkit: Facilitating study and implementation with reinforcement learning in social science multi-agent simulations. Unpublished master's thesis, Department of Computer Science, Iowa State University.

Glimn, A. F., & Stagg, G. W. (1957, april). Automatic calculation of load flows. Power Apparatus and Systems, Part III. Transactions of the American Institute of Electrical Engineers, 76(3), 817-825.

Goldfarb, D., & Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. Mathematical Programming, 27, 1-33.

Gordon, G. (1995). Stable function approximation in dynamic programming. In Proceedings of Twelfth International Conference on Machine Learning (p. 261-268). Morgan Kaufmann.

Grainger, J., & Stevenson, W. (1994). Power system analysis. New York: McGraw-Hill.

Guo, M., Liu, Y., & Malec, J. (2004, October). A new Q-learning algorithm based on the metropolis criterion. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 34(5), 2140-2143.

ICF Consulting. (2003, August). The economic cost of the blackout: An issue paper on the northeastern blackout. (Unpublished)

IEEE Working Group. (1973, November). Common format for exchange of solved load flow data. Power Apparatus and Systems, IEEE Transactions on, 92(6), 1916-1925.

Kallrath, J., Pardalos, P., Rebennack, S., & Scheidt, M. (2009). Optimization in the energy industry. Springer.

Kienzle, F., Krause, T., Egli, K., Geidl, M., & Andersson, G. (2007, September). Analysis of strategic behaviour in combined electricity and gas markets using agent-based computational economics. In 1st European workshop on energy market modelling using agent-based computational economics (p. 121-141). Karlsruhe, Germany.

Kietzmann, T. C., & Riedmiller, M. (2009). The neuro slot car racer: Reinforcement learning in a real world setting. Machine Learning and Applications, Fourth International Conference on, 0, 311-316.

Kirschen, D. S., & Strbac, G. (2004). Fundamentals of power system economics. Chichester: John Wiley & Sons.

Krause, T., & Andersson, G. (2006). Evaluating congestion management schemes in liberalized electricity markets using an agent-based simulator. In Power Engineering Society General Meeting, 2006. IEEE.

Krause, T., Andersson, G., Ernst, D., Beck, E., Cherkaoui, R., & Germond, A. (2004). Nash Equilibria and Reinforcement Learning for Active Decision Maker Modelling in Power Markets. In Proceedings of 6th IAEE European Conference 2004, modelling in energy economics and policy.

Krause, T., Beck, E. V., Cherkaoui, R., Germond, A., Andersson, G., & Ernst, D. (2006). A comparison of Nash equilibria analysis and agent-based modelling for power markets. International Journal of Electrical Power & Energy Systems, 28(9), 599-607.

Li, H., & Tesfatsion, L. (2009a, July). The ames wholesale power market test bed: A computational laboratory for research, teaching, and training. In IEEE Proceedings, Power and Energy Society General Meeting. Alberta, Canada.

Li, H., & Tesfatsion, L. (2009b, March). Capacity withholding in restructured wholesale power markets: An agent-based test bed study. In Power systems conference and exposition, 2009 (p. 1-11).

Lincoln, R., Galloway, S., & Burt, G. (2009, May). Open source, agent-based energy market simulation with Python. In Proceedings of the 6th International Conference on the European Energy Market, 2009. EEM 2009. (p. 1-5).

Maei, H. R., & Sutton, R. S. (2010). Gq(λ): A general gradient algorithm for temporal-difference prediction learning with eligibility traces. In In proceedings of the third conference on artificial general intelligence. Lugano, Switzerland.

McCulloch, W., & Pitts, W. (1943, December 21). A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biology, 5(4), 115-133.

Micola, A. R., Banal-Estañol, A., & Bunn, D. W. (2008, August). Incentives and coordination in vertically related energy markets. Journal of Economic Behavior & Organization, 67(2), 381-393.

Micola, A. R., & Bunn, D. W. (2008). Crossholdings, concentration and information in capacity-constrained sealed bid-offer auctions. Journal of Economic Behavior & Organization, 66(3-4), 748-766.

Minkel, J. R. (2008, August 13). The 2003 northeast blackout–five years later. Scientific American.

Momoh, J., Adapa, R., & El-Hawary, M. (1999, Feb). A review of selected optimal power flow literature to 1993. I. Nonlinear and quadratic programming approaches. Power Systems, IEEE Transactions on, 14(1), 96-104.

Momoh, J., El-Hawary, M., & Adapa, R. (1999, Feb). A review of selected optimal power flow literature to 1993. II. Newton, linear programming and interior point methods. Power Systems, IEEE Transactions on, 14(1), 105-111.

Moody, J., & Saffell, M. (2001, July). Learning to trade via direct reinforcement. IEEE Transactions on Neural Networks, 12(4), 875-889.

Moody, J., Wu, L., Liao, Y., & Saffell, M. (1998). Performance functions and reinforcement learning for trading systems and protfolios. Journal of Forecasting, 17, 441-470.

Naghibi-Sistani, M., Akbarzadeh-Tootoonchi, M., Javidi-D.B., M., & Rajabi-Mashhadi, H. (2006, November). Q-adjusted annealing for Q-learning of bid selection in market-based multisource power systems. Generation, Transmission and Distribution, IEE Proceedings, 153(6), 653-660.

National Electricity Transmission System Operator. (2010, May). 2010 National Electricity Transmission System Seven Year Statement (Tech. Rep.). National Grid Electricity Transmission plc.

Nicolaisen, J., Petrov, V., & Tesfatsion, L. (2002, August). Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. Evolutionary Computation, IEEE Transactions on, 5(5), 504-523.

Nicolaisen, J., Smith, M., Petrov, V., & Tesfatsion, L. (2000). Concentration and capacity effects on electricity market power. In Evolutionary Computation. Proceedings of the 2000 Congress on (Vol. 2, p. 1041-1047).

Overbye, T., Cheng, X., & Sun, Y. (2004, Jan.). A comparison of the AC and DC power flow models for LMP calculations. In System sciences, 2004. Proceedings of the 37th annual hawaii international conference on (p. 9-).

Peshkin, L., & Savova, V. (2002). Reinforcement learning for adaptive routing. In Neural Networks, 2002. IJCNN 2002. Proceedings of the 2002 International Joint Conference on (Vol. 2, p. 1825-1830).

Peters, J., & Schaal, S. (2006, October). Policy gradient methods for robotics. In Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on (p. 2219-2225).

Peters, J., & Schaal, S. (2008). Natural actor-critic. Neurocomputing, 71(7-9), 1180-1190.

Rastegar, M. A., Guerci, E., & Cincotti, S. (2009, May). Agent-based model of the Italian wholesale electricity market. In Energy Market, 2009. 6th International Conference on the European (p. 1-7).

Riedmiller, M. (2005). Neural fitted Q iteration - first experiences with a data efficient neural reinforcement learning method. In In 16th european conference on machine learning (pp. 317–328). Springer.

Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: the rprop algorithm.

Robbins, H. (1952). Some aspects of the sequential design of experiments. Bulletin American Mathematical Society, 58(5), 527-535.

Roth, A. E., Erev, I., Fudenberg, D., Kagel, J., Emilie, J., & Xing, R. X. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior, 8(1), 164-212.

Schaul, T., Bayer, J., Wierstra, D., Sun, Y., Felder, M., Sehnke, F., et al. (2010). PyBrain. Journal of Machine Learning Research, 11, 743-746.

Schweppe, F., Caramanis, M., Tabors, R., & Bohn, R. (1988). Spot pricing of electricity. Dordrecht: Kluwer Academic Publishers Group.

Sharpe, W. F. (1966, January). Mutual fund performance. Journal of Business, 119-138.

Sharpe, W. F. (1994). The Sharpe ratio. The Journal of Portfolio Management, 49-58.

Stott, B., & Alsac, O. (1974, May). Fast decoupled load flow. Power Apparatus and Systems, IEEE Transactions on, 93(3), 859-869.

Sun, J., & Tesfatsion, L. (2007a). Dynamic testing of wholesale power market designs: An open-source agent-based framework. Computational Economics, 30(3), 291-327.

Sun, J., & Tesfatsion, L. (2007b, June). Open-source software for power industry research, teaching, and training: A DC-OPF illustration. In Power Engineering Society General Meeting, 2007. IEEE (p. 1-6).

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: An introduction. MIT Press. Gebundene Ausgabe.

Sutton, R. S., McAllester, D., Singh, S., & Mansour, Y. (2000). Policy gradient methods for reinforcement learning with function approximation. In Advances in neural information processing systems (Vol. 12, p. 1057-1063).

Tellidou, A., & Bakirtzis, A. (2007, Novemeber). Agent-based analysis of capacity withholding and tacit collusion in electricity markets. Power Systems, IEEE Transactions on, 22(4), 1735-1742.

Tesauro, G. (1994). Td-gammon, a self-teaching backgammon program, achieves master-level play. Neural Computation, 6(2), 215-219.

Tesfatsion, L., & Judd, K. L. (2006). Handbook of computational economics, volume 2: Agent-based computational economics (handbook of computational economics). Amsterdam, The Netherlands: North-Holland Publishing Co.

Tinney, W., & Hart, C. (1967, Novemeber). Power flow solution by Newton's method. Power Apparatus and Systems, IEEE Transactions on, 86(11), 1449-1460.

Tsitsiklis, J. N., & Roy, B. V. (1994). Feature-based methods for large scale dynamic programming. In Machine learning (p. 59-94).

United Nations. (2003, December 9). World population in 2300. In Proceedings of the United Nations, Expert Meeting on World Population in 2300.

U.S.-Canada Power System Outage Task Force. (2004, April). Final report on the august 14, 2003 blackout in the united states and canada: Causes and recommendations (Tech. Rep.). North American Electric Reliability Corporation.

Veit, D., Weidlich, A., Yao, J., & Oren, S. (2006). Simulating the dynamics in two-settlement electricity markets via an agent-based approach. International Journal of Management Science and Engineering Management, 1(2), 83-97.

Vengerov, D. (2008). A gradient-based reinforcement learning approach to dynamic pricing in partially-observable environments. Future Generation Computer Systems, 24(7), 687-693.

Visudhiphan, P. (2003). An agent-based approach to modeling electricity spot markets. Unpublished doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology.

Visudhiphan, P., & Ilic, M. (1999, February). Dynamic games-based modeling of electricity markets. In Power Engineering Society 1999 Winter Meeting, IEEE (Vol. 1, p. 274-281).

Wang, H., Murillo-Sanchez, C., Zimmerman, R., & Thomas, R. (2007, Aug.). On computational issues of market-based optimal power flow. Power Systems, IEEE Transactions on, 22(3), 1185-1193.

Wang, J., Koritarov, V., & Kim, J.-H. (2009, July). An agent-based approach to modeling interactions between emission market and electricity market. In Power Energy Society General Meeting, 2009. PES 2009. IEEE (p. 1-8).

Weidlich, A., & Veit, D. (2006, July 7-10). Bidding in interrelated day-ahead electricity markets - insights from an agent-based simulation model. In Proceedings of the 29th IAEE International Conference.

Weidlich, A., & Veit, D. (2008, July). A critical survey of agent-based wholesale electricity market models. Energy Economics, 30(4), 1728-1759.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. In Machine learning (p. 229-256).

Wood, A. J., & Wollenberg, B. F. (1996). Power Generation Operation and Control (second ed.). New York: Wiley, New York.

Yao, J., Adler, I., & Oren, S. S. (2008). Modeling and computing two-settlement oligopolistic equilibrium in a congested electricity network. Operations Research, 56(1), 34-47.

Yao, J., Oren, S. S., & Adler, I. (2007). Two-settlement electricity markets with price caps and cournot generation firms. European Journal of Operational Research, 181(3), 1279-1296.

Zimmerman, R. (2010, March 19). MATPOWER 4.0b2 User's Manual (Version 4.0b2 ed.) [Computer software manual]. School of Electrical Engineering, Cornell University, Ithaca, NY 14853.

Zimmerman, R., Murillo-Sánchez, C., & Thomas, R. J. (2009, July). MATPOWER's extensible optimal power flow architecture. In IEEE PES General Meeting. Calgary, Alberta, Canada.

Appendix A

Open Source Power Engineering Software

To couple existing implementations of policy gradient reinforcement learning methods from the PyBrain machine learning library (Schaul et al., 2010) with scalable and extensible optimal power flow formulations, the Matlab[1] source code from Matpower was translated to the Python programming language for this thesis. With permission from the Matpower developers, the resulting package was released under the terms of the Apache License version 2.0 (Lincoln, Galloway, & Burt, 2009). This section briefly describes the project in the context of other open source Electric Power Engineering software to illustrate the contribution made.

A.1 MATPOWER

Since 1996, a team of researchers at the Power Systems Engineering Research Center at Cornell University have been developing Matpower – a package of Matlab workspace files for solving power flow and optimal power flow problems (Zimmerman, Murillo-Sánchez, & Thomas, 2009). Initial development was part of the PowerWeb project in which the team created a power exchange auction market simulator that could be accessed by multiple users simultaneously through a web-based interface. Matpower is available under a custom license that permits it to be used for any purpose providing the project and authors are cited correctly. It has become very popular in education and research and has an active mailing list which is moderated by Ray Zimmerman.

Matpower includes five power flow solvers for both AC and DC problems.

[1]Matlab is a registered tradeamark of The Mathworks, Inc.

Package	Language	Licence	PF	DCOPF	ACOPF	CPF	SSSA	TDS	SE	SP	GUI	RL
AMES	Java	GPL		•							•	•
DCOPFJ	Java	GPL		•								
MatDyn	Matlab									•		
MATPOWER	Matlab		•	•	•	•			•	•		
OpenDSS	Pascal	BSD	•							•	•	
PSAT	Matlab	GPL	•		•	•	•	•		•	•	
PYLON	Python	Apache	•	•	•				•	•	•	•
TEFTS	C					•		•		•		
VST	Matlab		•			•	•	•		•	•	
UWPFLOW	C					•				•		

Table A.1: Open source electric power engineering software feature matrix.

The default solver uses Newton's method (Tinney & Hart, 1967) with a full Jacobian matrix updated in each iteration. Two variations on the fast decoupled method (Stott & Alsac, 1974) described in Amerongen (1989) provide quicker convergence for certain networks. The standard Gauss-Seidel method (Glimn & Stagg, 1957) is provided for academic purposes and the DC solver provides non-iterative solutions. The properties of Matlab sparse matrices are fully exploited to allow the solvers to scale well to very large systems. All functions are run from the Matlab command-line or from within users programs and no graphical user interface is provided.

Starting with version 4.0, MATPOWER includes the Matlab Interior Point Solver (MIPS) that can be used for solving DC and AC optimal power flow problems (H. Wang et al., 2007). Previously, FMINCON from the Matlab Optimization Toolbox[2] was required or one of a suite of high performance closed-source solvers. TSPOPF is a collection of three AC optimal power flow solvers, implemented in the C programming language and released as Matlab MEX files. It includes the original implementation of the step-controlled interior point method from which MIPS was derived. MINOPF provides an interface to the Fortran based MINOS[3] solver, developed at the Systems Optimization Laboratory at Stanford University, and is available only for educational and research purposes. DC optimal power flow problems can be solved with a Quadratic Programming interface to MIPS or using a MEX interface to BPMPD – a commercial interior point method for linear and quadratic programming.

MATPOWER has an *extensible* optimal power flow formulation that allows additional optimisation variables and problem constraints to be introduced by the user. It is used internally to extend the standard optimisation formulation to support piecewise linear cost functions, dispatchable loads, generator PQ capability curves and branch angle difference limit constraints. Examples of possible additional extensions include: reserve requirements, environmental costs and contingency constraints.

MATPOWER currently requires Matlab (version 6.5 or later) which is a commercial software product from The Mathworks that is supported on all major platforms. However, with minimal alteration MATPOWER has been shown to run on GNU/Octave[4] version 3.2.3.

[2]Optimization Toolbox is a registered trademark of The Mathworks, Inc.
[3]MINOS is trademark of Stanford Business Software, Inc.
[4]GNU/Octave is an free program for numerical computation with strong Matlab compatibility.

A.2 MATDYN

MATDYN is an extension to MATPOWER developed by Stijn Cole from the Katholieke Universiteit Leuven for dynamic analysis of electric power systems. It was first released in 2009 under the same license as MATPOWER and the same programming style has been used. The MATPOWER case format is extended with structs for dynamic and event data. MATDYN uses MATPOWER to obtain a power flow solution that is then used in solving the system of differential algebraic equations representing the power system. Results for MATDYN are validated against those obtained from PSS/E[5] and the Power System Analysis Toolbox and show good correspondence.

A.3 Power System Analysis Toolbox

The Power System Analysis Toolbox (PSAT) is a Matlab toolbox for static and dynamic analysis of electric power systems developed by Federico Milano, currently an Assistant Professor at the University of Castilla in Spain. It is released under the terms of the GNU General Public License (GPL) version 2 and offers routines for:

- Power flow,
- Bifurcation analysis,
- Optimal power flow,
- Small signal stability analysis,
- Time domain simulation and
- Phasor measurement unit placement.

A large number of input data formats are supported through Perl scripts and simulation reports can be exported as plain text, Excel spreadsheets or LaTeX 2_ε code. PSAT may be run from the Matlab command-line or through a Matlab based graphical user interface. The graphical interface can be used with Simulink[6] to construct cases such as the network from the UK Generic Distribution System shown in Figure A.1. A slightly modified version of PSAT that can be run from the GNU/Octave command-line is also available.

[5]PSS/E is a registered trademark of Siemens Power Transmission & Distribution, Inc. Power Technologies International.

[6]Simulink is a registered trademark of The Mathworks, Inc.

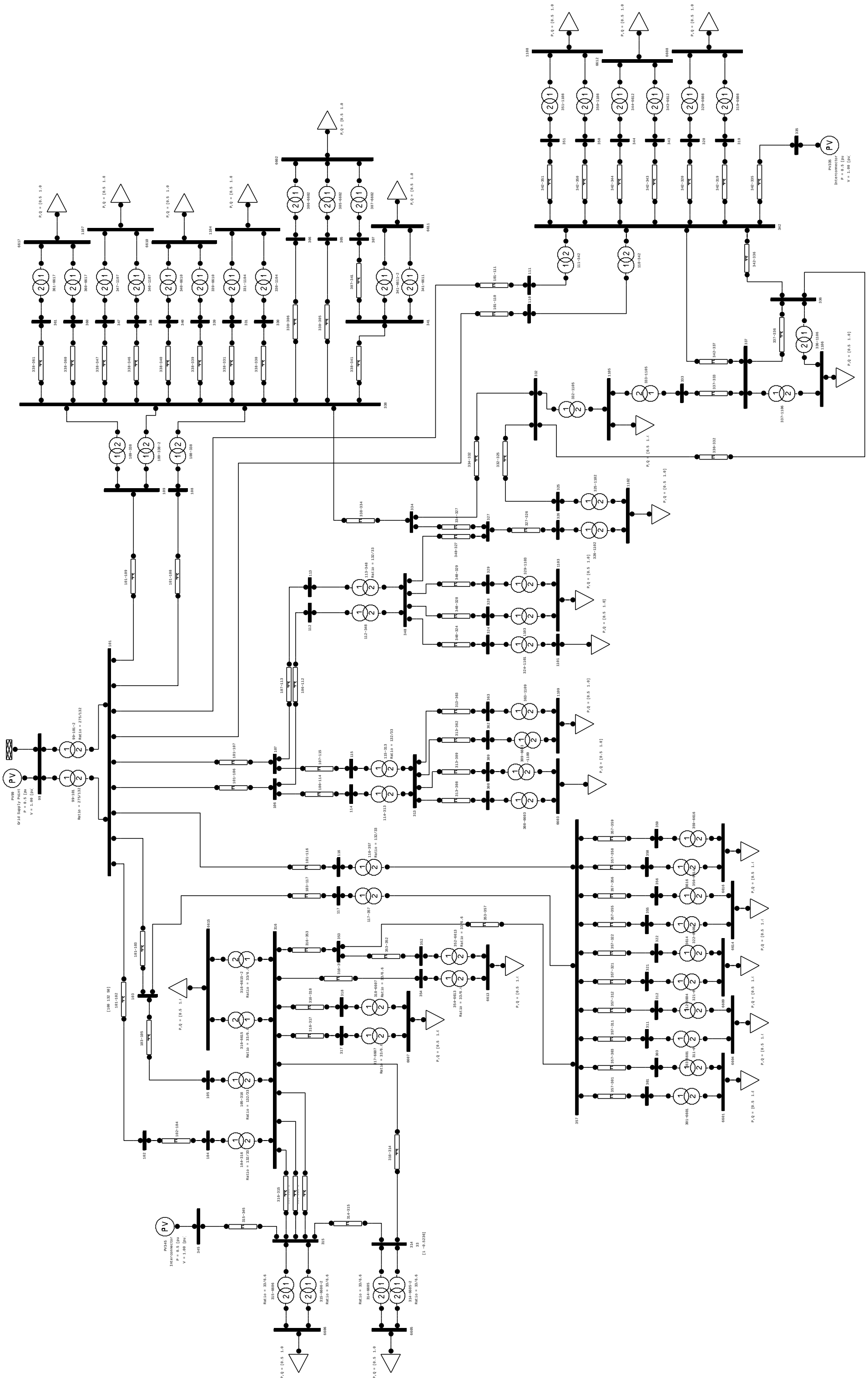

Figure A.1: UKGDS EHV3 model in PSAT Simulink network editor.

Optimal power flow problems are solved via an interface to the General Algebraic Modeling System (GAMS). GAMS defines optimisation problems using a high-level modelling language and has a large solver portfolio, including all of the major commercial and academic solvers. The interface can be used for solving single period optimal power flow problems where the objective function can model maximisation of social benefit, maximisation of the distance to the maximum loading condition or multi-objective of a combination of these. Multi-period optimal power flow is formulated as a mixed integer problem with linearised power balance constraints. The objective function models maximisation of social welfare, but is extended to include startup and shutdown costs.

Power flow and dynamic data are typically separated in electric power simulation tools, but in PSAT they are integrated. This combined with the large number of routines supported by PSAT can make the code base difficult to understand and modify. However, comprehensive documentation is included with PSAT and the mailing list is highly active. The price of GAMS licenses and the need for optimal power flow problems to be converted to the GAMS language before being solved could be considered barriers to its selection for certain projects.

A.4 UWPFLOW

UWPFLOW is a research tool for voltage stability analysis developed at the University of Waterloo, Ontario, and the University of Wisconsin-Madison. It is written in ANSI-C and is available as open source for research purposes only. The program can be run with the terminal command

```
$ uwpflow [-options] input_file
```

where `input_file` is the path to a data file in the IEEE common data format (CDF) (IEEE Working Group, 1973) that may contain High-Voltage Direct Current (HVDC) and Flexible Alternating Current Transmission System (FACTS) device data. Output is also in CDF and can include additional data for post-processing, including values for nose curve plots. An interface to UWPFLOW is provided with PSAT and can be used for bifurcation analysis.

A.5 TEFTS

The University of Waterloo also hosts TEFTS – a transient stability program for studying energy functions and voltage stability phenomena in AC/HVDC dynamic power system models. It too is written in ANSI-C and is licensed for research purposes only. An executable file for DOS is provided and the source package contains a simple example.

Voltage Stability Toolbox

The Voltage Stability Toolbox (VST) is a Matlab toolbox, developed at the Center for Electric Power Engineering at Drexel University in Philidelphia, for investigating stability and bifurcation issues in power systems. The source is available for any purpose providing that the authors are suitably cited. VST features routines for:

- Power flow,
- Time domain simulation,
- Static and dynamic bifurcation analysis,
- Singularity analysis and
- Eigenvalue analysis.

The feature matrix in Table A.1 shows the similar capabilities of VST and PSAT. It was developed around the same time and has the same goals for educational and research applications. However it does not have the same quality of documentation nor such an active community of users and developers as PSAT.

A.6 Distribution System Simulator

In November 2008, the Open Distribution System Simulator (OpenDSS) was released by the Electric Power Research Institute (EPRI) as open source. Development of OpenDSS began in April 1997 and it has been used extensively in distributed generation impact assessments. It is the only open source program designed for both distribution and transmission system simulation.

OpenDSS supports steady-state analysis in the frequency domain, including power flow, harmonics and dynamics. Arbitrary n-phase unbalanced circuit analysis is supported using an object orientated data model. Circuit elements are defined in Object Pascal and solutions are found using a linear sparse matrix solver written in C and C++. OpenDSS is available under the Berkeley Software Distribution (BSD) license, which allows use for almost any purpose. Circuits are defined in scripts, using a domain specific language, that may be executed through a graphical user interface or a Common Object Model (COM) interface. The user interface also provides circuit data editing, plotting and power flow visualisation tools.

The power flow solver is fast and can be configured for repeated studies using daily, yearly or duty-cycle data. The multi-phase circuit model allows complex fault conditions to be defined and three short-circuit analysis methods are provided. The heritage of OpenDSS is in harmonics and dynamics analysis and it does not support system optimisation.

A.7 Agent-based Modelling of Electricity Systems

The AMES (Agent-based Modeling of Electricity Systems) power market test bed is a software package that models core features of the Wholesale Power Market Platform – a market design proposed by the Federal Energy Regulatory Commission (FERC) in April 2003 for common adoption in regions of the U.S. (Sun & Tesfatsion, 2007a). The market design features:

- A centralised structure managed by an independent market operator,
- Parallel day-ahead and real-time markets and
- Locational marginal pricing.

Learning agents represent load serving entities or generating companies and learn using Roth-Erev methods (see Section 2.4.3) implemented with the Repast agent simulation toolkit (Gieseler, 2005). Agents learn from the solutions of hourly bid/offer based DC-OPF problems formulated as quadratic programs using the DCOPFJ package (Sun & Tesfatsion, 2007b) described in Section A.8, below.

The capabilities of AMES are demonstrated using a 5-bus network model in Li and Tesfatsion (2009a). The model is provided with AMES and a step-by-step tutorial describes how it may be used. AMES comes with a Swing-based graphical user interface with plotting and table editor tools and is released under the the Gnu GPL version 2.

A.8 DCOPFJ

To solve market problems defined in AMES, researchers at Iowa State University developed a stand-alone DC optimal power flow solver in Java named DCOPFJ. It formulates optimal power flow problems as convex quadratic programs which are solved using QuadProgJ. The same researcher developed QuadProgJ as an independent solver that uses the dual active set strictly convex quadratic programming algorithm (Goldfarb & Idnani, 1983). DCOPFJ requires generator costs to be modelled as polynomial functions, of second order or less and no sparse matrix techniques are employed to allow application to large systems.

A.9 PYLON

PYLON is a translation of MATPOWER and Matdyn to the Python programming language. It has extensions for agent-based electricity market simulation that provide features similar to those of AMES. Both the DC and AC forumlations of the extensible optimal power flow model (Zimmerman et al., 2009) from MATPOWER are implemented. Either a Python version of MIPS or an interface to the cp solver from CVXOPT can be used to compute solutions. The sparsity of the problems is exploited throughout the solution process using matrix packages from SciPy and bindings to SuperLU or UMFPACK for solver sparse set of linear equations and performing LU decomposition. Scripts are provided for reading and writing data files in PSS/E, MATPOWER and PSAT format. A wide variety of learing methods are available in PYLON due to its use of the PyBrain machine learning library (Schaul et al., 2010). PyBrain also provides the artificial neural network models used for policy function approximation and they may be accelerated using C extension modules from the ARAC project.

In addition to its market simulation capabilities, PYLON also features solvers for power flow problems (using Newton's method or fast decoupled methods), state estimation, continuation power flow and time domain simulation. PYLON includes a rudimentary GUI that uses the Tkinter library that is included with Python and thus imposes no additional dependencies. A more feature rich GUI is provided by plug-ins for Puddle – an extensible, GUI toolkit independant integrated development environment, created for the purposes of this thesis also. The use of matrix libraries from NumPy and SciPy has allowed PYLON (with the permission of the MATPOWER developers) to be released under the Apache license, version 2.0. This allows PYLON to be used in the developement of proprietary software as well as free and open source software since derivatives of the source code may be made available under more restrictive terms than the original Apache license. This is in contrast to “copyleft” licenses, such as the GNU GPL, that require the same rights to be preserved in modified versions of the work.

Appendix B

Case Data

This appendix provides the data for the electric power system models used in Chapters 5 and 6.

B.1 6-Bus Case

The data for the six bus case adapted from Wood and Wollenberg (1996, pp. 104, 112, 119, 123-124, 549) is presented in this section. The data was imported from the "case6ww.m" case file provided with MATPOWER. Figure **??** illustrates the structure of the model and shows the bus injections for the AC unit decommitment optimal power flow solution. Table B.1 lists the bus data, Table B.2 lists the generator data and Table B.3 lists the branch data.

Bus	P_d	Q_d	G_s	B_s	V_{base}	V_{max}	V_{min}
1	0	0	0	0	230	1.05	1.05
2	0	0	0	0	230	1.05	1.05
3	0	0	0	0	230	1.07	1.07
4	70	70	0	0	230	1.05	0.95
5	70	70	0	0	230	1.05	0.95
6	70	70	0	0	230	1.05	0.95

Table B.1: 6-bus case bus data.

B.2 IEEE Reliability Test System

This appendix provides the data from the IEEE Reliability Test System (Application of Probability Methods Subcommittee, 1979) that was imported from the "case24_ieee_rts.m" case file that is provided with MATPOWER and was originally contributed by

Bus	P_{max}	P_{min}	V_g	Q_{max}	Q_{min}
1	1.05	200	50	100	-100
2	1.05	150	37.5	100	-100
3	1.07	180	45	100	-100

Table B.2: 6-bus case generator data.

From	To	r	x	b_c	S_{max}	τ	θ_{ph}
1	2	0.1	0.2	0.04	40	0	0
1	4	0.05	0.2	0.04	60	0	0
1	5	0.08	0.3	0.06	40	0	0
2	3	0.05	0.25	0.06	40	0	0
2	4	0.05	0.1	0.02	60	0	0
2	5	0.1	0.3	0.04	30	0	0
2	6	0.07	0.2	0.05	90	0	0
3	5	0.12	0.26	0.05	70	0	0
3	6	0.02	0.1	0.02	80	0	0
4	5	0.2	0.4	0.08	20	0	0
5	6	0.1	0.3	0.06	40	0	0

Table B.3: 6-bus case branch data.

Bruce Wollenberg. Figure **??** illustrates the model struture and shows the bus injections for the AC unit de-commitment optimal power flow solution for all generators at marginal cost. Table B.4 lists the bus data, Table B.5 lists the generator data, Table B.6 lists the branch data and Table B.7 lists the generator cost data provided by Georgia Tech Power Systems Control and Automation Laboratory.

Bus	P_d	Q_d	G_s	B_s	V_{base}	V_{max}	V_{min}
1	108	22	0	0	138	1.05	0.95
2	97	20	0	0	138	1.05	0.95
3	180	37	0	0	138	1.05	0.95
4	74	15	0	0	138	1.05	0.95
5	71	14	0	0	138	1.05	0.95
6	136	28	0	-100	138	1.05	0.95
7	125	25	0	0	138	1.05	0.95
8	171	35	0	0	138	1.05	0.95
9	175	36	0	0	138	1.05	0.95
10	195	40	0	0	138	1.05	0.95
11	0	0	0	0	230	1.05	0.95
12	0	0	0	0	230	1.05	0.95
13	265	54	0	0	230	1.05	0.95
14	194	39	0	0	230	1.05	0.95
15	317	64	0	0	230	1.05	0.95
16	100	20	0	0	230	1.05	0.95
17	0	0	0	0	230	1.05	0.95
18	333	68	0	0	230	1.05	0.95
19	181	37	0	0	230	1.05	0.95
20	128	26	0	0	230	1.05	0.95
21	0	0	0	0	230	1.05	0.95
22	0	0	0	0	230	1.05	0.95
23	0	0	0	0	230	1.05	0.95
24	0	0	0	0	230	1.05	0.95

Table B.4: IEEE RTS bus data.

Bus	P_{max}	P_{min}	V_g	Q_{max}	Q_{min}	Type
1	20	16	1.035	10	0	U20
1	20	16	1.035	10	0	U20
1	76	15.2	1.035	30	-25	U76
1	76	15.2	1.035	30	-25	U76
2	20	16	1.035	10	0	U20
2	20	16	1.035	10	0	U20
2	76	15.2	1.035	30	-25	U76
2	76	15.2	1.035	30	-25	U76
7	100	25	1.025	60	0	U100
7	100	25	1.025	60	0	U100
7	100	25	1.025	60	0	U100
13	197	69	1.02	80	0	U197
13	197	69	1.02	80	0	U197
13	197	69	1.02	80	0	U197
14	0	0	0.98	200	-50	SynCond
15	12	2.4	1.014	6	0	U12
15	12	2.4	1.014	6	0	U12
15	12	2.4	1.014	6	0	U12
15	12	2.4	1.014	6	0	U12
15	12	2.4	1.014	6	0	U12
15	155	54.3	1.014	80	-50	U155
16	155	54.3	1.017	80	-50	U155
18	400	100	1.05	200	-50	U400
21	400	100	1.05	200	-50	U400
22	50	10	1.05	16	-10	U50
22	50	10	1.05	16	-10	U50
22	50	10	1.05	16	-10	U50
22	50	10	1.05	16	-10	U50
22	50	10	1.05	16	-10	U50
22	50	10	1.05	16	-10	U50
23	155	54.3	1.05	80	-50	U155
23	155	54.3	1.05	80	-50	U155
23	350	140	1.05	150	-25	U350

Table B.5: IEEE RTS generator data.

From	To	r	x	b_c	S_{max}	τ	θ_{ph}
1	2	0.0026	0.0139	0.4611	175	0	0
1	3	0.0546	0.2112	0.0572	175	0	0
1	5	0.0218	0.0845	0.0229	175	0	0
2	4	0.0328	0.1267	0.0343	175	0	0
2	6	0.0497	0.192	0.052	175	0	0
3	9	0.0308	0.119	0.0322	175	0	0
3	24	0.0023	0.0839	0	400	1.03	0
4	9	0.0268	0.1037	0.0281	175	0	0
5	10	0.0228	0.0883	0.0239	175	0	0
6	10	0.0139	0.0605	2.459	175	0	0
7	8	0.0159	0.0614	0.0166	175	0	0
8	9	0.0427	0.1651	0.0447	175	0	0
8	10	0.0427	0.1651	0.0447	175	0	0
9	11	0.0023	0.0839	0	400	1.03	0
9	12	0.0023	0.0839	0	400	1.03	0
10	11	0.0023	0.0839	0	400	1.02	0
10	12	0.0023	0.0839	0	400	1.02	0
11	13	0.0061	0.0476	0.0999	500	0	0
11	14	0.0054	0.0418	0.0879	500	0	0
12	13	0.0061	0.0476	0.0999	500	0	0
12	23	0.0124	0.0966	0.203	500	0	0
13	23	0.0111	0.0865	0.1818	500	0	0
14	16	0.005	0.0389	0.0818	500	0	0
15	16	0.0022	0.0173	0.0364	500	0	0
15	21	0.0063	0.049	0.103	500	0	0
15	21	0.0063	0.049	0.103	500	0	0
15	24	0.0067	0.0519	0.1091	500	0	0
16	17	0.0033	0.0259	0.0545	500	0	0
16	19	0.003	0.0231	0.0485	500	0	0
17	18	0.0018	0.0144	0.0303	500	0	0
17	22	0.0135	0.1053	0.2212	500	0	0
18	21	0.0033	0.0259	0.0545	500	0	0
18	21	0.0033	0.0259	0.0545	500	0	0
19	20	0.0051	0.0396	0.0833	500	0	0
19	20	0.0051	0.0396	0.0833	500	0	0
20	23	0.0028	0.0216	0.0455	500	0	0
20	23	0.0028	0.0216	0.0455	500	0	0
21	22	0.0087	0.0678	0.1424	500	0	0

Table B.6: IEEE RTS branch data.

Gen	C_{up}	a	b	c	Type
1	1500	0	130	400.685	U20
2	1500	0	130	400.685	U20
3	1500	0.01414	16.0811	212.308	U76
4	1500	0.01414	16.0811	212.308	U76
5	1500	0	130	400.685	U20
6	1500	0	130	400.685	U20
7	1500	0.01414	16.0811	212.308	U76
8	1500	0.01414	16.0811	212.308	U76
9	1500	0.05267	43.6615	781.521	U100
10	1500	0.05267	43.6615	781.521	U100
11	1500	0.05267	43.6615	781.521	U100
12	1500	0.00717	48.5804	832.758	U197
13	1500	0.00717	48.5804	832.758	U197
14	1500	0.00717	48.5804	832.758	U197
15	1500	0	0	0	SynCond
16	1500	0.32841	56.564	86.3852	U12
17	1500	0.32841	56.564	86.3852	U12
18	1500	0.32841	56.564	86.3852	U12
19	1500	0.32841	56.564	86.3852	U12
20	1500	0.32841	56.564	86.3852	U12
21	1500	0.00834	12.3883	382.239	U155
22	1500	0.00834	12.3883	382.239	U155
23	1500	0.00021	4.4231	395.375	U400
24	1500	0.00021	4.4231	395.375	U400
25	1500	0	0.001	0.001	U50
26	1500	0	0.001	0.001	U50
27	1500	0	0.001	0.001	U50
28	1500	0	0.001	0.001	U50
29	1500	0	0.001	0.001	U50
30	1500	0	0.001	0.001	U50
31	1500	0.00834	12.3883	382.239	U155
32	1500	0.00834	12.3883	382.239	U155
33	1500	0.00490	11.8495	665.109	U350

Table B.7: IEEE RTS generator cost data.