Application of Hybrid Intelligent Agents to Modelling a Dynamic, Locally Interacting Retail Market.

by
Alison Jane Heppenstall

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of Leeds
School of Geography
October 2004

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Abstract

The emergence of agent-based modelling from the field of artificial intelligence (AI) presents a new and alternative approach to geographical modelling. The vast potential offered by agent-based models in representing distributed complex systems, coupled with the increase in available computing power has resulted in agent-based models becoming an increasingly popular and powerful tool within geographical applications. These models offer distinct advantages over traditional empirical techniques through their characteristics of autonomy, flexibility and adaptability. There is an emerging recognition that the power of agent-based systems is enhanced when integrated with other AI-based and conventional approaches. The resulting hybrid models are powerful tools that combine the flexibility of the agent-based methodology with the strengths of more traditional modelling.

This research examines the application of a hybrid agent-based model to the case study of the retail petrol market. Detailed analysis of the real data was first performed before the construction of an agent-based model. Model performance was evaluated against real data from the UK for a three month period in 1999. On the basis of this evaluation, the agent model was further developed to incorporate consumer behaviour by the inclusion of a spatial interaction (SI) model and a network model. Suitable parameters for these models were derived through detailed analysis of the real data, numerical experimentation and experimentation on the real data. These developments improved the performance of the model. A genetic algorithm (GA) was constructed to provide an objective approach to deriving optimal parameters. There was a close agreement in the values selected by the GA and those derived by hand. This research clearly demonstrates that agent-based modelling has the ability to improve on existing geographical models. Further investigation is needed if this potential is to be fully realised for a range if geographical problems.

Contents

List of Figures

List of Tables

Acknowledgements

This research would not have been possible without the help, support and advice of many people to whom I am extremely grateful. First, thanks to my supervisors, Andrew Evans and Mark Birkin, for their advice and guidance over the last three years. A support role has been provided by my rogue RSG committee; thanks to Ian Turton, Jason Noble, Dimitris Ballas and Graham Clarke for their input at various stages.

This research would not have been possible if it were not for funding from the ESRC and GMap for their partnership in the CASE award. In addition, I was fortunate enough to hold a WUN scholarship which allowed time to be spent at Penn State University. My thanks to Mark Gahagen, James Macgill and in particular, David O'Sullivan.

Within the School of Geography, several people should be thanked. First, Pete Shepherd and his genius creation of"muffin Friday"; Hazel Parry for savvy fashion advice and holidays/conferences; Olga McFarland, petrol princess and repressed geek and Tim "Maurice" James, the happy Canadian. My thanks also to Bokhwan Kim, Charatdoa "Tok" Kongmuang, Jo Clarke, Alan Cundill, Elyze Smeets and members of both the MASS group and CCG.

Outside Geography, several people deserve a mention for support and encouragement (or is that bemusement?) during this thesis. A big thank you to Andrew "George" Gibb for fine whisky and slinky tunes. Thanks also to Jon Greenwood, Andrew Leach, Gill Dean, Sarah Haney, Jane Montgomery and Dominique Brown.

This thesis represents three years of my life in Leeds. These years would have been quite lacking if I hadn't been fortunate enough to indulge my obsession for music and play with the "better than Black Dyke" Otley Brass Band. My thanks to all the band, in particular to Jim Little for keeping order and humour in the back ranks!

To my family, I promise not to do another degree...just yet; it is a fools hope that Lauren, Megan, Caitlin and Leo will ever see this thesis as more than a colouring book, but I remain optimistic.

The largest and final thanks must go to Andrew Ross. For his perpetual support, encouragement, enthusiasm and big smile, I will always be grateful.

Chapter 1

Introduction

1.1 Aims and Objectives of the Research

There are a range of geographical problems that are spatially and temporally complex. Often within human geographical systems, these problems have a distributed network of, for example, people or companies making decisions that have an impact on others within the system. A good example of this can be found within retail markets. Here, companies make decisions on pricing whilst consumers make decisions on where to purchase one or more commodities. The competition between retailers to attract customers, and therefore make profit, results in complex spatial interactions between all participants within the market.

With such a high degree of complexity within geographical systems, it is unsurprising that traditional empirically-based techniques have failed to produce realistic representations of human behaviour. Many geographical systems possess highly non-linear relationships amongst a large number of variables and are also characterised by high degrees of imprecision and uncertainty. Cavezzali and Rabino (2003) argue that there is a need for new models to represent phenomena with high complexity whose structure is not entirely known or not mathematically expressible.

In recent years, the field of artificial intelligence (AI) has produced several problem solving algorithms inspired by natural genetic evolution and by human interactions (reviewed in Openshaw and Openshaw, 1997). The general methodologies and heuristics of AI can potentially add more intelligence to current geographical models which may ultimately lead to better geographical representations of the world (See, 1999).

Agent-based modelling is one of several new technologies that has emerged from the field of AI. Due to the advantages that they offer over traditional approaches, agent-based models have become an increasingly popular and powerful tool within geographical applications (O'Sullivan and Haklay, 2000). The agent paradigm offers the possibility to represent individuals, their behaviour and their interactions. This makes it possible to analyse a phenomenon as the result of interactions of autonomous entities (the agents).

This type of decomposition is a natural metaphor for complex systems that are always distributed with numerous autonomous decision making parts. The agent framework is particularly applicable to geographical systems, such as those described above, that are characterised by complex spatial and temporal dynamics. Multi-agent simulation makes it possible to create artificial environments that mimic real systems. The behaviour of the agents at the individual level is specified and through their interactions, large-scale features emerge. Furthermore, when different AI techniques, for example genetic algorithms (GAs), are fused with the power of agent-based models the result is a hybrid approach that often produces better solutions than the use of an agent techniques alone.

This thesis will examine a complex geographical system to assess the potential for hybrid agent-based systems in actually reproducing these complexities. The retail petrol market has been chosen as a case-study for several reasons. This market has the advantage of selling a single homogeneous product. This means that it does not possess any of the complex trade-offs in both pricing and consumer's purchases that are found in many retail markets, for example the grocery market. In addition, the retail petrol market is an important system in its own right. Petrol is a ubiquitous commodity valued by retailers and customers alike. Consumer sensitivity to petrol prices was clearly demonstrated in the UK during August and September 2000 when there was a "Petrol Crisis" consisting of consumer blockages of refineries and protests in reaction to soaring fuel taxes. Over the summer of 2004, the issue of petrol prices has again been in the news, this time due to rises in the price of crude oil resulting from a combination of political instabilities (the Iraq war, Nigerian rebels, Venezuelan strikes) and natural disasters (hurricanes in the Caribbean).

Typical approaches to modelling retail markets have involved the use of empirically based models, for example regression models and spatial interaction models. However, these methods are static and require geographically aggregated data. These techniques are also unable to model the impact of individual retailer or consumer behaviour at more than one scale. However, building an agent-based model would be computationally very intensive. Using traditional methods provides an easy way to incorporate well understood behaviour. Can a model be successfully developed and applied that combines the flexibility of the agent-based methodology along with the strengths of traditional modelling?

The overall aim of this research is to examine the ways in which agent-based models can be applied to modelling a dynamic, locally interacting retail market. To achieve the aim of this thesis, the following research objectives were formulated:

1. Review and discuss the current state of agent-based applications in the petrol market highlighting potential areas for research.
2. Analysis of the real market data to look for evidence of spatial and temporal variabilities and investigation of the suitability of empirical techniques to explain these variations.
3. Use agent technology to build a model to simulate the spatial and temporal variations in price observed in a single commodity retail market.
4. Assess model responses to different configurations, initial conditions and rule sets using both idealised and real data.
5. Investigation of objective techniques to select optimal values for the parameters in the agent model.
6. Provide recommendations for future work in the form of a research agenda.

1.2 Organisation of the Thesis

Table 1.1 presents an overview of the thesis. This relates the 6 research objectives to each of the chapters.

Objective	Chapter
1: Review and discuss the current state of agent-modelling and the modelling of the petrol market highlighting potential areas for research.	2: Retail Petrol Market 3: Agent-Based Models
2: Analysis of the real market data to look for evidence of spatial and temporal variabilities and investigation of the suitability of empirical techniques to explain these variations.	4: Real Data Analysis And Traditional Modelling
3: Use agent technology to build a model to simulate the spatial and temporal variations in price observed in a single commodity retail market.	5: Agent Model Development 6: Hybrid Model Development
4: Assess model responses to different configurations, initial conditions and rule sets using both idealised and real data.	5: Agent Model Development 7: Experimentation With Idealised Data 8: Experimentation With Real Data
5: Investigation of objective techniques to select optimal values for the parameters in the agent model.	9: Optimisation Using A Genetic Algorithm
6: Provide recommendations for future work in the form of a research agenda.	10: Conclusions And Recommendations For The Future

Table 1.1: The relationship of the research objectives to the chapters.

Objective 1: Review and discuss the current state of agent-modelling and the modelling of the petrol market highlighting potential areas for research.

Chapters 2 and 3 provide a detailed assessment of the petrol price market and a review of agent-based modelling. Chapter 2 examines available information on the retail petrol market and reviews current and past modelling approaches. The criticisms of these approaches leads to the introduction and review of agent-based modelling in Chapter 3. The basic introduction presented here provides the necessary building blocks for the development of the pure agent model in Chapter 5.

Objective 2: Analysis of the real market data to look for evidence of spatial and temporal variabilities and investigation of the suitability of empirical techniques to explain these variations.

Chapter 4) fulfils objective 2 by presents a detailed analysis of the real data used within the thesis. Identification of general patterns within the data and price variation within different geographical areas and station categories are also presented. Classifications for analysis of model results were also derived along with study areas. These will be used within subsequent chapters for detailed assessment of the hybrid and network hybrid models[1] (Chapter 8) performance. In addition, similar empirical techniques to those used within the literature were applied to the data. This work showed both the limitations and unsuitability of using such techniques for modelling the retail petrol market.

Objective 3: Using agent technology to build a model to simulate the spatial and temporal variations in price observed in a single commodity retail market.

The third objective of this thesis covers Chapters 5 and 6. The pure agent model is constructed in Chapter 5 and initial experimentation undertaken with individual parameters and separate rule sets. This an agent-based equivalent to the traditional modelling methods highlighted in Chapter 2 that do not take consumers into account. This model represents an attempt to build a pure agent solution. Despite both operating successfully and performing sensibly, the agent model is limited by the lack of any consumer behaviour.

Chapter 6 details the construction of a spatial interaction model for reproducing the behaviour of consumers. A simple network model was also developed to re-distribute the population on the basis of journey to work data. These additional models are linked to the agent model from Chapter 5 to provide a hybrid and network hybrid model which included more realistic consumer and market behaviour.

Objective 4: Assess model responses to different configurations, initial conditions and rule sets using both idealised and real data.

The fourth objective of this thesis is concerned with testing and validating the hybrid and hybrid-network models. This has already been partly undertaken in Chapter 5, the rest is presented in Chapters 7 and 8. A comparison of the performance of each of the three models is undertaken with robustness and sensitivity testing using both idealised (Chapter 7) and real data (Chapter 8).

Chapter 7 presents the experimentation with idealised data. By standardising the geography and population, the behaviour and sensitivity of the hybrid[2] model were tested through a series of experiments. These experiments were designed to assess whether the model would produce variations similar to those observed within the real market. Additional experimentation examined the sensitivity of the system to small changes in individual parameters. This work also served to validate part of the model performance.

[1]The hybrid model is the spatial interaction model linked to the agent model. The network hybrid model is a combination of the network, spatial interaction and agent model.

[2]As the population and geography were idealised, the network part of the model within this chapter was redundant.

Within Chapter 8 the performance of the pure agent, hybrid and network hybrid models are compared. The network hybrid model is used to perform more validation of the model by further experimentation with data from different geographical and temporal periods. Diffusion experiments are also performed to examine the sensitivity and robustness of the system when using the real data.

Objective 5: Investigation of objective techniques to select optimal values for the parameters in the agent model.

The fifth objective of this thesis is covered by Chapter 9. This chapter reviews various optimisation strategies and justifies the choice of a genetic algorithm (GA) for this application. The development of a genetic algorithm to couple to the hybrid and network hybrid models is detailed. This GA produces optimal values for each of the parameters used within the model. The optimal solution produced by the GA is compared with those derived in Chapter 8 and further tested on data from different geographical and temporal periods.

Objective 6: Provide recommendations for future work in the form of a research agenda.

The final objective of this thesis is covered by the last chapter. In Chapter 10 the main research findings are summarised and a set of generic guidelines for the application of hybrid agent-based modelling methods to geographical systems are presented. The chapter concludes with a set of recommendations for future research.

Chapter 2

The Retail Petrol Market

2.1 Introduction

One of the objectives of this thesis is to develop a model that can accurately represent the processes and dynamics of the petrol market. A vital component of this is a thorough understanding of the behaviour of retailers and consumers as well as other processes that may influence petrol prices. The first half of this chapter is concerned with presenting a breakdown of the petrol market, its structure, components of pump prices and published information on consumer and retailer behaviour. The second half of this chapter will provide a context to this research by reviewing the current literature on spatial and temporal price variations and spatial competition. The chapter will conclude with a discussion of the scope and limitations of the methods reviewed for modelling the retail petrol market.

2.2 The Retail Petrol Market

Petrol is one of the most valuable oil derived commodities, valued by retailers and customers alike. Despite pressures on natural resources there is a rising demand for petrol associated with an ever increasing individual mobility. At the end of 2002, there were 11,707 sites retailing over 36 billion[1] litres of motor fuel in the UK. This equates to an average of approximately 1,350 litres of fuel consumed by each car and van per annum. Consumers are becoming ever more aware of petrol prices. Internet sites such as the AA Petrol Busters (AA, accessed 2004) enable the consumer to have almost perfect knowledge of prices within their area. This has created a highly competitive and sensitive market, with organisations employing various strategies to maximise profits. This sensitivity to petrol prices was fully borne out in the UK during August - September 2000 with the "Petrol Crisis" and the associated fuel protests that were precipitated by soaring fuel taxes.

2.2.1 Structure of the Retail Petrol Market

The petrol retailing market in the UK is tightly controlled by a small number of companies. The "major players" can be distinguished by a high degree of vertical integration, i.e. activity at all

[1] Although a UK billion is a million million, the figure quoted here is a US billion, a thousand million (Source: OED).

stages of the petroleum supply chain, from exploration and drilling to extraction, pipelines, refining and processing. The sale of petrol at either company-owned forecourts or franchise/licence agreements is the public face of these companies' operations.

In 1999 - 2000 there were three "players" that dominated the UK petrol scene, Esso, BP and Shell (Table 2.1). Between them, they accounted for a combined share of 47% of volume sales. Texaco, Jet and Total/Fina/Elf formed a second tier within the market giving these five companies 62% of the retail petrol market. In addition to the standard petrol retailers, supermarkets and hypermarkets have a 26% share of volume sales in 2000, compared to 20% in 1996. The remaining 8% of the market was made up of smaller operators and independents. Of these, the most notable are Save and Q8, both with an estimated 1.5% of sales in 2000 (not tabulated), and Repsol and Murco, who both have 1% (not tabulated).

	1996	%	1998	%	2000	%	% Change
ESSO	6.20	17	6.66	18	6.65	18	+7.3
BP/Mobil	6.20	17	6.29	17	5.55	15	-10.5
Shell	5.11	14	5.55	15	5.18	14	+1.4
Texaco	2.55	7	2.99	8	2.96	8	+16.1
Total/Fina/Elf	-	-	-	-	2.95	7	n/a
Jet (Conoco)	-	-	-	-	1.48	4	n/a
Supermarkets/ Hypermarkets	7.29	20	8.14	22	9.61	26	+31.8
Others	9.12	25	7.40	20	2.96	8	-67.5
Total	36.47	100	37.02	100	36.97	100	+1.4

Table 2.1: Petrol and diesel retailers' brand shares in the UK, 1996 - 2000. Figures are in millions of litres. (Source: Mintel Intelligence Group Report, 2003).

2.2.2 Components of Petrol Pump Prices

There are several factors that control the price of petrol at the pump before any commercial pricing strategy is enacted. These include government tax and duty; the cost of petrol on the open market (crude oil prices and exchange rates) and the costs and profits of the wholesaler and retailer. The relative influences that these factors exert will be briefly reviewed in the following sections.

Fuel Tax and Duty

There are two purposes to the road fuel tax and duty in the UK. The first is an attempt to change travel behaviour i.e. reduce the amount that people use their cars in order to protect the environment; the second is to raise revenue. In 1999-2000, fuel duties (excluding VAT) raised 22.3 billion pounds in the UK. This represented 6 per cent of total government revenue (HM Treasury, 2000). Together, duty and tax currently make up approximately 75 - 80% of the pump price.

Cost of Petrol on the Open Market

In the Competition Commission (1990) report, BP, Esso and Shell were all asked to comment on the relationship of the petrol pump price and internationally-traded Brent crude oil. Each company acknowledged that there was a strong relationship between the two (only BP said that this relationship was direct). Both Esso and Shell stated that the relationship was close but not exact being subject to supply and demand pressures, time-lags and other competitive forces at various stages of the supply chain.

Quantifying this relationship between petrol prices and crude oil has been the focus of numerous studies. There is still no universal agreement on the precise relationship, but it is generally agreed that while an increase in crude oil prices quickly leads to an increase in petrol pump prices, a decrease in crude oil prices leads to a gradual decrease in petrol prices. This phenomenon is known as the *"rockets and feathers"* effect (Bacon, 1991). This effect will be examined in more detail in §2.5.2.

Exchange Rates

Crude oil is generally traded in US Dollars. The US Dollar/Pound Sterling exchange rate has varied between $1.40 = £1 and $1.67 = £1 since 1998. This variation alone would cause a 20% fluctuation in the pre-tax price of petrol in the UK when expressed in pence per litre. In times of fluctuation, the influence of exchange rates can therefore be considerable.

Wholesaler and Retailer Costs

Petrol retailing is an extremely competitive business. Breaking down a petrol price of 74.9p; 57p accounts for the duty and tax, 13.9p covers the cost of the product and the remaining 4p goes to the retailer, referred to as the gross margin (Mintel Intelligence Group Report, 2003). From this 4p, the wholesaler has to cover the costs of transporting the product from the refinery to the distribution terminal and storage. The retailer also has to meet such costs as running the site, employing staff and marketing.

2.3 Retailer Behaviour

The petrol industry is highly commercial with little published information detailing company strategies or profit margins. One of the central aims of this research is to build a model that can replicate the main processes and drivers of this system. This lack of information is a limitation that will be addressed at several stages throughout this thesis.

However, there are two sources of published information that can be exploited. The first is the Competition Commission's (1990) report covering all aspects of the petroleum industry. Despite being published in 1991, this weighty document contains invaluable information on price setting, sensitivity and possible strategies that retailers employ. The second source is the recent research by Ning and Haining (2003). Part of this work involved surveying both retailers and consumers about their respective behaviour.

The following sections will use these two studies to provide a synopsis of retailer behaviour. Consumer behaviour will be reviewed in §2.4.

2.3.1 Local Competitive Environment

When questioning retailers about the most important factor in setting their prices, Ning and Haining (2003) found that local competitor price levels were the most important influence. In a survey of 42 stations in and around Sheffield over 85% of stations said that they had one or more reference stations for setting a price. For 83% of stations, this reference station was the nearest neighbour, normally located on the same road. A similar finding was made by the Competition Commission report. They found that 42%[2] of petrol stations had competition located within 1/4 mile; 60% had competition within 1/2 mile and of these stations, 82% said that they knew their nearest competitor's price.

Examining price differences between neighbours (defined by a straight line distance) showed that in 60% of cases, the differences were less than 1p and in 82.7% of cases, differences were less than 2p per litre (Ning and Haining, 2003). Few stations had price differences of 3p and where these did occur, neighbouring stations were not located on the same road.

Supermarket stations were found to take other supermarkets as their main reference. One site manager commented:

> "If in a 3-4 mile radius there was another supermarket, this would be the main reference station." (Site manager interview, Ning and Haining, 2003).

Some multinationals consider supermarkets as a basic reference in price setting:

> "Esso stations aim to be within 2p per litre of Safeway." (Site manager interview, Ning and Haining, 2003).

However, if supermarkets are located at a distance, neither multinationals nor minors (smaller chains of stations and independents) consider their price in setting their own.

2.3.2 Influence of Locality

The location of a petrol station was identified by both studies as exerting a significant influence on price. For example, according to the Competition Commission's report, a typical rural station has an average throughput of 1 million litres per year. This is in contrast to 4 million litres at a station located on an average busy road site and 8 million litres at a busy supermarket. Stations with these lower throughputs require higher margins to cover operating costs such as increased delivery charges. This results in the setting of higher pump prices to recover overheads.

The amount of throughput is directly related to the number of potential consumers that stations have. Rural stations have a much smaller market than urban stations. Analysing potential consumers in urban areas, Ning and Haining (2003) found that supermarkets have the greatest number of potential consumers. Many of these consumers visit supermarket petrol stations as part

[2] The sample size is not given.

of a "one-stop shop" or because of the incentives and discounts offered linked to their grocery purchase. As a result of this, supermarkets were found to have the lowest median price and a narrower price range than other stations. The median price of petrol stations on main roads (in urban areas) was found to be lower than those on minor roads. City centres were also found to have the narrowest price spread as most sites charged the same price.

2.3.3 Price Setting

For many brand name and franchised corporations, it is the parent company that is responsible for pricing on the forecourt. However, in the case of supermarkets, distributors have no say and the pricing decisions are taken by the head office. For example, the Morrisons supermarket chain buys its petrol from Shell, but prices are not influenced by the distributor:

> "Even if Shell's petrol price goes up, we would stick at the lower price." (Site manager interview, Ning and Haining, 2003)

The Competition Commission report questioned several petrol brands about price setting. BP commented that prices were determined by competition, the effectiveness of which was shown by the low levels of profitability earned. They added that:

> "...individual wholesalers find it difficult to raise their prices above the level set by market forces because to do so, would lead to a sharp loss of sales volume. No company wants to put prices up because of the potential sales loss."

Therefore, when prices were increased, there was a natural tendency to "hang onto prices" as the input costs (typically crude oil prices) fell away. This response was representative of all the stations questioned.

However, neither of the resources surveyed contained any information on possible aggressive company strategies in pricing schemes.

2.3.4 Price Sensitivity

Price is obviously an important factor in determining sales volume and therefore profit. However, at what level would an increase or decrease in price at a competitor cause a station to also increase or decrease their price? Ning and Haining (2003) investigated this and the results are presented in Table 2.2.

	Price change (pence per litre)				Not relevant	Total
	0.5	1.0	1.5	2.0		
Number responding to price increases	6	17	1	6	12	42
Number responding to price decreases	7	15	1	7	12	42

Table 2.2: Reactions to competitors' price increase or decrease per litre, (From Ning and Haining, 2003).

From Table 2.2, it can be seen that most stations ignore a 0.5p per litre price difference and start responding to match the price increase when price differences reach approximately 1.0p per litre. A similar trend can be found in the downward shift in prices. Stations that selected "not relevant" can be divided into two categories; (1) supermarket stations that see themselves as "price leaders" offering the lowest price so other stations follow them; (2) franchised stations, where the site does not share any profit or bear any costs for a price increase or decrease. Site managers in these stations are not allowed to change prices without orders from the head office.

Supermarkets and Loyalty Schemes

Supermarkets are now a significant force in petrol and diesel retailing. In 2000, Tesco was the leading supermarket in terms of petrol station forecourts with 332 sites (in England and Wales), followed by Sainsbury's (223), Safeway (178) and Asda (146). Supermarkets have seen a gradual increase in market share as highlighted in Table 2.3.

Year	% Petrol	% Diesel
1996	21.8	15.4
1997	22.7	16.7
1998	24.0	17.5
1999	25.6	18.7

Table 2.3: Supermarket/hypermarket share of retail deliveries, by volume, 1996 - 1999 (Source: Department of Trade and Industry).

Initially offering established branded forecourts, the establishment of "own-label" forecourts was accompanied with a marketing strategy that focused on undercutting the competition. This ploy was a dual success both attracting customers to purchase fuel and helping to expand sales for the supermarket. This resulted in the major players, for example Esso, BP and Shell responding with enhanced loyalty schemes and price-matching policies, the most prominent of which was Esso's "Price Watch" [3] introduced in 1996. Table 2.4 highlights some of the current loyalty schemes offered by different supermarkets in response.

The effectiveness of such schemes are beginning to decline, apparently driven by a desire to obtain the best price at the pump. This can be seen through the announcement of Asda to abandon its loyalty card scheme and concentrate on promoting lower fuel prices (Independent on Sunday, 10th Feb, 2002).

2.4 Consumer Behaviour

Retailers aim to be as competitive as possible within their local area. The pricing level itself is dependent on external factors such as crude oil prices, type of station (e.g. multinational or supermarket) and location. The motivation behind this competitiveness is profit. This comes from consumers patronising a particular station. However, is a competitive price the main motivating factor for a consumer or do other factors come into consideration?

[3] ESSO promise to match any competitor's price within 3 miles.

Place	Scheme
Asda	No specific price scheme. Competitive prices in line with other supermarkets.
Morrisons	Earn 15 Morrisons' miles every full litre. Collect 4995 miles, get £5 shopping vouchers (equivalent to saving of 1.5p per litre).
Safeways	Petrol Payback: Save 2p per litre when spend £20 in-store; 5p when spend £50; 8p = £75; 12p = £100; 20p = £150.
Sainsburys	Nectar card scheme; earn 1 nectar point for every litre of fuel bought. Vouchers (500 nectar points min) enable different rewards.
Tesco	Earn one clubcard point for every £1 spent on petrol. 150 points = vouchers.

Table 2.4: Examples of petrol incentive schemes operated at different supermarkets (Source: Internet survey, April 2003).

2.4.1 Consumer Sensitivity

Ning and Haining (2003) surveyed households in the Sheffield area and asked whether they thought that there was a difference in quality between brands. Only 20% replied that they believed this to be the case. This leaves considerable scope for consumers to make decisions on where to buy petrol based on a range of other criteria. Factors that rated highly were price and convenience. The sensitivity of consumers was also tested. Households were asked to estimate the price difference per litre that would be necessary for them to switch to buying at a different station. 50% claimed to be sensitive to a 3p per litre or smaller differential (although the amount of this differential was not specified).

This work raises an interesting point about consumer knowledge. To react to prices, consumers would have to have perfect knowledge of the prices within their locality. However Brannon (2003) highlights that as consumers are unable to search costlessly for the lowest price in a market, perfect knowledge is unlikely (see also Dudey, 1990). This statement is, to a degree, invalidated by the rise of internet sites such as the AA Petrol Busters. However perfect knowledge in this context assumes that each household has access to the internet and knowledge of the site.

2.4.2 Consumer Petrol Purchasing

Further work by Ning and Haining (2003) revealed that petrol was most frequently bought as part of a trip to work. Shopping trips and social or recreational trips also accounted for a large proportion of journeys where petrol was bought. Very few respondents regularly bought petrol on school trips or made special trips to buy petrol. When asked where consumers purchase their petrol, the most common answer was the station nearest the consumer's home. No figures were presented on the number of people who buy petrol at supermarkets and no comment was made about the possible influence of loyalty schemes. However, the work of Haining (1986) indicated that consumers do not show high levels of loyalty to particular brands.

2.5 Approaches to Modelling the Petrol Market

The aim of the work presented here is to provide an understanding of the nature and scope of the methodologies used in current and previous research modelling the petrol market. Approaches can be divided into two broad areas; research that seeks to explain variations in price and studies concerned with analysing the impact of competition strategies. The following sections will provide an overview of both these areas using a mixture of theory and applications. Due to the considerable size of literature on spatial regression modelling and spatial competition, the following sections will be tightly focused on the petrol market. For detailed synopsis of spatial regression modelling and spatial competition, the reader is referred to Haining (1985); Norman (1986) and Greenhut *et al.* (1987).

In addition, this work is not meant to be a comprehensive review of all aspects of petrol modelling and will not cover areas that are not directly relevant to this thesis. For example, high frequency cycles (whether there is a weekly or monthly variation within petrol prices) and demand models occupy a substantial amount of the literature and will only be briefly referred to. For further information on these areas, the reader is directed to Castanias and Johnson (1993), Noel (2002) and Eckert (2002, 2003).

2.5.1 Spatial Regression Modelling

Spatial regression modelling is concerned with attempting to explain observed variations in price by seeking empirical relationships with other important variables such as crude oil price and population density. The following sections will present an overview of spatial regression modelling focusing in particular on its applications to the retail petrol market.

Early Work

In early "spaceless" models, spatial price variation in integrated markets was often explained by transport cost differences (Takayama and Judge, 1964). For example, the work of Samuelson (1952) grouped buyers and sellers into regions (the only spatial concession), but attributed price differences to varying supply and demand conditions and transport costs between markets. Other reasons put forward within the literature to explain spatial price variation were market power of producers and imperfect information on the part of the consumers.

Incorporating a spatial aspect allows individual buyers and sellers to occupy specific locations in bounded or unbounded (infinite) space. By taking this approach, sellers can locate in strategic places within the market and in relation to other sellers. Relationships between the market boundary and other sellers can then be defined in terms of either transport networks or straight line distances. As Fik and Mulligan (1991) highlight, these locations, for example, whether situated at the edge of the market or near other sellers, will have price implications. Consumers will vary both in their knowledge of the sellers (in terms of convenience and cost) and in their social-economic profile. Together, these different elements will characterise the geography of the market area and thus influence the magnitude and spatial distribution of price variations (Ning and Haining, 2003).

There has been a considerable amount of work on spatial variation. Faminow and Benson

(1990) showed that evidence for market integration is affected as soon as the market is specified in spatial terms (where sellers compete only with near neighbours and where consumers consider only nearby sellers). Haining (1983a) constructed a supply-side model for spatial pricing in a market in which sellers only compete with near neighbours. This model showed how prices would vary spatially as a consequence. Sheppard *et al.* (1992) developed a model in which consumers had choice sets that underlay the inter-site competitive structure in the market. In a later development of this model, Haining *et al.* (1996) demonstrated how prices vary spatially under different profit objectives given different assumptions about the choice-set structure linking the sites. This choice set also reflected the road network of the urban area, an aspect of spatial competition also studied by Fik (1991a).

Spatial price variation studies are dominated by empirical studies, the most common of which have focused on groceries (see Campbell and Chisholm, 1970; O'Farrell and Poole, 1972; Parker, 1974). However, studying groceries can be complicated as consumers rarely buy just one product. This means that there can be complex trade-offs in both pricing and in what consumers purchase. When explaining price variation, it can be difficult to disentangle product inhomogeneity from effects associated with location and the local competitive environment of the retailer (Ning and Haining, 2003). One of the advantages of modelling the retail petrol market is that it is relatively simple. It is a single homogeneous product and thus does not exhibit any of the trade-offs seen in the grocery market. This homogeneity may also explain why, despite marketing efforts, consumers tend not to show high levels of loyalty[4] (Haining, 1986).

Applications to the Retail Petrol Market

Almost all of the work examining price variations for petrol has stressed competitive factors. For example, Claycamp (1966) showed the importance of local competition effects in a metropolitan area of southeastern USA. Slade (1992) presented similar evidence from his study of petrol prices in Vancouver. Other studies exploring the effects of local competition and price wars on price setting include those by Allvine and Patterson (1972) and Schendel and Balestra (1969). More recently, Haining (1983b) fitted a regression model to petrol price data in southwest Sheffield during an intensive price war and Ning and Haining (2003) examined supply and demand side variables in interdependent retail petrol markets. The study of Haining (1983b) and a similar study by Plummer *et al.* (1998) on St Cloud Minnesota found site and location variables were important. Prices were lowest at sites that concentrated only on selling petrol, had a relatively large number of pumps, sold non-branded petrol and enjoyed a high level of accessibility within the urban area.

Clustering of firms near the centre of a bounded market is often found wherever consumers have limited information on prices, where the acquisition of such information is not costless, and where competition is not intense (Dudey, 1990; Economides, 1993). This characterises many types of retailing activity in urban areas and has implications for spatial price variation. In the context of grocery retail pricing, Fik (1988) remarks

[4] Although there is no evidence, it is likely that people have preferences for supermarkets due to in-store discounts linked to food purchases.

> "...spatial structure and firm density are the predominant forces which shape the geography of price."

The more isolated a retailer, the higher the price. This is presumably because the retailer is providing convenience and overall lower costs to a geographically defined subset of consumers. Fik (1991b) observed that the clustering of retailers promotes lower overall prices and that price patterns reflect the relative location of firms and the distance to the nearest and next-nearest rivals. The larger the number of intermediate sellers between any two sellers the weaker the linkage between their prices indicating more than a simple distance decay effect in the structure of spatial competition (Mulligan, 1989). When assessing maps of petrol retailers Ning and Haining (2003) found that there are typically tight clusters (at intersections of roads), strings (along principal routes) and isolated sites (on side roads). This suggests that petrol retailing only partially meets the conditions that induce clustering within an urban market. When exploring spatial price variation, the petrol market can be viewed as having an added spatial complexity.

2.5.2 Temporal Price Variations

In addition to examining spatial variations, there has also been a focus on the variations in the petrol market over time due to a variety of internal and external factors. Some of these factors have already been highlighted in §2.2.2, for example fuel tax, duty and crude oil prices. Examining pricing asymmetries over time has formed a significant part of the work on temporal price variations. According to Galoetti *et al.* (2003), these studies generally differ in one or more of the following aspects; the country under scrutiny (most often the UK and USA); the time frequency and period of data used; the focus on wholesale and retail gas prices or oil and gasoline prices and finally the dynamic model employed in the empirical investigation.

The hypothesis that petrol prices increase at a faster rate than they decrease in times of turbulence within international crude oil markets was first investigated by Bacon (1991) on the UK market and termed the *"rockets and feathers"* effect. In every recurring period of tension in the price of oil, there has been renewed interest about the level of petrol prices, the magnitude of its cost components (including retailers' margin), and the taxes that contribute to keep those prices high and sluggish. The debate centres on the fact that petrol prices do not decrease at the same rate as crude oil prices.

Manning (1991) further investigated this effect by assessing the impact of changes in oil prices on UK petrol retail prices. The results showed some evidence of this phenomenon. Karrenbrock (1991) examined the empirical relationship between the wholesale and retail petrol prices whilst Shin (1994) related the average wholesale price of oil products to the price of oil; both studies based on data from the US market. More recent studies have focused on the development of the dynamic model and the incorporation of improved data. Borenstein and Cameron (1997) used weekly data to confirm that retail petrol prices react more quickly to increases in crude oil prices than to decreases. Reilly and Witt (1998) re-investigated the UK market emphasising the role of the $ - £ exchange rate and potential asymmetries associated with it. Evidence was found to confirm a positive relationship.

Mitchell *et al.* (2000) took a slightly unusual perspective by investigating the effects of season, holiday and the weather on petrol prices. The results were intriguing; February to May had the lowest prices, no apparent holiday effect was distinguished and a relationship between mood/weather and petrol buying was apparent in two of the four cities investigated. Unfortunately, the degree of influence that inflation may impose is not investigated.

The most recent work on this area was undertaken by Galoetti *et al.* (2003). His work differed from those above by using comparable data to assess different international markets; application of a suitable dynamic regression model allowed distinction between short-lived variations and asymmetries. Finally, unlike Reilly and Witt (1998), the exchange rate was explicitly accounted for. The results found overwhelmingly that the rockets and feathers effect dominates the price adjustment mechanism of petrol markets in many European countries.

The models referred to above are all empirically based using mean price, not variance and disregarding spatial effects. The work of Reilly and Witt (1998) is representative of such approaches. This model was developed for examining the relationship between the net retail price, crude oil price and exchange rate. In its simplest form it can be expressed as:

$$\begin{aligned} \Delta p_t &= \alpha + \beta_1 dc \times \Delta c_t + \beta_2 \Delta c_t + \beta_3 dx \times \Delta x_t + \\ &\quad \beta_4 \Delta x_t + \gamma_1 P_{t-1} + \gamma_2 c_{t-1} + \gamma_3 x_{t-1} + \mu t + \varepsilon_t \end{aligned} \tag{2.1}$$

where:

- C is the crude oil price.
- P is the UK retail petrol price.
- X is the dollar/sterling exchange rate.
- Δ is the first difference operator.
- ε_t is an error term.

The small letters denote the natural log of the variable with a capital letter, e.g. $c = \log(C)$. The set of β coefficients gives information on the short-run effects of crude oil prices, the exchange rates and the net retail price of petrol. Long terms effects are monitored by γ parameters. Two binary variables, dc and dx, are defined to be 1 when the growth in related variable c or x is positive and 0 when it is negative. The coefficients are found by fitting the equation to the data. If the coefficients β_1 and β_3 are found to be non-zero then there is an asymmetry in the price changes. The term μt is a deterministic time trend included to capture the effects of other costs on the retail petrol price. t is the time and μ is a constant.

The research on temporal price variations serves to emphasise an important point: successful modelling of petrol prices cannot be undertaken in isolation (i.e. by just looking at the relationship between two variables); account must be taken of both internal and external factors. The retail petrol market is complex and for accurate modelling, account must also be taken of other factors, for example, company strategies, effects of geography and consumer behaviour.

2.5.3 Spatial Competition

Spatial competition can be defined as the attempts of retailers within a geographical area to be more successful (by cheaper prices, marketing strategies etc) than their competitors in the local neighbourhood. The study of spatial competition has attracted many researchers since Hotelling (1929) published his much quoted paper. The various modifications and extensions of the original model have introduced alternative cost structures (d'Aspremont *et al.*, 1979; Economides, 1986), examined circular and two dimensional markets (Salop, 1979; Tabuchi, 1994; Veendorp and Majee, 1995), markets with non-uniform customer distributions (Tabuchi and Thisse, 1994) and markets with multiple firms (Eaton and Lipsey, 1975; De Palma *et al.*, 1987). Except for a few papers (e.g. Precott and Visscher, 1977; Neven, 1987; Gupta, 1992), most studies on spatial competition assumed that firms choose locations simultaneously, that locations can be changed instantly and with little cost and that market demand is fixed, thus ignoring some important strategic dimensions of the location decision.

A vast amount of the reviewed literature has been directed at location analysis, choosing sites and suitable strategies to optimise profits (see Tabuchi, 1999; Zhou and Vertinsky, 2001; de Frutos *et al.*, 2002; Brekke and Straume, 2004; Camacho-Cuena *et al.*, 2004). These approaches use idealised environments and are typical of many of the models used within spatial competition. For example, they have uniform pricing, consumer distribution and demand. The environment that they compete within is normally a derivative of Hotelling's linear city (competitors equally spaced along a line) or Salop's (1979) circular market (competitors equally spaced around a circle). The work of Zhou and Vertinsky (2001) is unusual in considering some dynamic aspects by looking at a growing market where demand increases with time. Although this is an advancement, this is still a highly idealised problem. Many of these studies are also concerned with the wholesale market (one company selling to another) rather than the retail market (one company selling to many individual buyers).

Research on spatial competition within retail markets generally differs in scope, but works are linked by the following two issues; what determines the equilibrium pattern of locations of firms and what are the properties of the equilibrium prices if there is spatial competition between firms? Two conclusions are generally reached; retail shops have a tendency to be densely located in areas with a high population density and equilibrium prices tend to be lower if there is a high density of sellers (Fik, 1991b; Clemenz and Gugler, 2002).

2.5.4 Spatial Competition and the Retail Petrol Market

In an attempt to place the spatial modelling of prices onto a sounder theoretical basis, concepts have been borrowed from economics. These ideas provide a useful framework within which to begin modelling retail markets. For example, economists usually begin an analysis of a particular market by assuming that it resembles the standard, perfectly competitive model. Here the goods sold are homogeneous, there are many atomistically small buyers and sellers who have perfect price information, and prices are perfectly flexible. The theoretical implications of a perfectly competitive market are that firms will earn no economic profit and operate at an efficient level (Greenhut *et al.*, 1987).

No market contains all of the characteristics of the perfectly competitive ideal (d' Aspremont, 2000). The retail petrol market lacks several of these characteristics. For example, while there are many different buyers of petrol, there are relatively few sellers in a given market. Each station does have some degree of market power, meaning that it is not constrained to charge the same price as its rivals. While the petrol is itself largely homogeneous, firms can distinguish their product in other ways by offering a superior location or services such as cheap car washes or location next to a supermarket. This allows each seller of petrol a modicum of market control.

An important feature of spatial competition is the assumption that each consumer will make purchases at the shop where total costs (consisting of price and transport costs) are the smallest. This is also based on the assumption that the consumer has perfect knowledge. Despite advertising boards and postings, neither customers nor retailers ever have perfect information at any time. Brannon (2003) highlights that consumers cannot costlessly search for the lowest price in a market[5] and firms also expend resources trying to find their competitors prices as well.

Consequently, each shop has a local monopoly whose geographical size depends on the prices charged by the nearest competitors and the transport costs consumers have to incur at different shops in a given area (Clemenz and Gugler, 2002; Ning and Haining, 2003). With increasing distance from competitors, petrol stations are able to charge higher prices (Competition Commission Report, 1990). These higher prices are also dependent on the availability of a local market. This trend is particularly evident in rural areas.

Brannon (2003) highlights that most empirical research undertaken on petrol markets suggests that the primary problem in the market is not predatory pricing, but rather a propensity towards price collusion. Borenstein and Shepard (1993) and Borenstein and Cameron (1997) suggest that in an oligopolistic setting where there is uncertainty about competitors' costs, prices may be sticky downwards because retailers engage in focal point pricing. This argument is a refinement of the "trigger price" model of oligopolistic coordination proposed by Green and Porter (1984). In response to a negative cost shock, a firm may be reluctant to change its price before its competitors have, implying that the old price offers a focal point. Retailers are not assumed to exercise similar restraint after a positive shock. This last mode of behaviour is described as the *"rockets and feathers"* effect (described earlier in §2.5.2).

Modelling the Retail Petrol Market

There are comparatively few studies concerned with modelling the petrol market. One of the exceptions to this comes from the work of Sheppard *et al.* (1998). This study exemplifies how spatial competition methods use highly idealised conditions. Sheppard *et al.* (1998) empirically experimented with different strategies for attaining profit. Firms were uniformly distributed, each with a sub-market (the part of the city which is the nearest to them). Demand was uniform and the share that each firm received dependent on their pricing relative to their neighbours or the other firms within their market. Using this situation, Sheppard *et al.* (1998) assigned different strategies for maximising the rate of profit or the total profit. The results showed that maximising the rate of

[5] Development of web-based price awareness sites for example the AA Petrol Busters (AA, accessed 2004), partly invalidate this statement. Customers are able to search within their locality for the cheapest petrol at little expense or inconvenience to themselves.

profit gave better results than attempting to maximise the total profit.

This approach carries several criticisms. No account was taken of any geographical influences, all the stations operate the same rules and the market is assumed uniform. It would be impossible to apply a model this idealised to a real market.

Another example of using spatial competition to model the petrol market comes from the work of Chan *et al.* (2003). Chan *et al.* examined the location of petrol stations and pricing within Singapore using two empirical models. The island was divided into grids and the potential demand (based on factors such as population, median income and number of cars) was calculated using one model. Stations were located where the demand was greatest. The various parameters were adjusted so the location of the predicted stations matched the actual station locations. The second model assessed the market share. The amount each station sold was dependent on various factors such as distance between stations, market conditions and price. To set the price, one of two possible profit strategies was used. The first strategy allowed each station to maximise its own profit, the second enabled a chain of petrol stations to maximise their profits. The various parameters for different strategies were estimated based on the real data. The conclusion reached was that market share is positively influenced by price, location and number of pumps.

The models presented by Chan *et al.* (2003) and Sheppard *et al.* (1998) are both steady state models. Both make assumptions about the probabilistic distribution of market share based on factors such as price and distance and use this to calculate profit levels. As a result of this approach, both models are specific to a certain set of locations and markets. They can be used to model the impact of a station closing, but could not be applied to dynamic real data scenarios, for example the consequences of a rapid increase in crude oil prices.

2.5.5 Artificial Intelligent Approaches to Petrol Price Modelling

The petroleum industry has been amongst one of the first to use AI techniques, for example DIP-METER ADVISOR and PROSPECTOR are often mentioned as being examples of early classic systems (Alvarado *et al.*, 2004). The focus of these systems are strategic with multi-agent architectures being found in applications where distributed decision-making is advantageous, for example flexible manufacturing control, planning and scheduling, diagnosis, process design, modelling and diagnosis and supply chain management (Braunschweig and Gani, 2002; Sheremetov *et al.*, 2004; Sherstuk, 2004). Of the literature surveyed, there was no published work on the application of artificial intelligence to modelling any aspect of the retail petrol market.

Several conclusions can be drawn from this. The focus of many of the empirical studies was to quantify the relationship between petrol price and one other variable. There are no attempts to model the processes surrounding price setting or interaction between stations in a competitive neighbourhood. This tight focus in empirical studies could simply be a result of a lack of detailed data and suitable computing resources. Whatever the reason for this focus, both data and computational resources have greatly improved and there is now the opportunity to use new techniques to explore previously barren areas.

2.6 Conclusion

The purpose of this chapter has been to present an overview of the retail petrol market and demonstrate how previous research has attempted to model it. A detailed synopsis has been supplied on the structure of the market, the main players and the components that contribute towards petrol prices. With information drawn from Ning and Haining (2003) and the Competition Commission (1990) report on consumer and retailer behaviour, the petrol market can be viewed as being a complex system with many processes combining at different spatial and temporal levels towards the final price. The information gathered provides a firm basis on which to construct a model that is suitable for modelling the intricacies of petrol pricing.

Current approaches taken towards modelling the petrol price market were reviewed. These were presented in two main sections, spatial regression modelling and spatial competition. The limitations associated with each approach were neatly summarised by Plummer (1996) who commented that:

> "Typically, these models are largely confined to aggregate inter-regional and inter-sectoral long-run analyses or are limited to a range of institutional structures in which firms are considered to act as if they were independent entities operating in spatial markets".

For example, in spatial regression modelling, detailed geography and customer behaviour are generally disregarded, whilst the models used in spatial competition are highly idealised with uniform pricing, demand and consumer distribution. None of the approaches reviewed sought to model the full complexities of the system, for example temporal asymmetries in pricing were researched with no account of geographical factors.

Such models as those reviewed are mathematically based and as such present certain general problems (Pave, 1994). For example, empirical models link up parameters that are all on the same scale of analysis. It is not possible to make the behaviours executed at the 'micro' level correspond with the global variables measured at the 'global' level. Equations used within these models are generally complex, containing large numbers of parameters that are both difficult to estimate and lacking realism. These models also tend to be over calibrated to the degree that they are essentially large regression equations with more meaning in the coefficients than the variables. By their nature, mathematical models only consider quantitative parameters. Valuable information can be input to a model by use of qualitative data.

Finally, it is difficult to take into account the actions of individuals and therefore the modifications to the environment which results from their behaviour. As Ferber (1999) highlights,

> "If we consider actions only in terms of their measurable consequences at the global level, or of their probability of appearance, it will be difficult to explain phenomena emerging from the interaction of these individual behaviours, in particular all those relating to intra- and inter-specific cooperation."

Retail markets are extremely important, but economists have few practical tools for analysing the way dispersed buyers and sellers affect the properties of markets. There is a need to both improve and extend current approaches if the retail petrol price market is to be successfully modelled. An alternative technique that will hopefully rectify many of the problems highlighted will be introduced in Chapter 3.

Chapter 3

Agent-Based Systems

3.1 Introduction

Chapter 2 presented an overview of the petrol price market. In particular the characteristics of the market were described along with a review of current modelling approaches. The conclusion was drawn that the retail petrol market is a complex system with various parameters exerting different levels of influence. None of the research reviewed took account of this, rather the emphasis was on modelling the relationship between petrol prices and one other variable, normally crude oil prices. Although the research of Ning and Haining (2003) identified spatial factors as being an important part in determining petrol pricing, almost none of the models took this into account. The empirical basis of many of these models prohibits individual stations being modelled. This means that the impact of different pricing strategies at different geographical scales cannot not be assessed.

Chapter 2 ended with a call for a different modelling perspective for the petrol price market that would rectify the limitations of previous empirically-based approaches. There are a number of software paradigms that are available for modelling complex systems. These include object-orientation (Booch, 1994) and component-ware (Szyperski, 1998). However, when it comes to modelling complex systems as typified by the petrol market, they fall short in two ways. The relationships between computational entities are too rigidly defined and there are insufficient mechanisms available for representing the systems' inherent organisational structures (see Jennings, 2000, for further details).

One technique which avoids many of these problems is the use of agents. Agents are a relatively new paradigm for developing software applications. The origins of the concept evolved from the Concurrent Actor model of Hewitt (1977). This model proposed the concept of self-contained "actors" that communicated with other concurrently executing actors through messages. Their vast potential in designing and building complex systems (Jennings, 2000) coupled with the increase in computing power and the advantages that they offer over traditional approaches has resulted in agent-based models becoming an increasingly popular and powerful tool within geographical applications (O'Sullivan and Haklay, 2000). Current applications can be found within the scope of commercial (shopbots, Greenwald and Stone, 2001), industrial (air traffic control systems, Kinny and Georgeff, 1996) and geographic applications (Transims, (Los Alamos National Laboratory, accessed 2004) and Sprawlsim, (Torrens, accessed 2004)). This section presents a brief overview of the nature of agents, their suitability for modelling complex systems and a brief review of current agent applications.

3.2 Definition

There is no universally agreed definition of an agent (see Franklin and Graesser, 1996, for additional discussion) with researchers continually debating whether definition should be by an agent's application or environment (Goodwin and Wright, 1993; Brenner *et al.*, 1998). With an ever-increasing list of agents appearing (Nwana, 1996), the most useful characterisation comes from Wooldridge and Jennings (1997):

> "An agent is an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives."

Autonomy is a fundamental notion for agents (Castelfranchi, 2001). All agents are inherently endowed with autonomy (Ferber, 1999). Autonomy allows agents to function as an independent unit performing a variety of actions necessary to achieve its objectives whilst responding to changes in both other agents and its environment (Castelfranchi, 2001; Roozemond, 2001). This is extremely useful as it allows the agents to take into account local and updated information providing them a firmer basis on which to make future decisions.

Jennings *et al.* (1998) widened Wooldridge's definition to include the following characteristics:

- **Social ability:** communication with other agents.

- **Reactivity:** perceiving their environment and reacting to changes in it.

- **Proactiveness:** capability to exhibit goal-directed behaviour by taking the initiative.

These characteristics, along with the definition of Wooldridge and Jennings (1997) emphasising autonomy and flexibility, provides a set of attributes that should be present within a single software entity to provide the power of the agent paradigm. They distinguish agent systems from related, and often integrated, software paradigms such as object-oriented systems, distributed systems, and expert systems (see Jennings, 2000, for a discussion). This is not to say that this list is exclusive; within an application each agent could potentially have a different set of attributes but would contain the essential characteristics outlined by Wooldridge and Jennings above. However, some agents may have additional characteristics and for certain types of applications, some attributes will be more important than others (Jennings *et al.*, 1998). This means that there can be many different types of agents within one simulation.

3.3 Agent Architectures

An agent architecture is a particular methodology for building agent systems that specifies how agents may be decomposed into component modules, and how these modules interact with their environment via sensory input (Roozemond, 2001). This definition differs from the architecture of a collection of agents where the focus is on the communication, coordination and cooperation between agents. The most common architecture used is a Multi-Agent System (MAS) (see Stone and Veloso, 1998; Ferber, 1999, for a discussion of other architectures). MAS are concerned with the behaviour of agents aiming at solving a given problem (Jennings, 2000). MAS can be defined as a loosely coupled network of problem solvers that work together to solve problems that are beyond the individual capabilities or knowledge of each problem solver (Sen *et al.*, 1998). These problem solver agents are autonomous and may be heterogeneous in nature.

MAS have several defining characteristics. Each agent has incomplete information, or capabilities for solving the problem (i.e. each agent has a limited viewpoint), there is usually no global system control, data is decentralised and computation is asynchronous. In addition, as Tsvetovat and Carley (2002) have highlighted, these architectures inherently possess the following properties that are invaluable in simulating complex systems:

- **Spatial realism:** Many artificial intelligence (AI) based simulations are built on the concept of agents or cellular automata operating on a grid of a specified shape. Interactions are based on the concept of proximity, defined by distance between agents on the grid. However, the choice of grid shape and type of neighbourhood is often arbitrary and often does not carry any recognition of realism. Within MAS the agents are governed by the formal structure of the organisation and the agents' beliefs about the informal structure. This allows modelling of ideas and concepts to be based on what is observed in reality.

- **Temporal realism:** The majority of simulations are synchronous - based on the idea of time periods and "turn-taking". This provides an adequate approximation of simple interactions, but does not model simulations of protracted interactions. More realistically, real world interactions are of an asynchronous nature. MAS structures allow this behaviour.

- **Information flow realism:** In MAS, agents do not have perfect knowledge about the world. The way for them to obtain knowledge is to ask other agents - or obtain the information via an exchange interaction. This type of behaviour is especially important in a simulation where the agents are competitive.

- **Task realism:** To be successful in modelling a complex system and completing an assigned task, the MAS must accurately represent the processes present in the real system. In modelling emergent phenomena such as market behaviour, a simulation environment that contains large numbers of agents is required. If the agents are too "simple", not possessing enough knowledge of processes within the system, they will not be able to replicate the observed behaviour.

Using a MAS architecture brings the advantages of robustness, efficiency and the ability to solve problems in which data, expertise, or control is distributed (Jennings, 2000). This is of particular use within applications that are geographical (i.e. spatially distributed) in nature.

3.4 Interactions

Using an agent perspective requires multiple agents to represent the decentralised nature of the problem. This requires interaction, either to achieve the objectives or manage dependencies that ensue from being in a common environment (Jennings, 2000; Castelfranchi, 2001). Interactions develop out of a series of actions whose consequences in turn have an influence on the future behaviour of the agents (Ferber, 1999). These interactions can range from simple information exchanges to cooperation, coordination and negotiation in order to arrange independent activities and are a central feature of an agent (Batty *et al.*, 2003). There are numerous ways in which interactions can be organised, for example message-passing systems and blackboards (see Haverkamp and Gauch, 1998; Ferber, 1999, for a detailed discussion). Implementation of these organisations is dependent on the application under consideration. Agents can simply operate by passing or requesting information between themselves as required.

Jennings (2000) conceptualises these interactions as taking place at the *knowledge level* where the interactions form some sort of passing of information. As these interactions form part of some wider goal, there is typically some underpinning organisational context between them which defines the nature of their relationship and consequently their behaviour (Cavezzali and Rabino, 2003). In many cases these relationships are subject to ongoing change. The social interaction between agents means that existing relationships evolve and new relations are created. This flexibility and partial control is quite different from hard-wired engineering that can be found in non-agent approaches.

3.5 Agent Behaviour

Each agent in an agent architecture has internal states and behavioural rules. The internal state could relate to variables associated with the agent, for example, "I am thirsty" while the behavioural rules simply dictate the procedure of behaviour to be taken e.g. "I am thirsty. Go to the tap, fill up a glass of water and drink it". Some of these states can be fixed over the life of the agent while others change through interaction with other agents or with the external environment (Epstein, 1996). This allows many different systems to be represented. For example, shopbots are electronic agents whose primary function is to search for information on products and merchants on the Internet (Kephart *et al.*, 2000), Doran (2003) created agents with beliefs to model the interaction of groups with different religious views and Axelrod and Scott Bennett (1997) used agents to examine cooperation and competition to determine which nations would aggregate into alliances.

Additionally, agents can simply be the facilitators of decisions made by other models or use information from other models to help inform their decision. For example, Sprawlsim (Torrens, accessed 2004) was partly developed as a MAS to help researchers and public planners experiment with ideas about suburban sprawl. The model is built in a hybrid fashion; top-down dynamics are handled by traditional aggregate land-use and transportation models while bottom-up dynamics are simulated with cellular automata (for urban infrastructure) and MAS (representing population dynamics). Beltratti *et al.* (1997) constructed an agent-based model for market simulation that used artificial neural networks to inform the agents about the state of the market. The research of Parsons and Wooldridge (2002) focused on using game theory to allow the agents to make the best decision based on information available to it.

Giving agents the ability to learn allows them to modify their behaviour and maintain their performance in a dynamic environment. Learning can be achieved through symbolic (rule based) or sub-symbolic (connectionist based) mechanisms that define how and what they learn. If an agent exists in a dynamic environment, or is tasked with learning concepts, it may be necessary for the agent to modify its learning behaviour in order to maintain its performance. The work of Holland and Miller (1991) on modelling the stock market emphasised the importance of allowing agents to learn. They used genetic algorithms and classifier systems to generate the best possible strategies for the agents to take. This approach was also taken by Palmer (1994).

3.6 Environments

As the definition of Wooldridge and Jennings (1997) states, agents are situated in some sort of environment. There are many types of environment ranging from distributed to highly centralised (Ferber, 1999) and just as there can be many different types of agent in a simulation, there can also be many environments (Batty and Jiang, 1999). Russell and Norvig (1995) suggested the following environmental properties:

- **Accessible v inaccessible:** Accessible environments are where the agents obtain current information about the environment. Most real-world environments (physical world and the Internet), are not accessible in this sense.

- **Deterministic v non-deterministic:** A deterministic environment is one where any action has a single guaranteed effect; there is no uncertainty about the state that will result from performing an action.

- **Static v dynamic:** Static environments are assumed to remain unchanged, except by the performance of the agent. With dynamic environments, processes other than the agent control the outcome.

- **Discrete v continuous:** Discrete environments have a fixed number of actions and perceptions within it.

It is simpler to construct an agent if the environment is accessible, the quality of the agents depends on the quality of the information available.

The non-deterministic environment captures several important aspects of environment. Firstly, agents have a limited sphere of influence (partial control over their environment) and secondly, actions performed by agents are done with the aim to bring about some desired state of affairs. Non-determinism captures the fact that some agents can fail. Almost all realistic environments are non-deterministic because of their complexity.

Dynamic environments have two important properties for agents. An agent trying to decide on a course of action at time t_1 cannot assume that the environment is the same as it was at an earlier time, t_0. Even if the state of the agent is the same at both times the decision on a course of action may be different because of the different environment. This means that for an agent to perform an appropriate action, it must perform information gathering actions to determine the state of the environment. Wooldridge (2002) argues that static environments are easier to design for because the agents only need to perform information gathering once. If the information that an agent gathers is correct and it correctly understands the effects of its actions, it can accurately predict the effects of its actions. In a static environment, the agent never needs to worry about coordinating or synchronising its actions with other processes.

Discrete environments are seen to be easier to design agents for than continuous environments. This is for several reasons; digital computers are discrete-state systems, they can simulate continuous systems, but have some degree of mismatch - information that a discrete state agent uses in order to select an action in a continuous environment will be made on the basis of information that is inherently an approximation.

3.7 Suitability for Modelling Complex Systems

A crucial aspect of any research design is using the most appropriate model for the problem. There can be many different options, but within software design Jennings (2000) outlines that the most powerful abstractions are:

> "those that minimise the semantic distance between the units of analysis that are intuitively used to conceptualise the problem and the constructs present in the solution paradigm"

Of course, the degree of conceptualisation will depend on the complexity of the system under examination.

Agent-orientated approaches advocate the decomposition of problems into small components. Decomposition is a natural metaphor for real complex systems that are always distributed and have multiple loci of control (Meyer, 1988). Agents are autonomous, they know when they should be acting and when they should update their state. This self-awareness reduces control complexity since the system's control "know-how" is taken from a centralised repository and localised inside each individual problem solving component (Jennings, 2000). Additionally, as decisions are made by the agents about what actions should be performed, this allows decision making to be responsive to the agent's actual state of affairs, rather than some external entity's perception of this state.

The retail petrol market was introduced in Chapter 2. This system can be viewed as being complex. In simplistic terms, changes in petrol prices at a local level are influenced by location and number of competitors and amount of custom. At a regional or national level, influences such as taxation and crude oil prices become significant. Company policies for large chains of stations, for example Esso, are also likely to act at this level. All these factors interact to form the final price.

This type of system lends itself to being represented by an agent structure. As Batty and Jiang (1999) comment, "agent-based simulation is perhaps most appropriate where local spatial operations are the focus"

This is then particularly appropriate for the petrol pricing market where the presence of the local competitive environment (see Chapter 2) appears to be a driving factor. Jennings (2000) emphasises this strong relationship between "notions of subsystems and agent organisations". The agent structure is concerned with the interaction of a number of components and their role in solving the problem. Complex systems also involve changing relationships between different parts of the system; this requires that various components are treated as a single conceptual unit. Using a distributed approach allows individual areas of expertise to be encoded into particular agents, thus modelling the real-world problem in an "intuitive, modular and therefore expandable manner" (Anumba *et al.*, 2002). An agent-mindset provides a suitable abstraction for managing and representing organisational relationships.

3.8 Potential Drawbacks

Agent-based systems have a great deal to offer the modelling of complex real-world systems such as the retail petrol market. However, the nature of the agent paradigm leads to a number of problems that are common to all agent-based applications.

Firstly, there may be no overall system controller (Jennings *et al.*, 1995). An agent based solution may not be appropriate for domains where global constraints have to be maintained, for example, in domains where a real-time response must be guaranteed. Secondly, there are no global perspectives. This is due to an agent's action being determined by its local state. However, since in almost any realistic agent system, complete global knowledge is not a possibility, this may mean that agents make globally sub-optimal decisions.

Decisions about the number, pattern and timing of interactions depend on a complex interplay of the agent's internal state. For example, the agent's perception of the environment and organisational context that exists when the decision is made. A combination of these factors makes it difficult to predict the outcomes of the behaviour. This relates closely to the degree of variability that the agent may request and what the result is; the nature and outcome of an interaction cannot be determined at the onset. Another source of unpredictability in agent-orientated systems is due to emergent behaviour. The end behaviour of agents cannot be deconstructed in terms of the behaviour of the individual components (Ferber, 1999).

However, depending on the application, some of these disadvantages can be avoided by using interaction protocols. These can include mechanisms such as game theory, whose properties can be formally analysed by adopting rigid and preset organisational structures. The aim of this type of procedure is to reduce the unpredictability of the system. However, the drawback is that the power of the agent based approach becomes limited (Jennings, 2000).

3.9 Application Domains

Agents are autonomous entities that can act independently or with other agents depending on their task. They can therefore encapsulate a wide variety of entities from humans to robots to software agents (Batty and Jiang, 1999). This is reflected in the variety of current agent applications. Many examples can be found including air traffic control (Kinny and Georgeff, 1996), planning and scheduling (Sheremetov *et al.*, 2004) and creating special effects in the film industry, for example the Massive program (Massive Ltd, accessed 2004) used to generate battles in the Lord of the Rings films.

The following sections will provide an overview of some of the current applications within geographical and economic research areas.

3.9.1 Geographical

The use of agent-based systems in geographical applications is ever increasing. Unlike the earlier generation of quantitative models in geography (for example spatial interaction models), software development is guiding agent formulation (Batty and Jiang, 1999). There has been, and is still ongoing, a considerable effort to provide easy to use software platforms. Tobias and Hofmann (2004) critiqued numerous platforms and produced a list of four that they felt were sufficiently optimised towards social science applications. These platforms are RePast (University of Chicago Social Science Research Computing, accessed 2003), SWARM (Swarm development group, accessed 2003), Quicksilver (Burse, accessed 2003) and VSEit (Brassel, accessed 2003). Each of these platforms allowed agents to be modelled as free and complex objects (for other details on selection criteria see Tobias and Hofmann, 2004). However, despite the increasing usability of these systems, many applications use custom built agent-systems. This is partly due to the absence of one widely accepted platform and the difficulties associated with customisation for the requirements of different applications.

The main bulk of agent research within geographical applications can be found within the scope of "human" geography. For example, Schelhorn *et al.* (1999) developed STREETS, an agent-based model of pedestrian movement in urban environments utilising the SWARM platform. Research at the Centre for Applied Spatial Analysis has produced models for simulating the movement of pedestrians around the Tate Gallery and Notting Hill carnival (CASA, accessed 2004). The work of Beneson *et al.* (2001) used an agent model to examine the movement of households within Tel Aviv for socio-economic reasons. Barros and Sobreira (2002) adopted a similar approach by using an agent-based model to focus on the process of consolidation of inner-city squatter settlements within a peripherisation process.

The concept of simulating patterns at an individual level is not solely restricted to pedestrian movements and migration patterns. The TRansportation ANalysis SIMulation System (Los Alamos National Laboratory, accessed 2004) models, at an individual level, simulations of traffic flow. Its overall aim is to produce a more realistic simulation to enable changes to transportation networks. In a similar manner, the Urban Traffic Control (UTC) system developed by Roozemond (2001) used agents to adjust traffic control systems.

In addition to these applications, agent-based models are also being used to model processes at large or multiple spatial scales, thereby allowing researchers to experiment with different ideas. Sprawlsim (Torrens, accessed 2004) is such a tool, developed to enable researchers and planners to experiment with urban sprawl at a city scale. Evans and Kelley (2004) developed an agent-based model to perform multi-scale analysis of landcover change and Parry *et al.* (2004) are currently developing an agent model to simulate insect movements over varying geographical scales.

3.9.2 Economic

One of the niches in which the use of agents is rapidly increasing is that of electronic commerce (Crabtree, 1998; Guttman *et al.*, 1998). Here, economic applications involving agents can be broadly divided into two main areas; the exchange of information and bulk market simulations.

Kephart *et al.* (2000) claims that within the next decade, the Internet could be populated with billions of agents exchanging information goods and services with one another and people. In these applications, agents help to "grease the wheels" that must turn in order that goods and services can be bought and sold across the Internet (Parsons and Wooldridge, 2002). The most popular and well researched class of wheel-greasing agents are shopbots, agents which search the Internet on behalf of consumers, comparing prices across dozens of web sites (Kephart *et al.*, 2000; Markopoulos and Kephart, 2002). Shopbots are valuable to both consumers and suppliers alike. For consumers, costs (in terms of both time and money) are cut, for suppliers the costs of advertising etc are reduced. As a result of the success of these agents and their potential to affect the ways markets operate, further research has investigated the impact of shopbots in single commodity markets, modelling the behaviour of both buyers and sellers using game theory techniques (Kephart *et al.*, 2000).

The use of game theory to analyse complex interactions in MAS is an area that, although not directly relevant to the work within this thesis, should be briefly alluded to. Game theory is a branch of mathematical analysis developed to study decision making in conflict situations. There is a considerable amount of work on developing optimal game plans and strategies for electronic markets. In particular, this research has been directed at auctions (Tesauro and Bredlin, 2001) and dynamic pricing (Greenwald and Stone, 2001). Typically, this research uses multiple complex games to produce optimal strategies (Walsh *et al.*, 2002). This is only a small part of the ongoing research. For more information on the application of game theory, the reader is referred to Kreps (1990). For information on applications with MAS, the work of Kraus (2002) provides a good introduction.

With the increase in popularity of economic agents such as shopbots, there has been a corresponding increase in the number of electronic "market-places" and related research into electronic agents. Table 3.1 summarises the main functions or roles of such agents and the corresponding business examples or research projects.

The information presented in Table 3.1 shows that the major market players of electronic commerce are buyers, sellers, and intermediaries. Within these applications, agent technologies are applied to a range of aspects of market places including using agents to model demand, sales and the exchange of information.

The main literature within this application area is concerned with electronic commerce. As such, these applications model the market aspect, but neglect spatial effects. In contrast the geographical applications reviewed are successful at modelling the spatial aspects of the given problem. However, there were no geographical examples that involved retail markets. Finally, there are no examples within the literature surveyed of economic agent-based models that take into account spatial and temporal effects.

	Main Function or Roles
1	**Search** for information on products and merchants; Mysimon.com, Auctionwatch.com, Shopbinder.co.kr
2	**Filter** unimportant commercial messages; Adsubtract.com
3	**Gather and Analyse** information on customers and merchants; Ffly.com
4	**Match** qualified commerce party; Kabash, Fastparts.com
5	**Notify and Push** event; Digitialimpact.com
6	**Monitor and Report** creation, change, and deletion of information; Netmind.com, Books.com
7	**Support Communication** between business and customers; Artificiallife.com,
8	**Personalise and recommend** interface, contents, products and services; Personalogic.com,Technoagent.co.kr,My.yahoo.com
9	**Make or support a decision** on bidding, pricing and negotiation; Auction.bot, eMediator
10	**Network** among consumers, merchants and manufacturers; Napster.com, OPEN4U.co.kr

Table 3.1: Main functions and examples of agents in electronic commerce from Jin and Lee (2001)

3.9.3 Hybrid Agent-Based Systems

The agent framework by itself provides a powerful methodology for modelling complex systems. However, when combined with other techniques, this framework can be enhanced. MAS can be combined with both traditional and artificial intelligence (AI) techniques as desired. There are two strong reasons to hybridise a MAS. Firstly, to improve the strategies implemented by the agents through learning and secondly to incorporate well known system behaviour such as market processes. For these reasons, there is considerable ongoing research into exploiting these hybrid systems. Several different applications are briefly highlighted here to provide an idea of the scope and nature of these applications.

There are many examples of the use of additional techniques (both traditional and AI) for improving the learning of agents. Holland and Miller (1991) and Palmer (1994) used genetic algorithms (GAs) and classifier systems to generate the best possible strategies for agents to take. Beltratti *et al.* (1997) employed artificial neural networks (ANNs) to enable his agents to learn about markets more quickly. Parsons and Wooldridge (2002) further developed this idea by incorporating game theory into their model to enable agents to make strategic decisions based on the available information. In a similar vein, He *et al.* (2002) employed heuristic fuzzy rules and fuzzy reasoning mechanisms in order to determine the best bid for an agent to make given the state of the marketplace.

Incorporating specific behaviour usually involves hybridising the agent model with a "traditional" technique. Within geographical applications, there are few applications that use traditional models alongside agent technology. Haff (2001) developed an "intelligent" model of hillslope development linking empirical models to agents. Sprawlsim (Torrens, accessed 2004) used traditional landuse and transportation models as part of its methodology. These traditional models are often used because they represent well understood dynamics of individual processes. They can just be dropped into the model, leaving the agents to represent the more complicated and less understood interactions between different parts of the system. The sparsity of hybrid agent models within geographical applications indicates an area ripe for further development.

3.10 Conclusion

A review has been presented of agent-based systems and their suitability for modelling complex systems. Through the brief summary of current applications, their adaptability and flexibility have been demonstrated. These characteristics, coupled with the ability to model at an individual level and the autonomous behaviour of the agents, make agent-based systems a powerful tool for modelling the retail petrol price market which overcome many of the problems associated with empirical models (see Chapter 2).

Of the architectures reviewed, a multi-agent system (MAS) provides the most natural metaphor to use for simulating the processes in the petrol retail market. Within a MAS, independent agents (petrol stations) can be created and assigned individual rule sets (corporate policy). Co-operation between agents for sharing information (reacting to local prices) can also be built in as required. Ideas of spatial variability, local competition, autonomous behaviour and imperfect knowledge are naturally incorporated in a MAS and so this makes an ideal framework for this application. These ideas will be developed further in Chapter 5 where the construction of the agent model is presented

Chapter 4

Real Data Analysis and Traditional Modelling

4.1 Introduction

The aim of this thesis is to examine the ways in which agent-based models can be applied to modelling a dynamic, locally interacting retail market. The review of the literature performed in Chapter 2 revealed that due to the commercial nature of the petrol price market, there is very little published information on pricing strategies or factors that are taken into consideration when setting petrol prices. This creates a situation where there is a great deal of data, but no information on potential strategies; the only exception to this is the Esso Price Watch (see §2.3.4). The most useful indicators come from the Competition Commission report on petrol pricing (Competition Commission, 1990). Part of this report questioned petrol retailers on the factors that played a part in setting prices. In their responses they highlighted movements in international prices, changes in competitors' prices and the experience and judgement of those making decisions. Further research within the report identified distance to nearest competitor, price differences between neighbours, awareness of competitors' prices and locality as important factors. Almost all of these factors can be quantified through close examination of the real data. However, without detailed information, it is virtually impossible to model an individual's decision to change the price.

It could be speculated that this lack of information about the processes that contribute to the setting of petrol prices has shaped current research into petrol price modelling. Certainly, the bulk of research has centred on using empirically based models, generally regression, to investigate the relationship between price and one other variable, normally crude oil prices. Part of the work within this chapter will be to extend this regression modelling using variables flagged as important from both the Competition Commission's report and real data analysis.

This chapter will be divided into three main sections. The first section will provide a preliminary examination of the data and construction of petrol station classifications. These will be used for further investigation in the second section and in subsequent chapters. The second section will provide a thorough investigation of the real data quantifying important factors cited by the Competition Commission's report. The final section will use this information in the construction of a regression model.

4.2 Data Preparation

4.2.1 Description of Data Set

The data set was provided by GMap[1]. It consists of daily petrol price readings taken throughout the months of July, August and September 1999. Geographically, the data set covers the UK and includes the main petrol retailers, for example internationals such as Esso, BP, Texaco and Shell; supermarket garages such as Sainsbury's, Asda, Tesco and Morrisons, and numerous "independents". The prices of the four main petrol types (unleaded, super unleaded, diesel and four star) are recorded; in total, there are over 16,000 data points.

4.2.2 Data Selection

The geographical coverage of the data set is quite substantial. Study areas will therefore be selected for exploration. These will be West Yorkshire (Figure 4.1) and a larger area comprising of West, South and North Yorkshire (Figure 4.2). This area will be termed the Yorkshire region[2]. The bulk of the analysis will be performed upon West Yorkshire with further validation carried out on the Yorkshire region. The Yorkshire region was chosen because of the geographical differences between counties. For example, both West and South Yorkshire have motorways and a mixture of urban and rural areas. In contrast, North Yorkshire contains no motorways and is predominantly a rural county. These differences will be further examined in §4.5. Using both the West Yorkshire and Yorkshire regions in experimentation will provide suitable tests of the model behaviour, robustness and sensitivity of parameters as well as helping to provide validation of the model.

Each petrol station can potentially sell four types of fuel (four star, diesel, unleaded and super unleaded). Unleaded is the largest data set; it is sold at every petrol station within the study area (Figure 4.2). Simulations will concentrate solely on unleaded petrol; other fuel types will not be examined.

4.3 Visualisation using GIS

The data set is both spatial and temporal. To understand the patterns and processes occurring suitable visualisation techniques need to be employed. The geographical information system (GIS) software of ArcGIS will used be to used to map results. In addition, a suite of statistical tests will be used to quantify the model results. These will be further detailed in §5.3.7.

The use of a GIS in a project containing spatial information is very helpful. High quality maps containing additional layers of contextual information can be produced for interpretation. A second advantage that a GIS package such as ArcGIS can offer is the creation of an interpolated price surface. The motivation for using interpolated price maps is to create a surface that may be easier to interpret than different coloured points on a map. It also gives some idea, albeit qualitative, of the petrol prices that a consumer is likely to pay in one particular area. The GIS can then be used to overlay several maps and produce an overall image (in ArcMap) of the area with

[1] A GIS consultancy company based in Leeds: www.gmap.com

[2] Humberside/East Yorkshire was not included as North, West and South Yorkshire possessed all the characteristics required for further analysis. These were a large geographical coverage and a mixture of rural-urban areas.

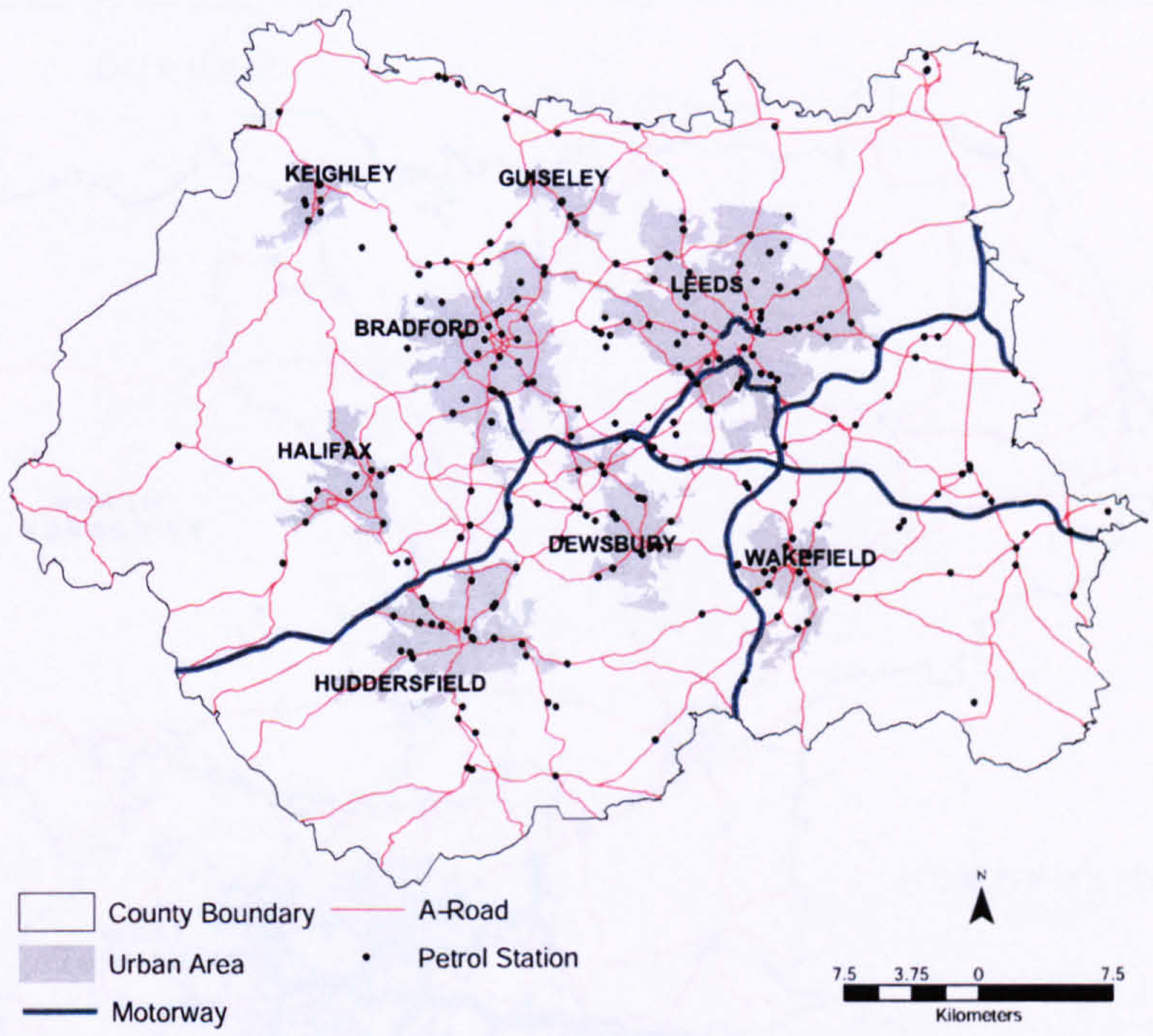


Figure 4.1: Location of urban areas, motorways and spatial distribution of petrol stations selling unleaded petrol within West Yorkshire.

useful information such as location of petrol stations, motorways and A-roads. An example of one of these maps is presented in Figure 4.3.

The interpolation technique used was Inverse Distance Weighting (IDW). Experimentation revealed that visually, the number of points used in the interpolation did not effect the appearance of the final map. As can be seen from Figure 4.3, the price surface does not cover the entire study area. It was decided that only the points that fall within the study areas would be used for interpolation. The main criticism with taking this approach is that the system has many drivers, for example, local competition and population density. Stations located at the edge of the study area will not have the same degree of influences as those located in the centre. However, these stations are typically situated in rural areas and would therefore not be expected to exert considerable influence. An approach that could have been taken would be to create a buffer of a given size around the study area. This would provide the edge stations with a higher amount of influences. However, as the system is dynamic and the degree of the system drivers cannot be easily quantified, the question would then become how large a buffer should be created. The effects of only using the petrol stations within West Yorkshire will be investigated in §5.4.2.

4.4 Real Data Analysis

The following sections present the results of analysis performed on the real data. In building a model to represent a system as complex as the petrol market, it is important to have as much knowledge about the characteristics of the data as possible to avoid assumptions and generalisa-

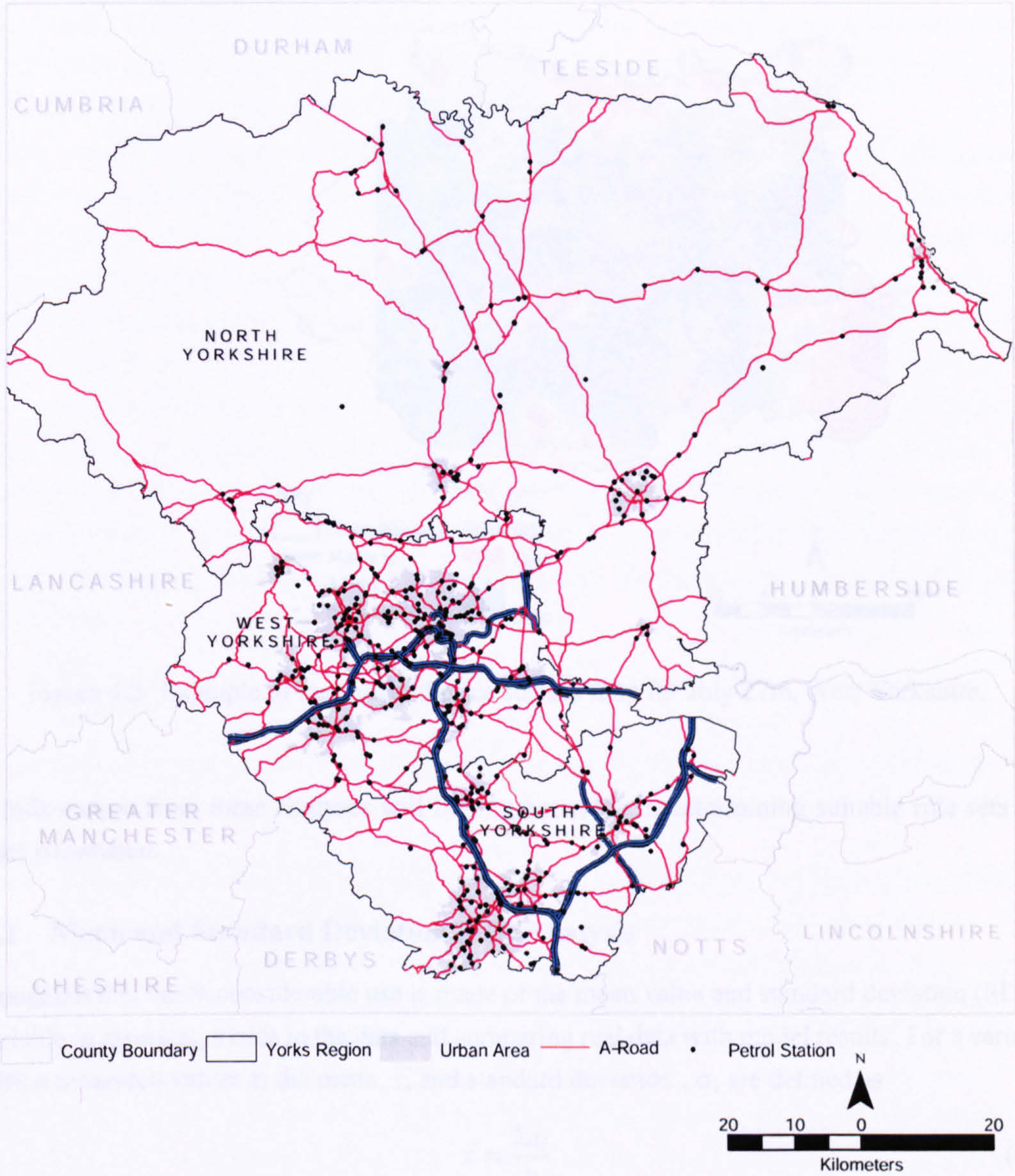


Figure 4.2: Location of urban areas, motorways and spatial distribution of petrol stations selling unleaded petrol within the Yorkshire region.

tions from having to be made. Chapter 2 presented known information about the petrol market detailing factors that are influential in the setting of petrol prices. These included the effect of local competition and the influence of locality.

The analysis undertaken within the following sections will be on two levels. The first preliminary analysis will be aimed at identifying broad patterns and possible limitations within the data. Classifications will then be developed and discussed. These will be used within the second, more rigorous section of analysis. This will aim to formalise a number of the influences that have been identified as important. This will include an analysis of price variations within the real data as well as calculation of the metrics representing price change, distance to neighbours and number of neighbours.

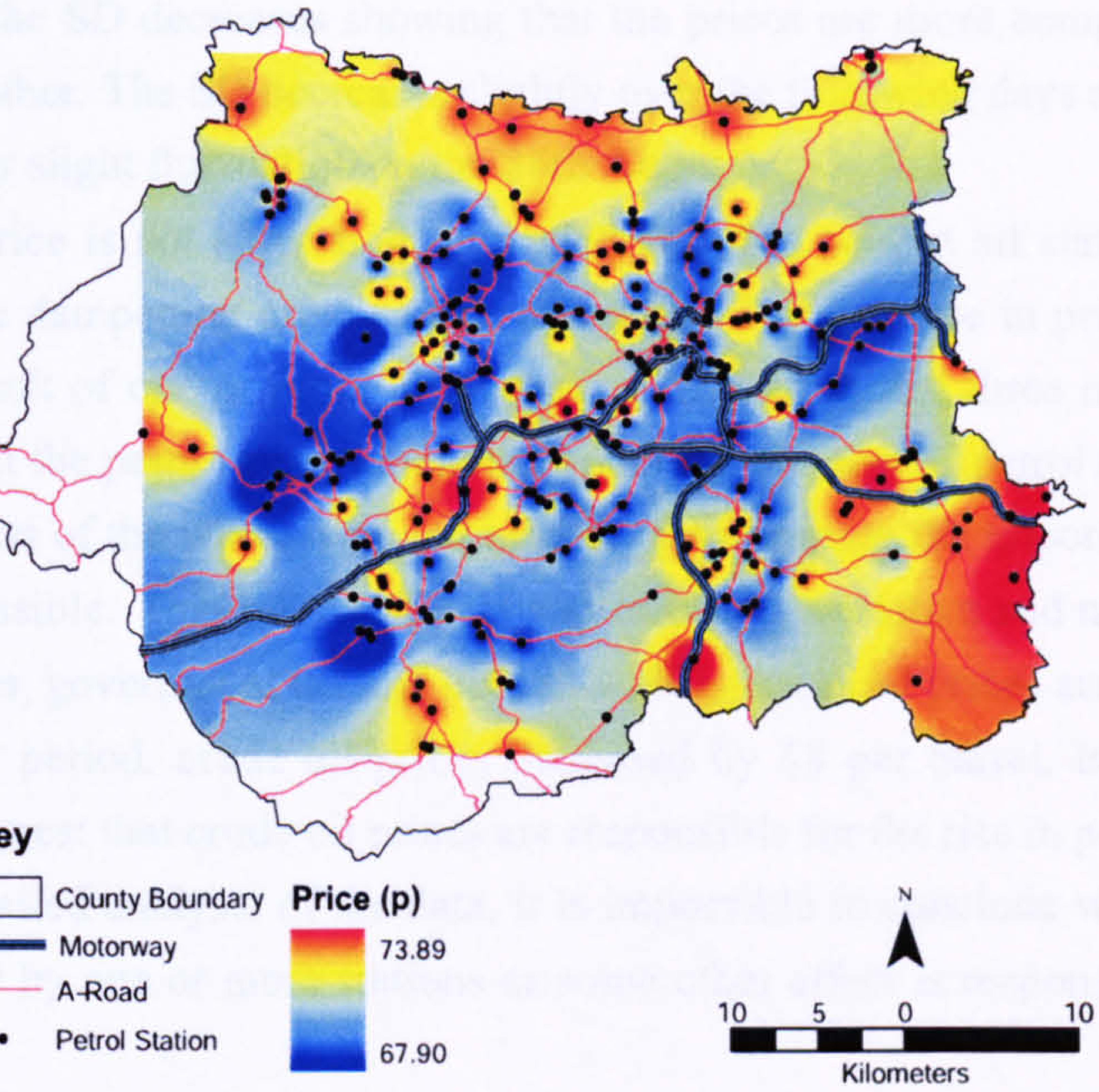


Figure 4.3: Example of interpolated price surface map for July 27th, West Yorkshire.

Information from these analyses will be of great value in determining suitable rule sets and model parameters.

4.4.1 Mean and Standard Deviation (SD) Analysis

Throughout this thesis considerable use is made of the mean value and standard deviation (SD) of a variable in assessing trends in the data and comparing real data with model results. For a variable x with n measured values x_i the mean, $\bar{x}$, and standard deviation , σ, are defined as

$$\bar{x} = \frac{\Sigma x_i}{n} \tag{4.1}$$

and

$$\sigma = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{(n-1)}}. \tag{4.2}$$

The mean price of a given area gives an overall metric of the pricing in that area. Figure 4.4 shows that over time, the mean prices for both (a) West Yorkshire and (b) Yorkshire region follow the same pattern. The initial trend of the mean price decreases over the first few days (27th July - 30th July) from 71p to 70.5p. This is followed by an increase of approximately 2.5p over the period 5th August - 19th August with the prices peaking around the 23rd August at 73p. The prices then remain almost constant (with only minor fluctuations) until the end of the data set. The initial standard deviation (SD) on 27th July for West and Yorkshire region is large, indicating that at the beginning of the period, there is a larger spread of prices. This suggests that the stations in a neighbourhood are not being very competitive in pricing or that there are large variations in price between different parts of the study area. This variation may be the result of the system being in a state of flux with not all stations changing their prices on the same day. After the price increase

(5th August - 19th August), the SD decreases showing that the prices are more competitive, i.e. within a small range of each other. The SD decreases slightly over the following days as the prices gradually settle before another slight fluctuation on the 26th August.

The consistency of the price is not surprising; unleaded petrol is sold at all stations and is therefore more reactive to the dampening effects of competition. The increase in price over the study period could be the result of one or more influences. §2.2.2 identified three main factors that control the price of fuel at the pump; government tax and duty, the cost of petrol on the open market and the costs and profits of the wholesaler and retailer. Determining the importance of the latter of these factors is impossible. This information is commercially sensitive and not available in the public domain. However, government tax and duty as well as crude oil prices are published. Over the course of the study period, crude oil prices increased by $8 per barrel, tax and duty remained static. This does suggest that crude oil prices are responsible for the rise in petrol prices. However, without a more detailed analysis of the data, it is impossible to conclude whether this, an aggressive pricing strategy by one or more stations or some other effect is responsible for the price increase.

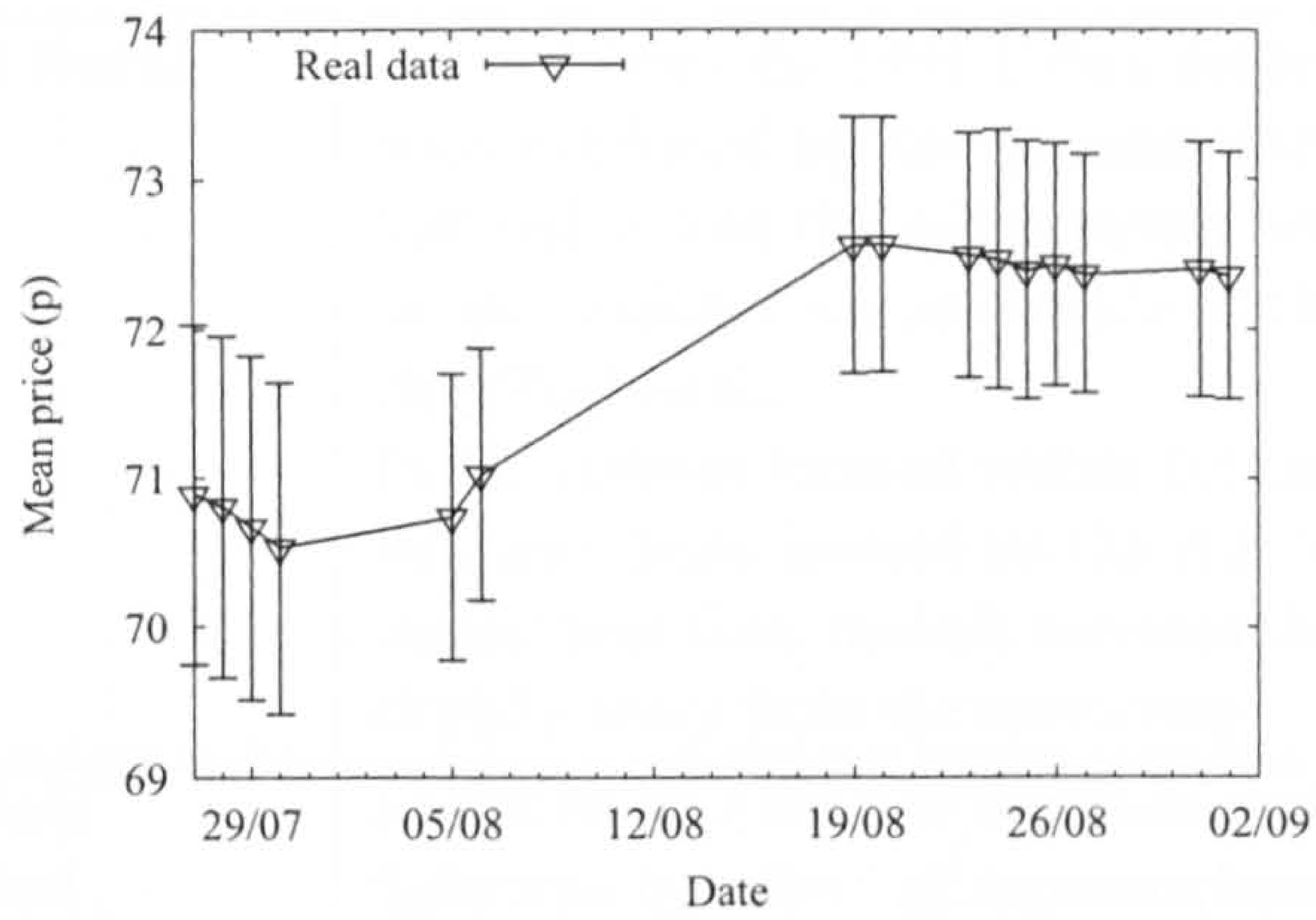
4.4.2 Limitations within the Data Set

During initial analysis, several limitations with the data set became apparent. These could have important influences on model development and results. These limitations are summarised below:

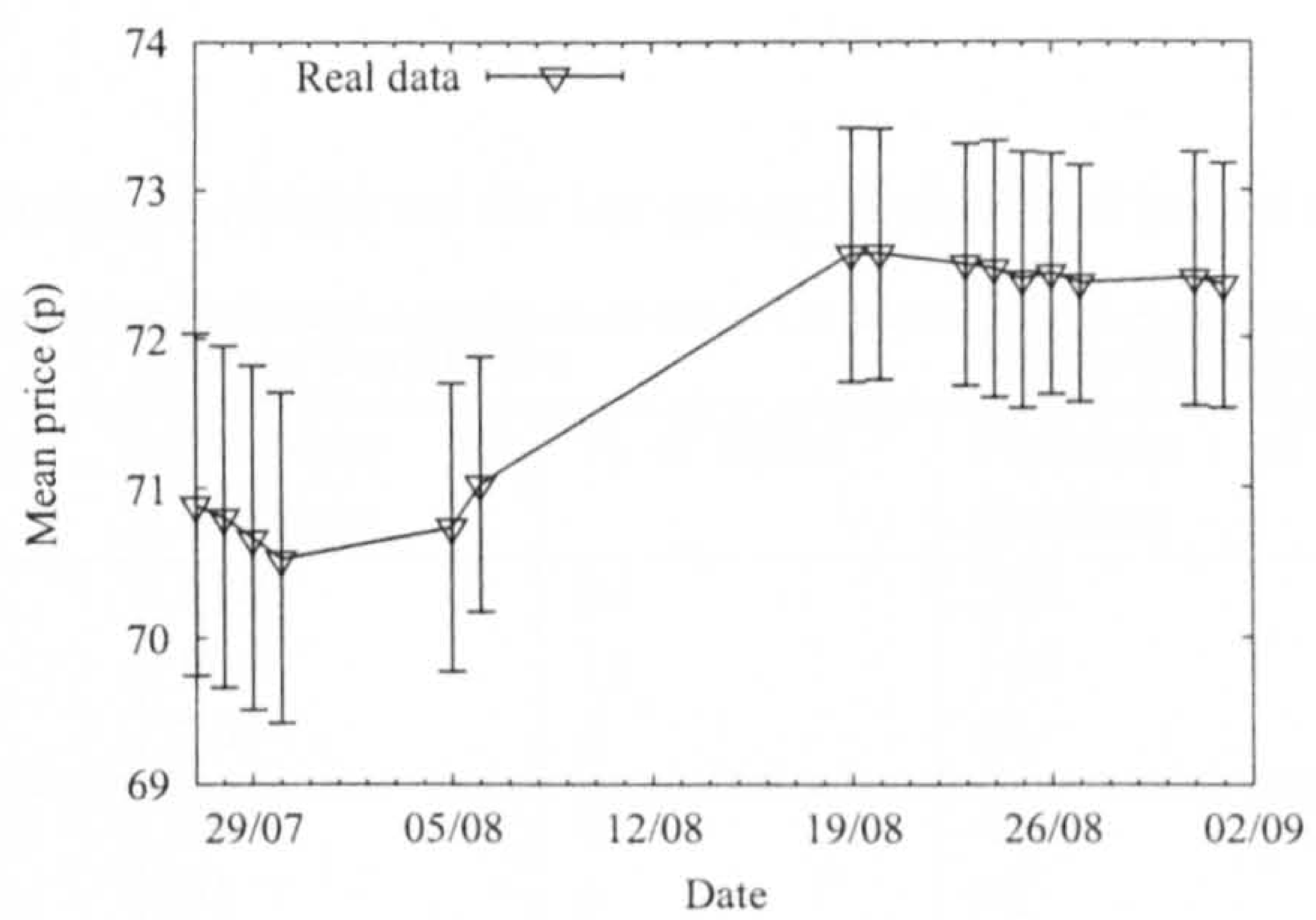
- Although comprehensive in spatial coverage, temporally there are large gaps within the data (see Figure 4.4). This makes identification of patterns and rules quite difficult. Even on days when measurements are recorded there may not be information for every station. However, it will be shown that despite this, the general patterns are still apparent from the data available.

- Within the middle of the data set (6th August - 19th August), the prices rise sharply in a period where there are no recordings. These rises are not generally characteristic of petrol prices; this renders this section of the data unsuitable for studying typical near equilibrium conditions. However, when the prices stabilise again, the data remains steady and can be used. To this end, the data set will be split into two sections. Initial analysis with the model will use the first half of the data (27th July - 6th August) and the second half (23rd August - 1 September) will be used to further test and validate final model versions.

- There are a few significant errors within the prices recorded between each day, for example within 24 hours, the price of unleaded can rise 25p at the same station! To counter this, two approaches will be taken. Where there is sufficient data, an average of the previous three days' prices will be used. On days where this is not possible, a filter will be built into the model that will remove these anomalies. Deciding at which price to filter out the data will be difficult as price differences of 4p over one day (based on initial analysis) may not be unreasonable. There is a danger that valuable data maybe removed. This will be discussed further in § 5.3.4.



(a) West Yorkshire



(b) Yorkshire Region

Figure 4.4: Mean price and standard deviation (indicated by the vertical bars) of real data over time for (a) West Yorkshire and (b) Yorkshire Region. The gaps in the data series are explained in §4.4.2.

4.5 Data Classification

Dividing the data into different classifications will contribute towards understanding patterns and trends within the data as well as further model development. The data was segregated in two ways based on geographical location and the "type" of petrol station. The criteria used to create these divisions are outlined in Table 4.1. Table 4.2 gives the breakdown of stations in each cateogory for West Yorkshire and the Yorkshire region.

Geographically, petrol stations were divided according to location, i.e. motorways or within urban or rural areas as shown in Figure 4.5. (Higher resolution maps of West Yorkshire showing both classifications can be found in Appendix A.) If a station was located in a rural or urban area and was on a motorway, it was classified as a motorway station. These classifications represented the main geographic characteristics of the area. Figure 4.5 shows that North Yorkshire has a much lower density of petrol stations than West or South Yorkshire. The stations are predominately

Classification	Validation
Urban and Rural	Derived from the 1991 Urban settlement boundaries produced by the government[a]. All areas that fell within the urban extent were classified urban. Points located outside of this area were classified rural.
Motorway	Petrol stations located within 0.5km of a motorway and those located on the A1. This distance should hopefully include services that are located slightly away from the motorway.
Multinational	Selection by a list [b] of multinational companies.
Supermarket	Selection by a list [c] of supermarkets.
Minor	All the remaining petrol stations.

[a]This data came from the Office of National Statistics (www.ons.gov.uk)
[b]http://en.wikipedia.org/
[c]http://en.wikipedia.org/

Table 4.1: Explanation of classifications for the geographical and petrol station "type" categories.

	West Yorkshire		Yorkshire Region	
Classification	**Number of stations**	**% of Total**	**Number of stations**	**% of Total**
Urban	434	84	710	76
Rural	63	12	189	20
Motorway	20	4	36	4
Multinational	265	51	483	52
Supermarket	31	6	62	7
Minors	221	43	309	33
Total[a]	517	100	935	100

[a]This is the total number of stations, including obsolete and out of industry.

Table 4.2: % of petrol stations within each classification for West Yorkshire and the Yorkshire region.

rural with a few urban and motorway (A1) stations. West Yorkshire contains the largest number of stations. These are mainly urban with a few rural stations on the edge of the county and several motorway stations. The greatest proportion of stations within South Yorkshire are urban.

The petrol stations were also split according to whether they were multinationals, supermarkets or minor stations (Figure 4.6). Figure 4.6 shows that North Yorkshire has the fewest supermarket stations, but a wide coverage of minor stations, especially in the more remote areas. The multinational stations appear to cluster around urban areas. West and South Yorkshire have the largest number of urban centres. Within or close to these centres are the highest density of supermarket stations. Multinationals are again clustered around the urban centres in both counties. Minors have the greatest spatial distribution, being located in both rural and urban areas.

In total there are 517 petrol stations within West Yorkshire and 935 within the Yorkshire region. However, many of the petrol stations do not have a price recorded or are registered as "obsolete" or "out of industry". These stations total 89 for West Yorkshire and 231 for the Yorkshire region. Additionally, discrepancies within the original data set between postal address and geographical

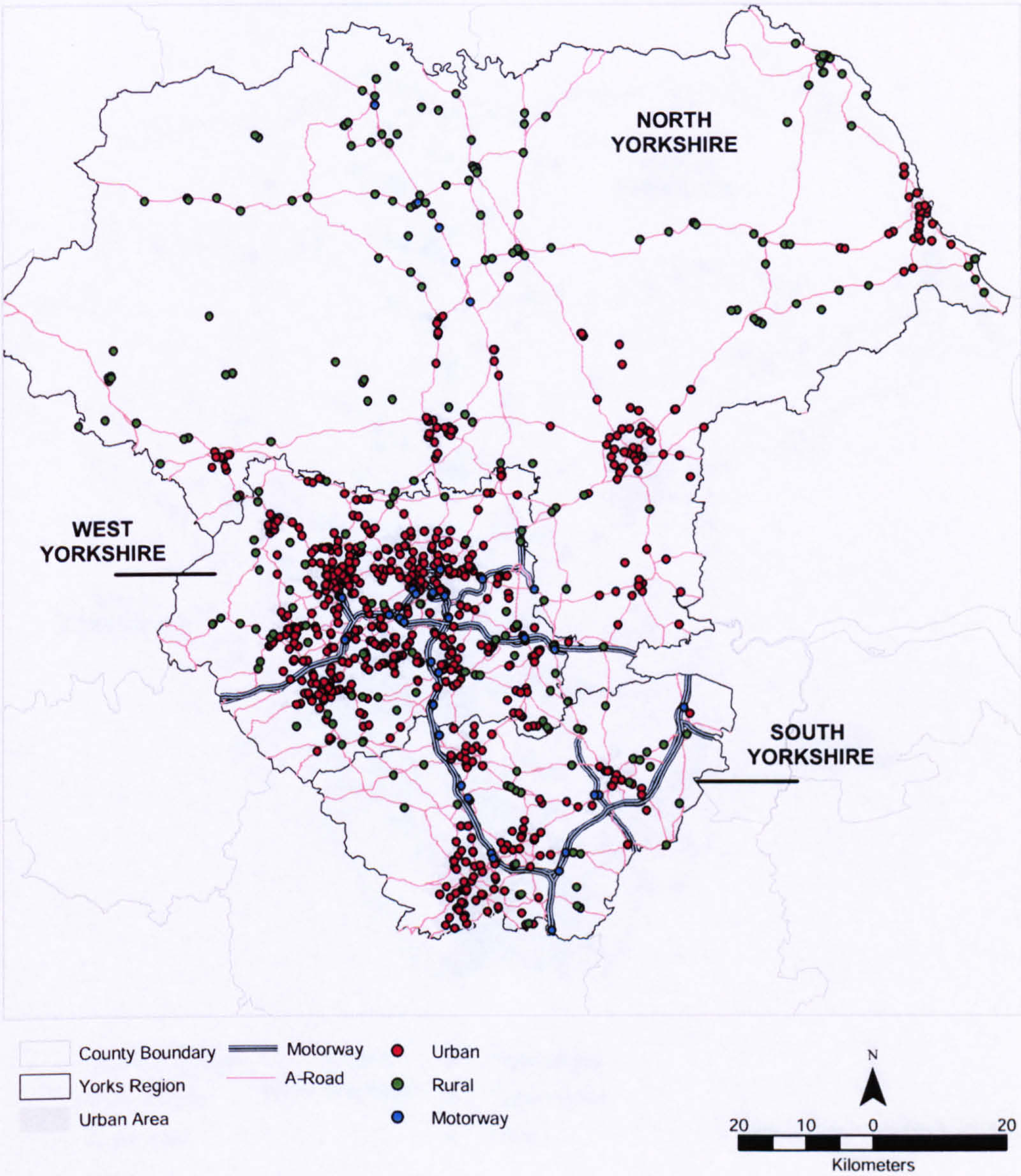


Figure 4.5: Division of petrol stations in the Yorkshire region according to the geographical classifications.

county, meant that a small number of garages (24) were actually located outside of the West Yorkshire boundary (52 outside of the Yorkshire region boundary). Stations such as these are not included in the figures below which show only the stations which are used in the model runs. These numbered (with the removal of the obsolete stations), 428 for West Yorkshire and 704 for the Yorkshire region.

4.5.1 Limitations

Creating different classifications within the data set is, to a degree, subjective. The criteria used to segregate the data will obviously influence the results. For the current classification, the limitations are outlined below:

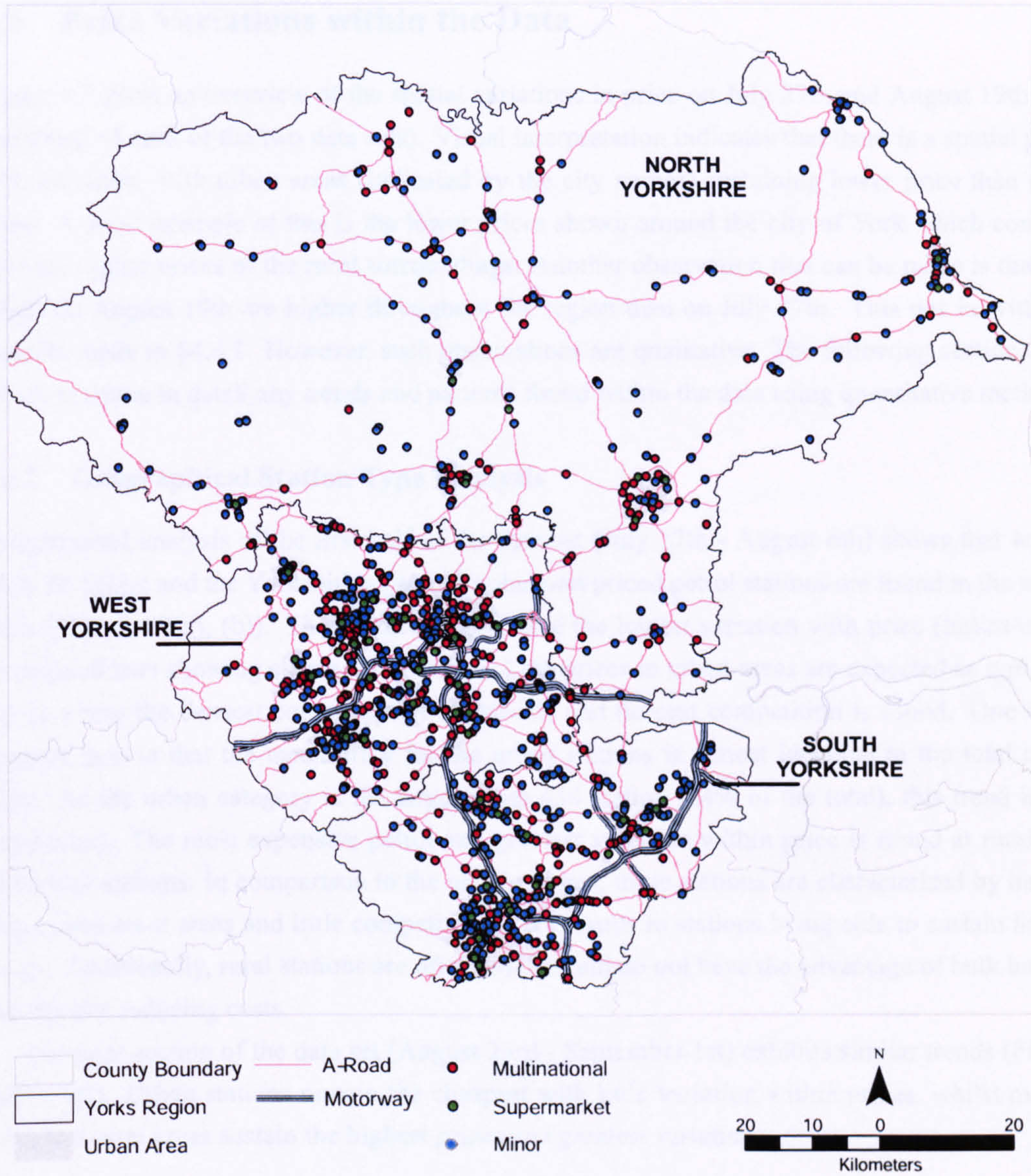


Figure 4.6: Division of petrol stations in the Yorkshire region according to the petrol station "type" categories.

- The classification of the urban - rural divide is based on settlement and population size as utilised by the UK government[3]. This is one of many ways available to classify these areas.
- The lack of classification of national and independent petrol stations is a limitation. There was no information within the database to indicate whether a petrol station was independent or otherwise. Rather than making an arbitrary divide which might not be correct, these smaller national chains and independent stations were classified together in the minors category.

[3]The website is www.statistics.gov.uk/geography/ urban_rural.asp

4.6 Price Variations within the Data

Figure 4.7 gives an overview of the spatial variations in price on July 27th and August 19th (the beginning of each of the two data sets). Visual interpretation indicates that there is a spatial price differentiation, with urban areas (indicated by the city names) sustaining lower price than rural areas. A good example of this is the lower prices shown around the city of York which contrast with the higher prices of the rural surroundings. Another observation that can be made is that the prices on August 19th are higher throughout the region than on July 27th. This ties in with the remarks made in §4.4.1. However, such observations are qualitative. The following sections will aim to examine in detail any trends and patterns found within the data using quantitative methods.

4.6.1 Geographical Station Type Analysis

Geographical analysis of the first half of the data set (July 27th - August 6th) shows that within West Yorkshire and the Yorkshire region, the cheapest priced petrol stations are found in the urban areas (Figure 4.8(a), (b)). These stations also have the lowest variation with price (indicated by the vertical bars showing standard deviation). Low prices in urban areas are expected as typically this is where the densest concentration of stations and fiercest competition is found. One other point of note is that the mean price for the urban stations is almost identical to the total mean price. As the urban category is the largest with 434 station (84% of the total), this trend is not unexpected. The most expensive petrol and greatest variation within price is found at rural and motorway stations. In comparison to the urban centres, these stations are characterised by having larger catchment areas and little competition. This results in stations being able to sustain higher prices. Additionally, rural stations are often smaller and do not have the advantage of bulk buying and thereby reducing costs.

The later section of the data set (August 23rd - September 1st) exhibits similar trends (Figure 4.8(a), (b)). Urban stations remain the cheapest with little variation within prices, whilst motorways and rural areas sustain the highest prices and greatest variation.

Price Distribution

The distribution of petrol prices within each category cannot be solely learnt from an analysis of the mean and standard deviation. Calculating the mean of the prices will result in any extreme values being masked. Quantifying the price distribution allows firmer conclusions to be drawn about internal pricing structures. For example, are rural stations much more expensive than urban stations? Frequency graphs were used to plot the price distribution of rural, urban and motorway stations for July 27th and August 19th in the West and Yorkshire study areas (Figure 4.9).

The results for July 27th (Figure 4.9(a) and (c)) shows a clear bimodal distribution in prices with the majority of stations being either 69p or 71p. Urban stations dominate the price structure, but as noted in §4.6.1, this category contains the most stations. This category also has the greatest number of stations with the lowest prices (67p and 68p). Conversely, a higher proportion of rural and motorway stations have prices of 72p, 73p and 74p.

By August 19th (Figure 4.9(b) and (d)), all the prices have increased (the majority of stations

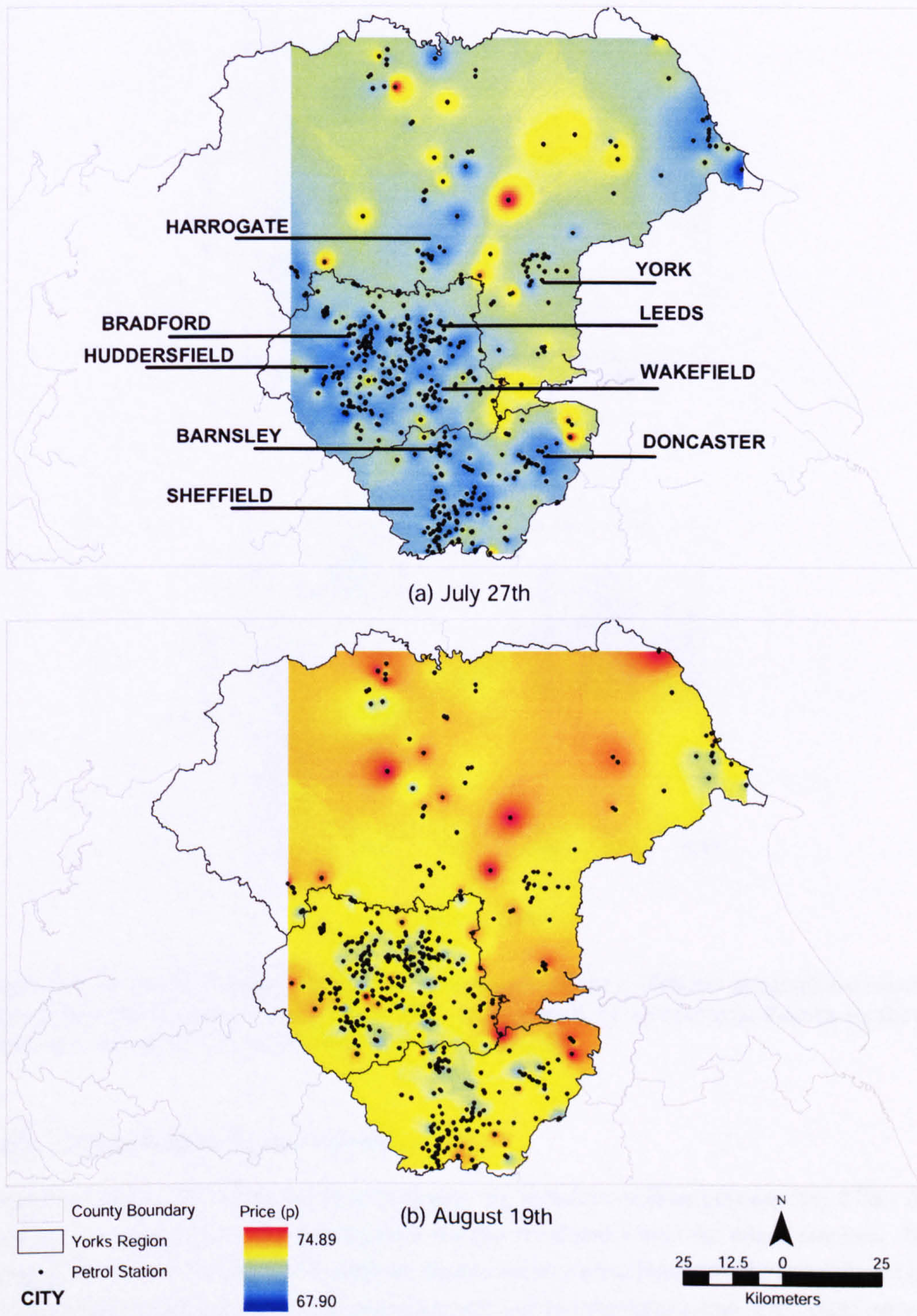
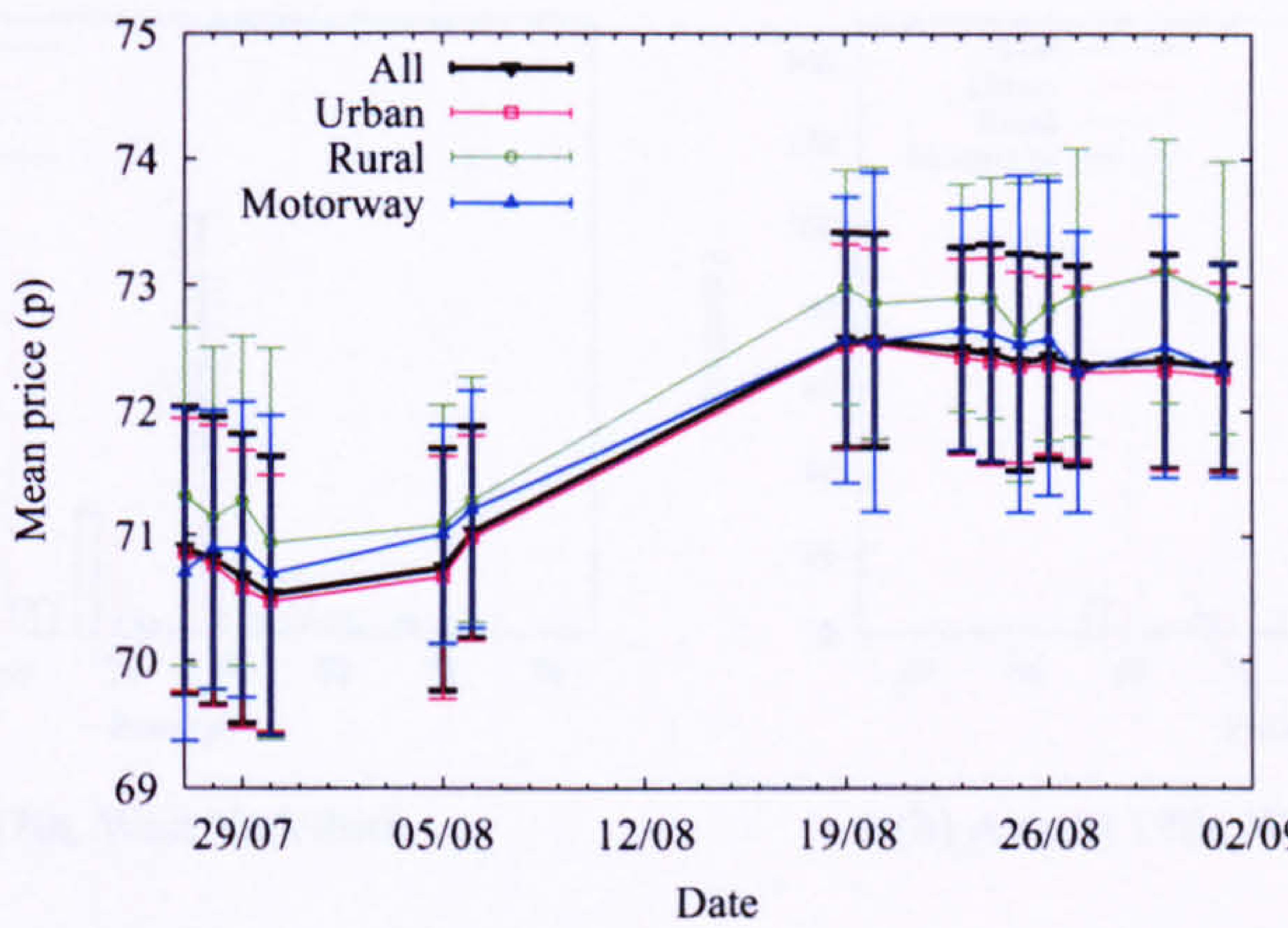
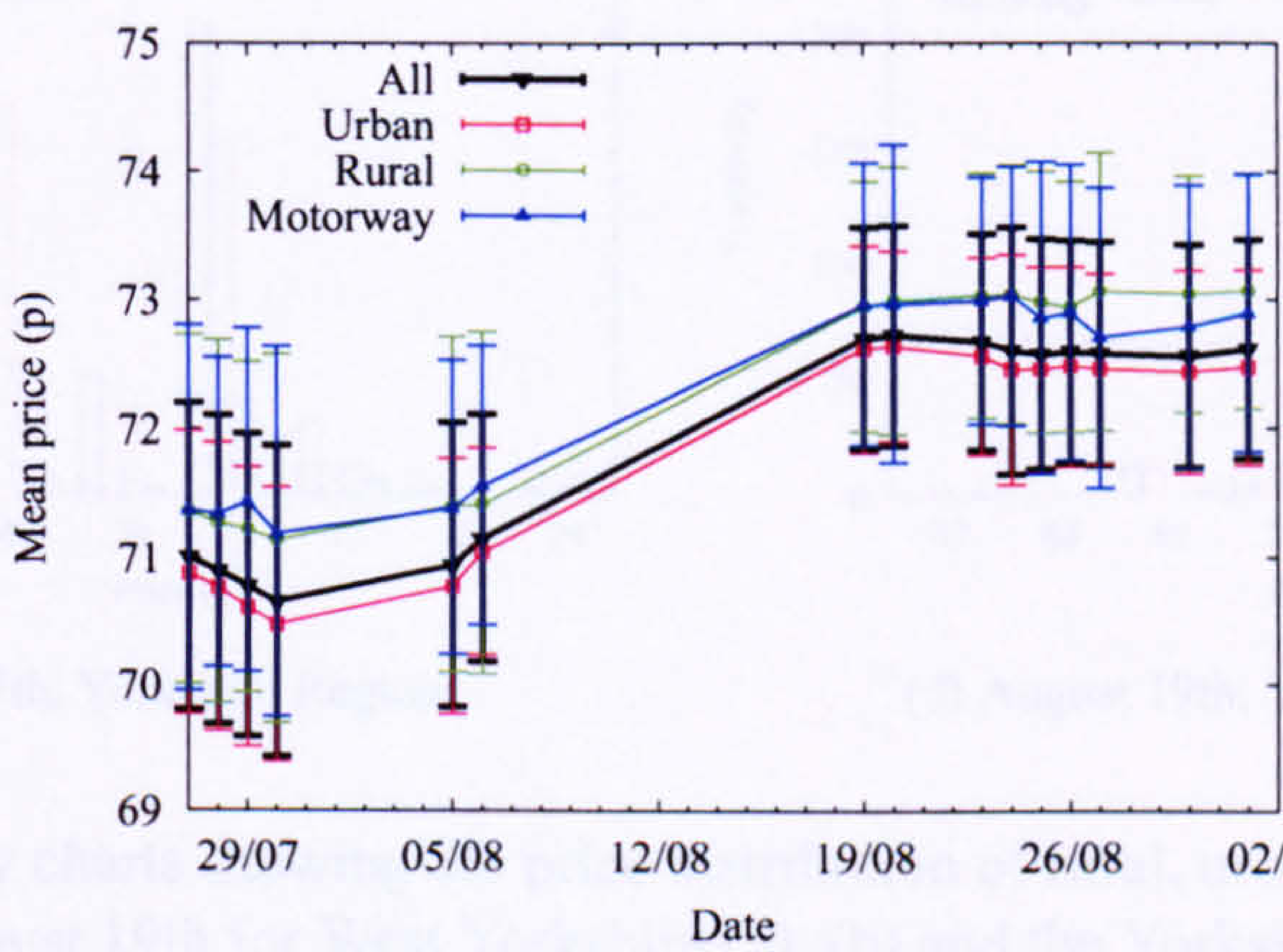


Figure 4.7: Interpolated price surface for the Yorkshire region on (a) 27th July and (b) 19th August.

are now either 71p or 72p and a greater number have increased their price to 73p or 74p). The distribution has changed to unimodal, possibly a reaction to the increase in petrol prices. It is possible that the prices are not stable at this point. Figure 4.4 shows that is likely to be the case, as the petrol prices continue to rise after August 19th.



(a) West Yorkshire



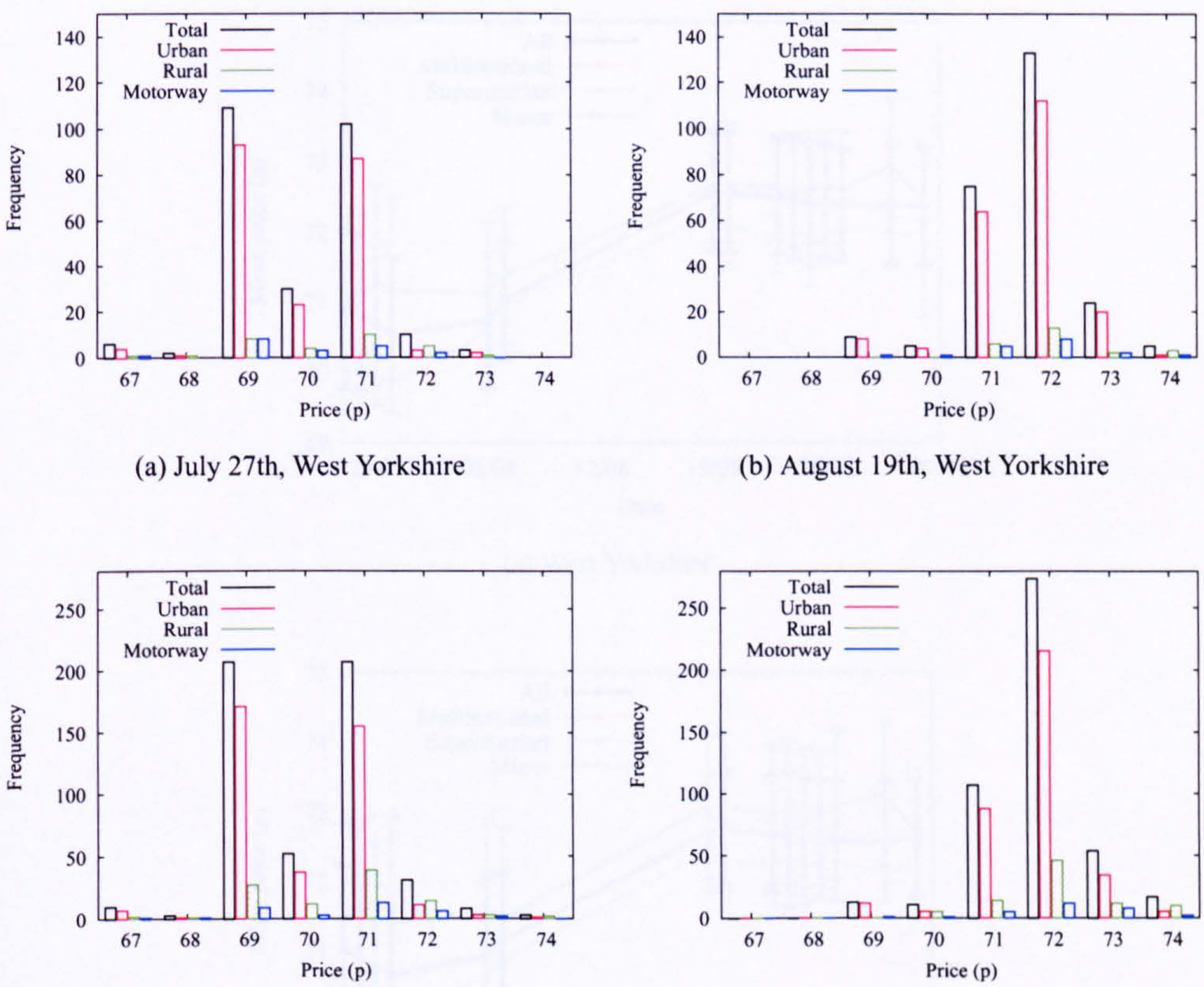
(b) Yorkshire region

Figure 4.8: Graph showing the trends of the mean price within the different geographical classifications throughout data set. Standard deviation is represented by vertical bars. Results are shown from West Yorkshire (a) and the whole Yorkshire region (b).

4.6.2 Petrol Station Type Analysis

Figure 4.10 shows that within the West Yorkshire and Yorkshire regions between July 27th - August 6th, the most expensively priced petrol stations are found within the minor category. This category also had the largest price variation. Supermarkets were consistently the cheapest and had the smallest standard deviation (with an average value of 1p over the duration of the study period), showing that they are competitively priced. They are also the most affected by the rise in price (i.e. their prices increase the most, but they are still the cheapest). The multinational category mirrored the total price. As found with the urban category in §4.6.1, this category is the largest with 265 stations (51% of the total).

The trends within the later section of the data set (August 23rd - September 1st) are the same for the two study regions (Figure 4.10(a), (b)) and mirrored the patterns found in the early data set



(a) July 27th, West Yorkshire

(b) August 19th, West Yorkshire

(c) July 27th, Yorkshire Region

(d) August 19th, Yorkshire Region

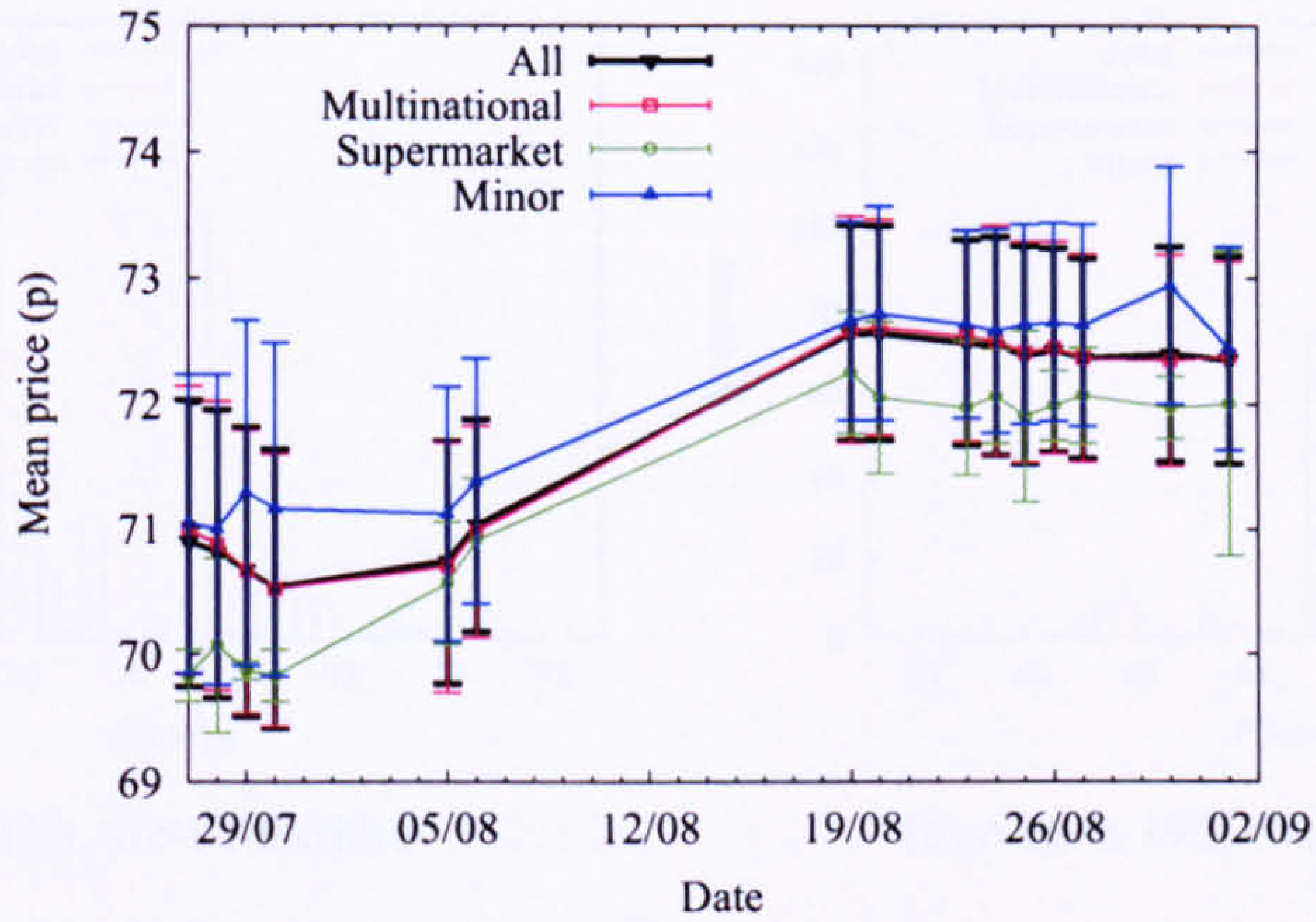
Figure 4.9: Frequency charts showing the price distribution of rural, urban and motorway stations for July 27th and August 19th for West Yorkshire (a), (b) and the Yorkshire region (c), (d).

(July 27th - August 6th). The most expensive stations were found within the minor category, the cheapest in the supermarket category. The minor category still had the greatest variation in price with supermarkets remaining competitive (indicated by a small SD).

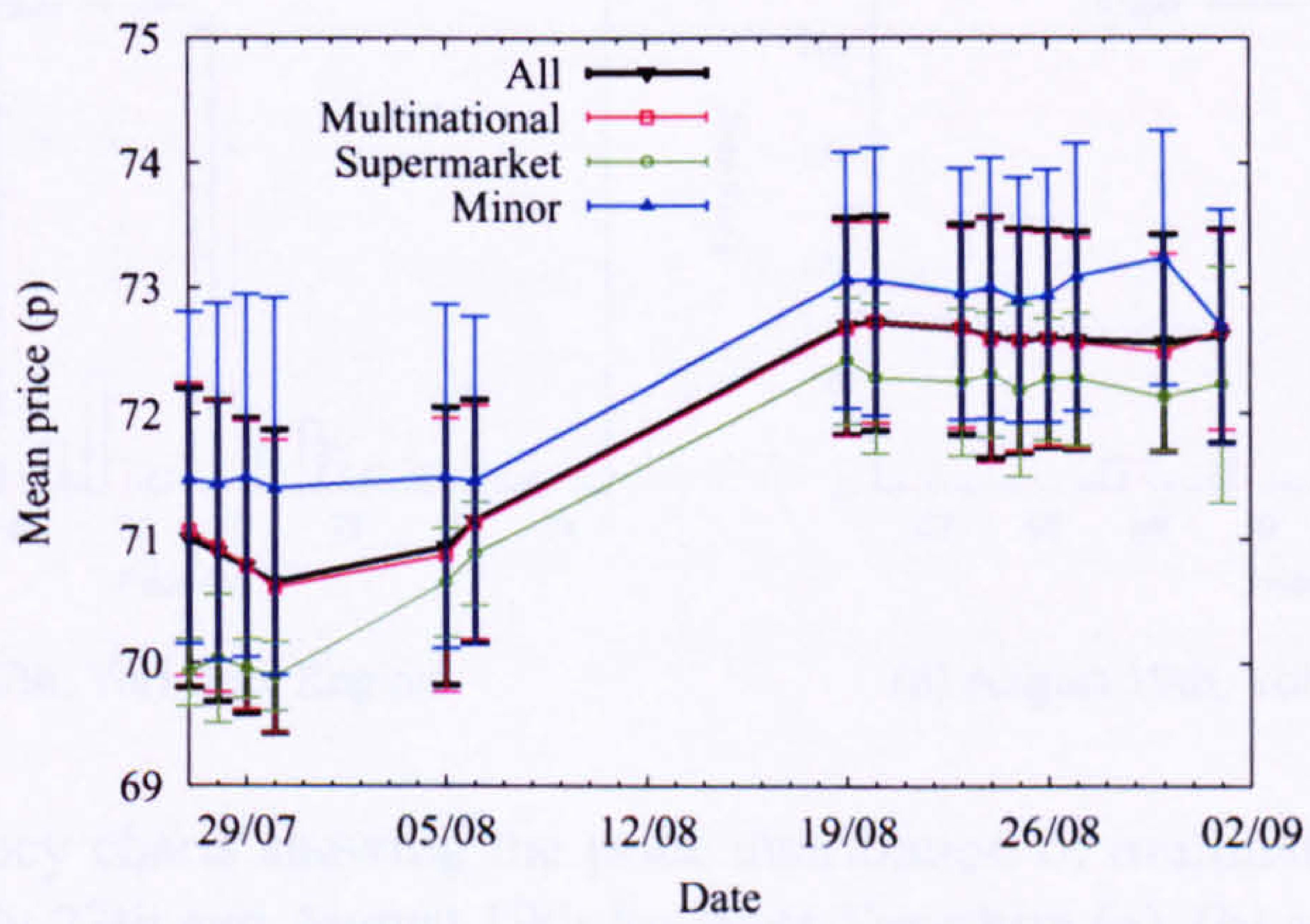
Price Distribution

The results from July 27th (Figure 4.11(a) and (c)) show a marked similarity to those presented in §4.6.1. There is a clear bimodal distribution with the majority of prices 69p or 71p. The largest category, multinational stations, dominate the pricing structure. Overall, the supermarket class is the cheapest with the majority of stations priced at 69p or less. The minor category is the most expensive group with a large proportion of the other stations 71p or more. These results correlate with Figure 4.10 which shows the supermarket and minor categories as the cheapest and most expensive respectively.

The distribution on August 19th has become unimodal (Figure 4.11(b) and (d)). The same pattern was found in the analysis of the geographical classifications. Again, the reason for this change in price distribution may be related to the increase in prices.



(a) West Yorkshire



(b) Yorkshire region

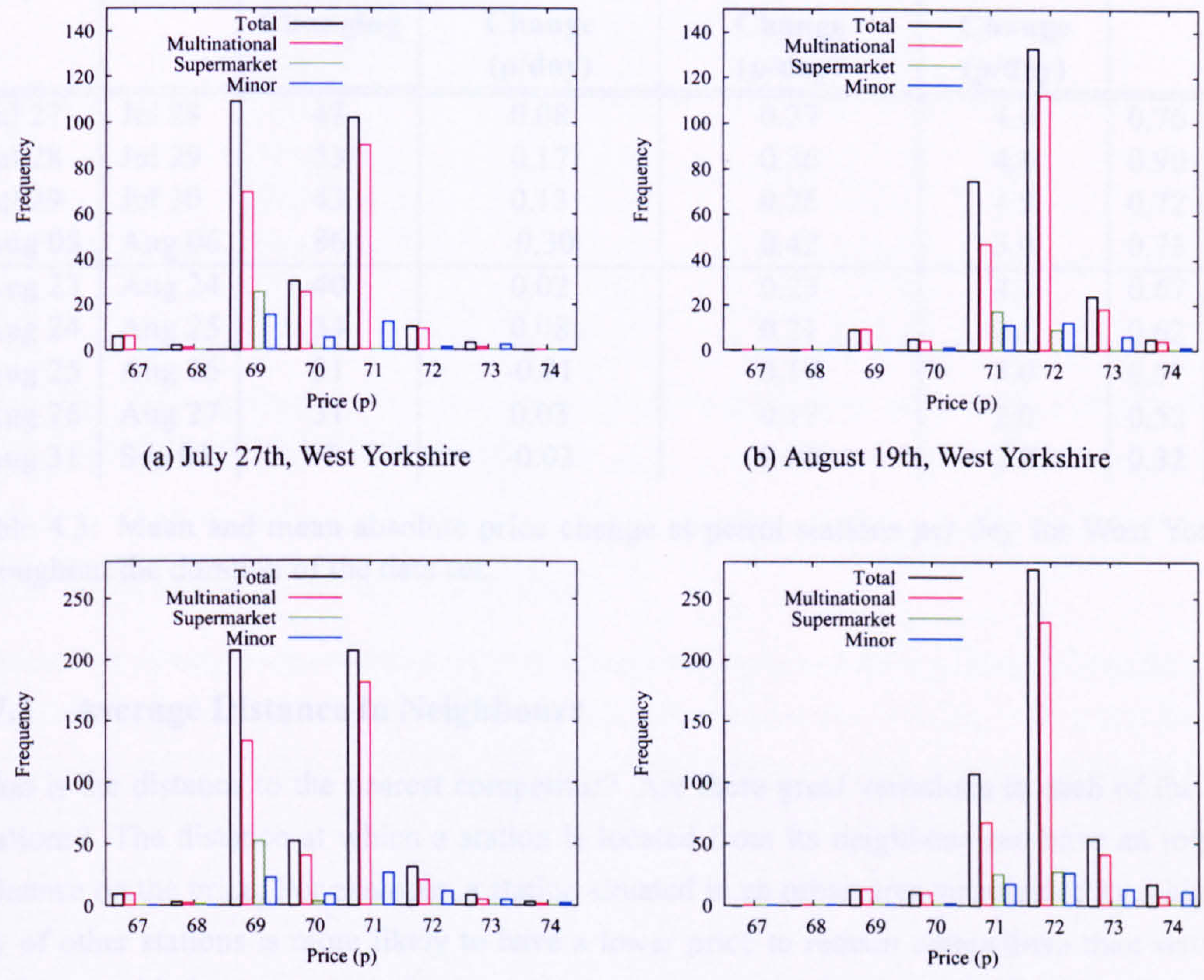
Figure 4.10: Graph showing the trends of the mean price within the different petrol station classifications throughout data set (July 27th - September 1st). Standard deviation is represented by vertical bars. Results are shown from West Yorkshire (a) and the Yorkshire region (b).

4.7 Further Analysis

Statistics collated by the Competition Commission (1990) gave valuable information about factors that influence the size of the price differential between stations. These statistics included information on, for example, location to nearest competitor and price differences between competitors (see §2.3.1). The purpose of this section is to perform analysis on the real data to find out the distance to the nearest competitor, average price changes and average number of neighbours. This information will be invaluable when deciding on parameters and rules within model development.

4.7.1 Average Price Change

The average price change per day (how much all the stations change on average) and the absolute price change per day (the magnitude of the changes at the stations that are changing) were both



(a) July 27th, West Yorkshire (b) August 19th, West Yorkshire

(c) July 27th, Yorkshire Region (d) August 19th, Yorkshire Region

Figure 4.11: Frequency charts showing the price distribution of multinational, supermarket and minor stations for July 27th and August 19th for West Yorkshire (a), (b) and the Yorkshire region (c), (d).

calculated. Using a combination of these values can be useful. For example, the average price change could be zero (thereby masking large changes), whereas the absolute value will show exactly how much the price is changing on average per day.

The values for the Yorkshire region were very similar to those seen in West Yorkshire. The discussion will therefore concentrate on West Yorkshire. The values for the Yorkshire region are tabulated in Appendix B.

Table 4.3 shows that the mean price change for West Yorkshire ranges between -0.03p to 0.30p. On a given day only about $1/5$ of the stations actually change price. The mean absolute price is higher, ranging between 0.06p - 0.42p. This suggests that it is not simply a case of all prices rising or falling on a given day. The changes that do occur appear to be extreme, for example, the maximum price change that occurs between July 27th - July 28th is 4.6p. Some of these maximum changes are sufficiently large to be a little suspicious. This information is useful, but provides no insight into any spatial patterns in variation.

Day 1	Day 2	Number Changing	Mean Change Change (p/day)	Mean Absolute Change (p/day)	Max Price Change (p/day)	SD	RMS
Jul 27	Jul 28	43	0.08	0.27	4.6	0.76	0.77
Jul 28	Jul 29	53	0.17	0.36	4.4	0.90	0.92
Jul 29	Jul 30	43	0.13	0.25	4.5	0.72	0.73
Aug 05	Aug 06	86	-0.30	0.42	3.0	0.73	0.78
Aug 23	Aug 24	40	0.02	0.23	4.1	0.67	0.67
Aug 24	Aug 25	34	0.08	0.21	3.1	0.62	0.62
Aug 25	Aug 26	31	-0.01	0.17	3.0	0.57	0.57
Aug 26	Aug 27	31	0.03	0.17	2.0	0.52	0.52
Aug 31	Sep 01	7	-0.03	0.06	2.0	0.32	0.32

Table 4.3: Mean and mean absolute price change at petrol stations per day for West Yorkshire throughout the duration of the data set.

4.7.2 Average Distance to Neighbours

What is the distance to the nearest competitor? Are there great variations in each of the classifications? The distance at which a station is located from its neighbour can have an important influence on the price. For example, a station situated in an urban area surrounded by a high density of other stations is more likely to have a lower price to remain competitive than stations in rural areas with fewer competitors. Analysis was performed using the classifications on July 27th for West Yorkshire and the Yorkshire region. The patterns for both regions were very similar with the main difference being the distance between stations. This is a reflection of the inclusion of the predominately rural and geographically larger county of North Yorkshire. Due to the similarity in results, only those pertaining to the main study area of West Yorkshire will be discussed. The values for the Yorkshire region are tabulated for reference.

Table 4.4 shows that the greatest distance between stations for West Yorkshire occurs in rural areas (an average of 1122.41m) with a much smaller distance in the urban category (638.06m). The average distance between a motorway station and its neighbour is 482.41m. This figure is perhaps lower than expected. However, when examining the distribution and density of the stations within the geographical classification (see Figure 4.5) it is apparent that the motorways within West Yorkshire are closely located to urban areas.

The average distance between a multinational station and its neighbour is 666.64m. This is greater than the distance of the urban station to its neighbour. Multinationals are located throughout West Yorkshire (see Figure 4.6) and this large distribution produces the largest mean in the petrol station type category. The supermarkets' average distance is smaller than the multinationals (467.33m); these stations tend to be located around urban areas with a higher density of stations. The minor category, with a similar distribution, yet a higher density of stations, has a similar value to that of the multinationals.

Classification	Average Distance (m)	
	West Yorkshire	Yorkshire Region
Urban	638.06	793.99
Rural	1122.41	1817.09
Motorway	482.41	1079.84
Multinational	666.64	963.94
Supermarket	467.33	574.94
Minor	775.04	1247.96
Total	701.04	1056.62

Table 4.4: Average distance in metres between petrol stations in each of the classifications for West Yorkshire and the Yorkshire region.

4.7.3 Average Number of Neighbours

The number of neighbours within a neighbourhood is a useful indicator of likely degree of competitiveness. For example, Esso have a Price Watch policy that states that they will match the lowest price of any station within 3 miles (5km). In parameterising a model, it would be useful to know how many stations are in a neighbourhood. Table 4.5 shows that with an increasing neighbourhood size, the number of neighbours also increases. For a given size, neighbourhoods in West Yorkshire contain more stations on average than the Yorkshire region as a whole. This reflects the more urban nature of West Yorkshire compared to North Yorkshire. Only for very large neighbourhoods (40km or more) is this no longer true. This is simply because these neighbourhoods stretch well beyond West Yorkshire and thus include additional stations from North and South Yorkshire. The figures in Table 4.5 shows that for Esso to keep their Price Watch promise, they have to check the prices of approximately 18 neighbouring stations every day within West Yorkshire and 14.8 in the Yorkshire region.

Neighbourhood (km)	Mean Number of Neighbours	
	West Yorkshire	Yorkshire Region
1	2.0	1.8
2	4.4	3.8
5	18.0	14.8
10	56.1	43.2
15	110.0	81.8
20	162.0	123.0
25	203.8	163.9
30	233.1	204.4
40	257.1	282.2
50	261.6	360.1

Table 4.5: Average number of neighbours within a given sized neighbourhood for the West and Yorkshire region.

4.7.4 Neighbourhood Statistics

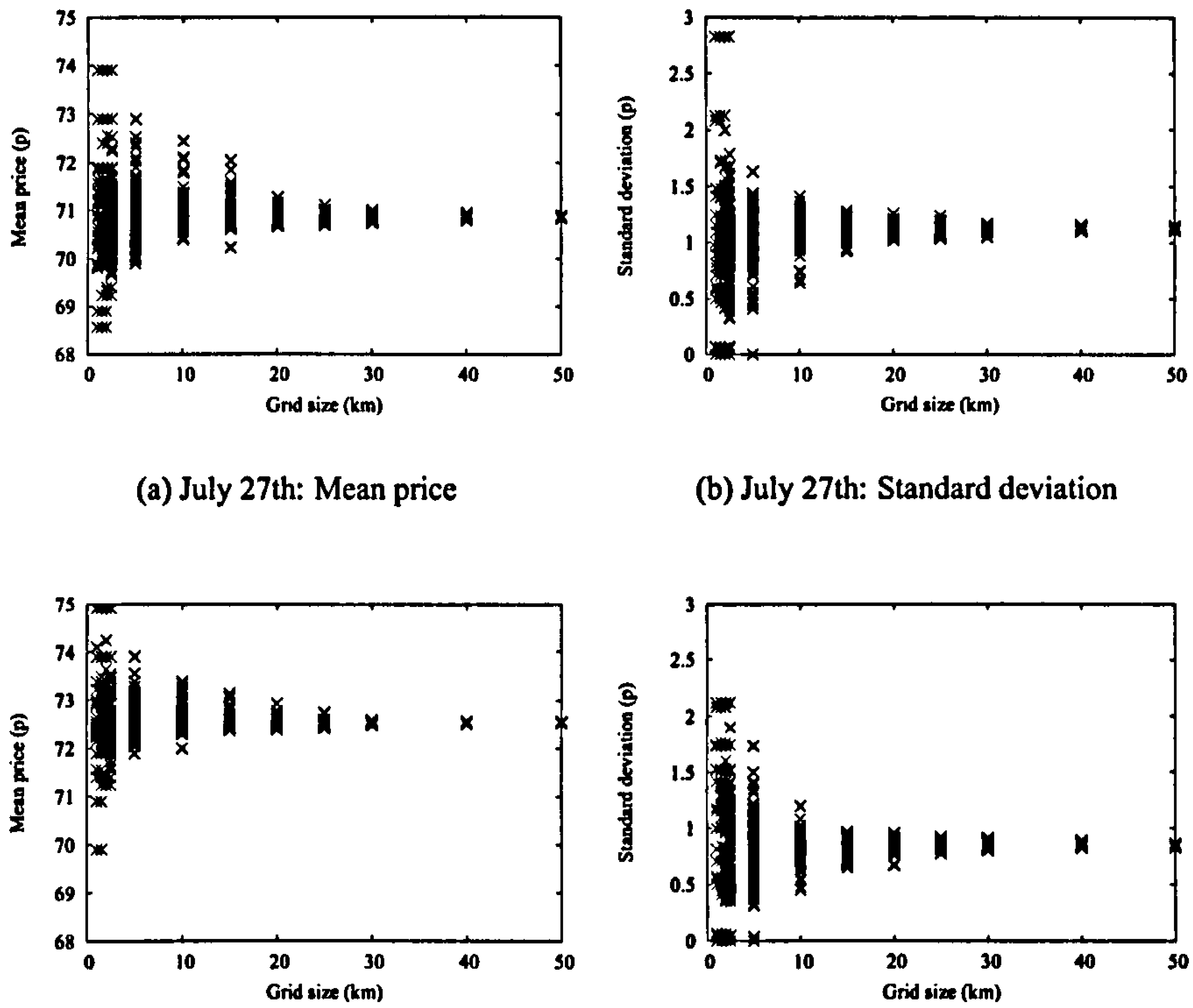
The average distance to the nearest neighbour and number of neighbours in a given area has been calculated. These give invaluable information on the spatial distribution of stations but neither of these techniques indicate whether there are any strong price differentials within a neighbourhood. To understand this, the mean price and range of prices (represented by the standard deviation (SD)) will be calculated for different neighbourhood sizes. The analysis will be performed on West Yorkshire for both July 27th and August 19th.

Around each station, a circular neighbourhood will be taken of a given radius and the mean price and SD calculated in this neighbourhood. The results should indicate the presence of any strong price differentials. This approach is entirely objective. Each station has a neighbourhood within which other stations are assumed to exert the largest influence. The downside with this method is that a proportion of the stations will be duplicated. An alternative approach would be to place grids of varying sizes over the study area and calculate the SD and mean price. The pitfall with adopting this technique is that there is a chance that the stations which exert the largest influence will be omitted.

Figures 4.12(a) and (c) show the mean price in each neighbourhood. Due to the price rise between the early and late data sets, the mean prices on August 19th are much higher than July 27th. Other than this, the patterns found are the same. Within the smaller neighbourhoods (10km and below) the majority of the mean prices are between 70p - 72p. It could be hypothesised that this is caused by the urban stations. There is also a large number with means of 68.5p (possibly supermarkets) and 74p (possibly rural areas). With an increase in neighbourhood size, these extremes are averaged out and once the scale goes beyond 25km, there is little further variation.

The SD's (Figures 4.12(b) and (d)) for July 27th and August 19th show that for smaller neighbourhoods (below 10km), there is a larger range of prices. As found with the mean price, the majority of SD's are concentrated between two values (0.5 - 1.5). Within these neighbourhoods, all the prices would be expected to be within a 1p range. Some of these smaller neighbourhoods have a larger range (2.0 - 3.0), these could be indicative of rural areas or stations located towards the edge of the study area. With an increase in scale, extreme values are averaged out. From 30km - 50km, there is little change in the amount of variation.

This work has been useful in indicating the spatial scale at which variation occurs within the real data. If a scale larger than 25km is used, variations are smoothed out. West Yorkshire is 60km in width, using a 25km neighbourhood (50km in total) therefore takes in almost all of the area. Most of the variability in price are apparent at a scale of 10km and less. If the system was driven by solely distance and price, little variation would be expected within smaller neighbourhoods. Some small neighbourhoods do have a small SD showing that there is little variability in price. However, there are also a few neighbourhoods with high SD and hence large price variations. This shows that there are other factors at work, for example, customer loyalty and different pricing strategies. These local variations in price will be further investigated in the following sections.



(a) July 27th: Mean price

(b) July 27th: Standard deviation

(c) August 19th: Mean price

(d) August 19th: Standard deviation

Figure 4.12: Mean price and standard deviation of petrol station neighbourhoods plotted as a function of neighbourhood size on July 27th and August 19th for West Yorkshire.

4.7.5 Analysis of Esso Stations

The Esso Price Watch policy provides the only published information on a petrol retailer's strategy. In Chapters 5, 7, 8 and 9, the Esso stations will be assigned the Price Watch rules as part of the experimentation. Figure 4.13 shows the mean price of all of the stations (except Esso) and the Esso stations. It can be seen that Esso sustain a lower price (by approximately 1p) than the other stations. This is not surprising as the Price Watch policy means Esso stations are keeping in check with the lowest prices in their neighbourhood. The SD is also smaller than the rest of the stations indicating that the Esso prices show less variation, presumably because they are more competitive on price.

Figure 4.14(a) shows that the Esso stations on July 27th have a unimodal distribution. This peak corresponds to the lower of the peaks of the bimodal distribution of the all the other stations. This provides further evidence of Esso's competitive behaviour as a result of the Price Watch Policy. On August 19th (Figure 4.14(b)), the prices at the Esso stations in common with all the other stations, has risen. There are now two peaks rather than a single peak in the price of Esso stations, with a split between 71p and 72p. However, the Esso stations remain competitive; for example, there are almost as many Esso stations selling petrol at 69p as the rest of the stations and very few Esso stations are selling petrol at more than 73p.

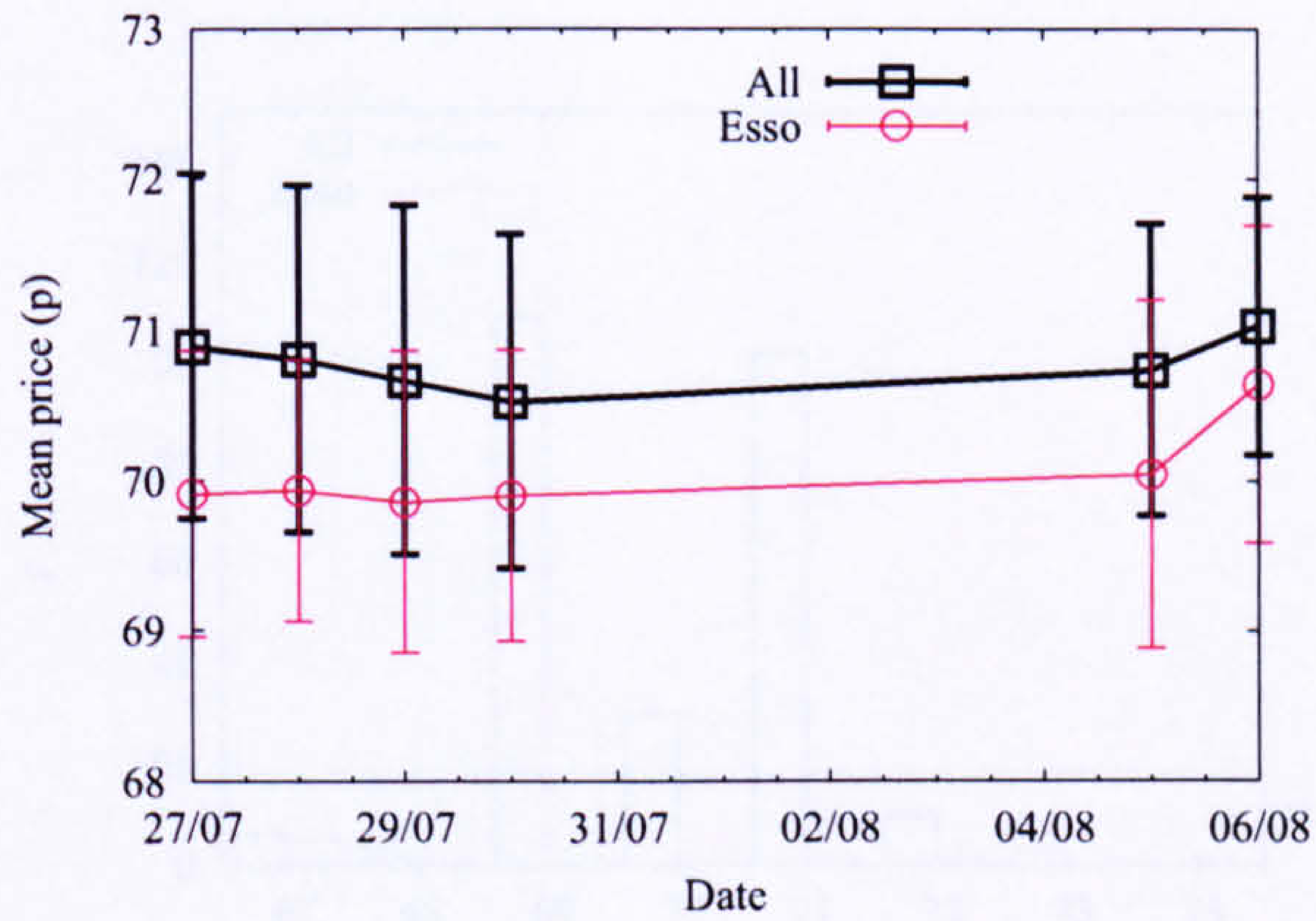


Figure 4.13: Mean price against time over the period July 27th to August 6th for West Yorkshire. The vertical bars indicate the standard deviation in the price. Results are shown for all stations and also just for the Esso stations.

Figure 4.15 presents statistics from the neighbourhood of 5km around the Esso stations (5km equates to the 3 miles cited in the Price Watch policy). The mean neighbourhood price is plotted against the price of the Esso station. The line shows where these two prices are equal. With the exception of two stations, the price at each Esso stations is cheaper than the mean price of its competitors. This provides good evidence that implementation of the Esso Price Watch results in cheaper prices at Esso stations.

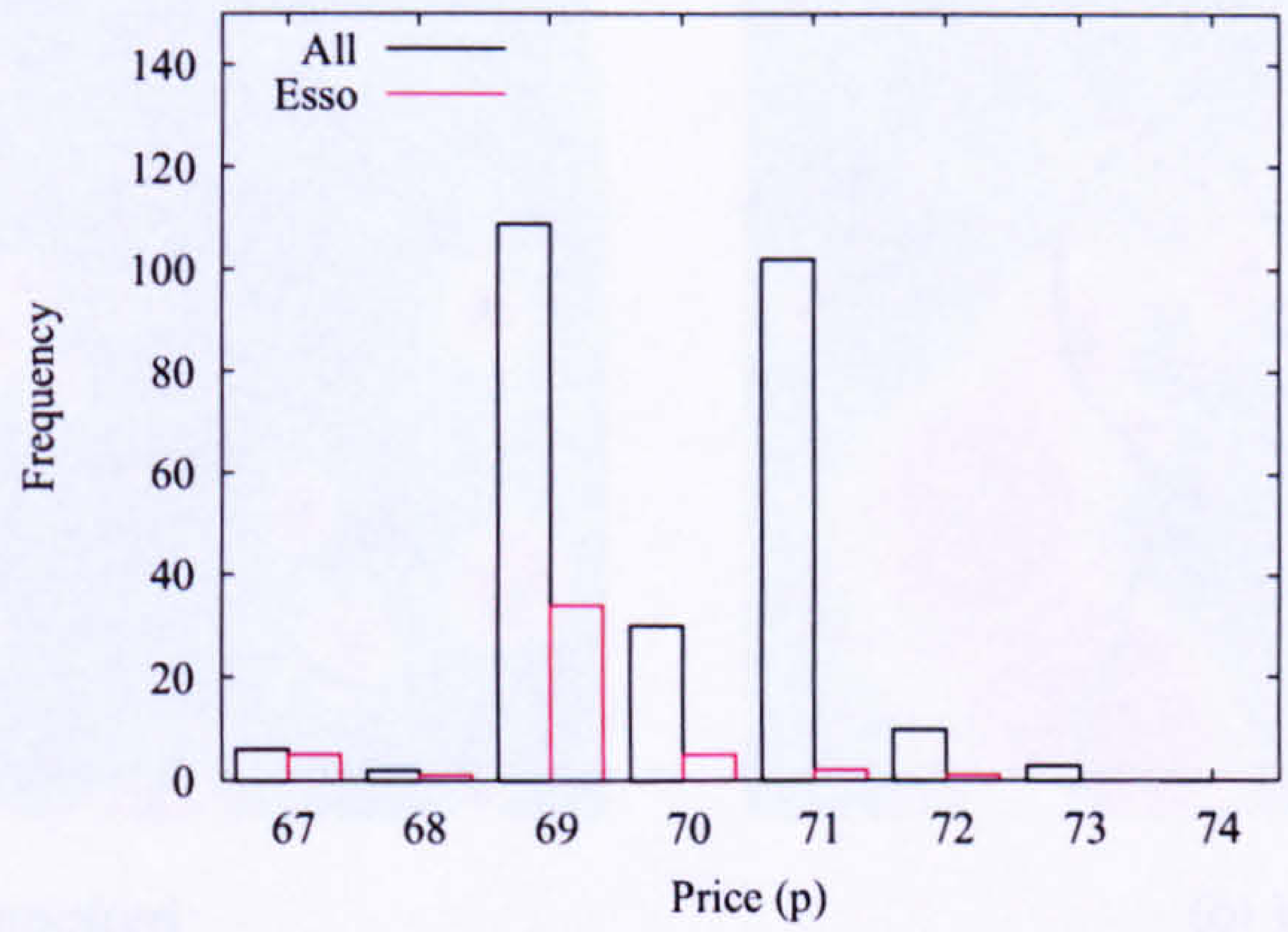
4.7.6 Visual Evidence of Behaviour

Several conclusions have been made about the real data based on the above analysis and review of the literature. For example, stations react to price changes in their local neighbourhoods, supermarket stations are the cheapest, there is a great deal of price variation within urban areas and so on. The aim within this section is to provide snapshots of the real data to support such statements.

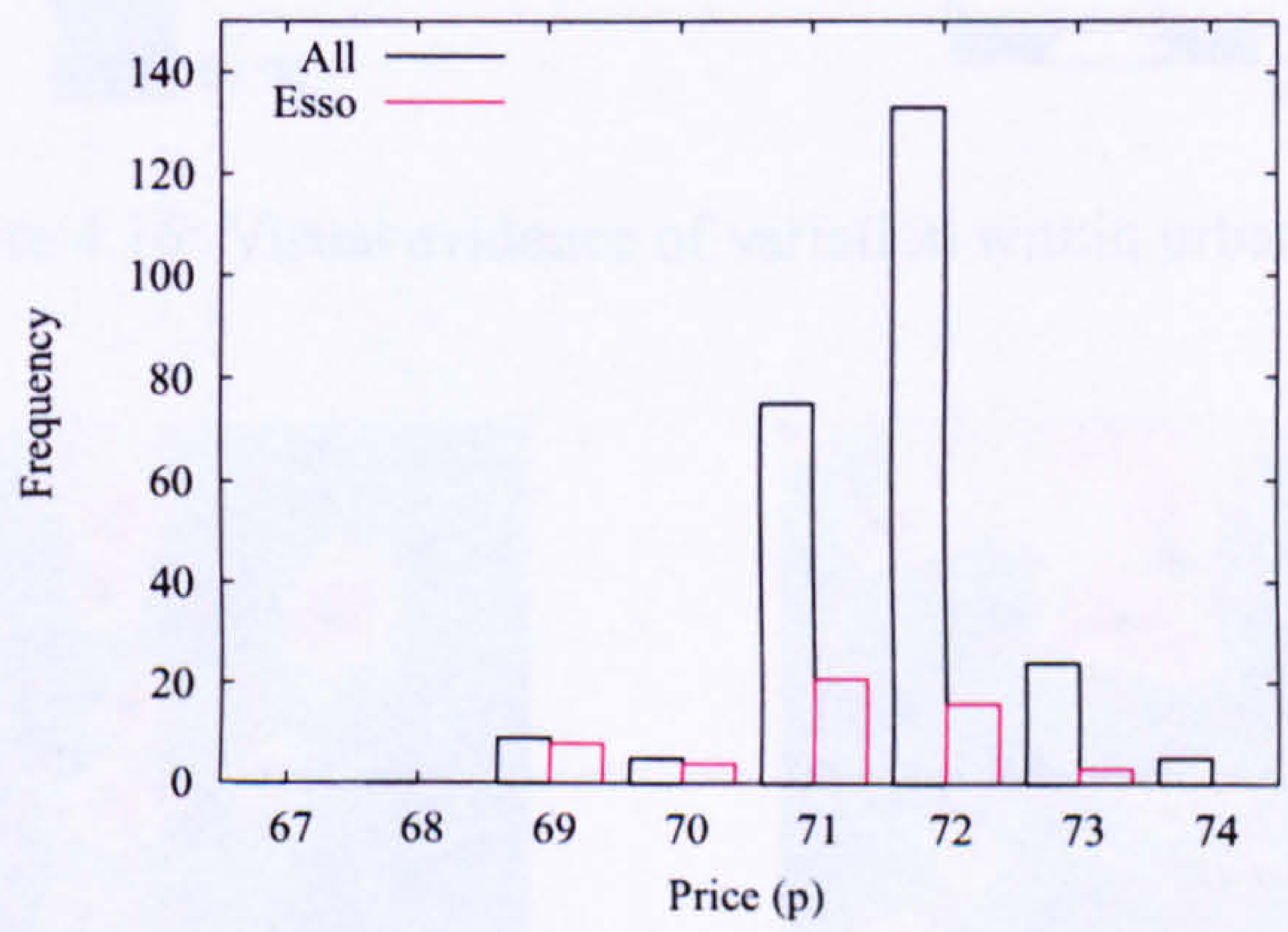
Figure 4.7 showed that there is clear price differential between rural and urban areas. However, it failed to highlight the variation that exists within urban areas. Figure 4.16 shows this variation within two cities, Bradford and Wakefield. Although both cities are cheaper than rural areas, the prices can range between 3-4p. Figure 4.16 also shows that high prices are sustained on the main arterial roads leading into urban areas. This is particularly prominent around Wakefield.

Statistical analysis of the real data identified urban and supermarkets as the cheapest stations (see §4.6.2). Figure 4.17 shows the prices of petrol stations in the city of Leeds over a 3 day period (July 27th - July 29th). The supermarket stations (green) are cheaper than all of the other stations with the exception of Esso (red). These stations, operating the Price Watch policy, are very competitively priced. Over the course of this 3 day period, the supermarkets do not react to changes in price at neighbouring stations. Esso however, do remain competitively priced but react to changes in the prices of nearby stations.

To prove that stations do react to their neighbours, Figure 4.18 shows the reaction of several stations to a price change in their neighbour over 3 days (all the stations fall under the minor



(a) July 27th



(b) August 19th

Figure 4.14: Frequency charts showing the price distribution of Esso and non-Esso stations for July 27th (a) and August 19th (b) for West Yorkshire.

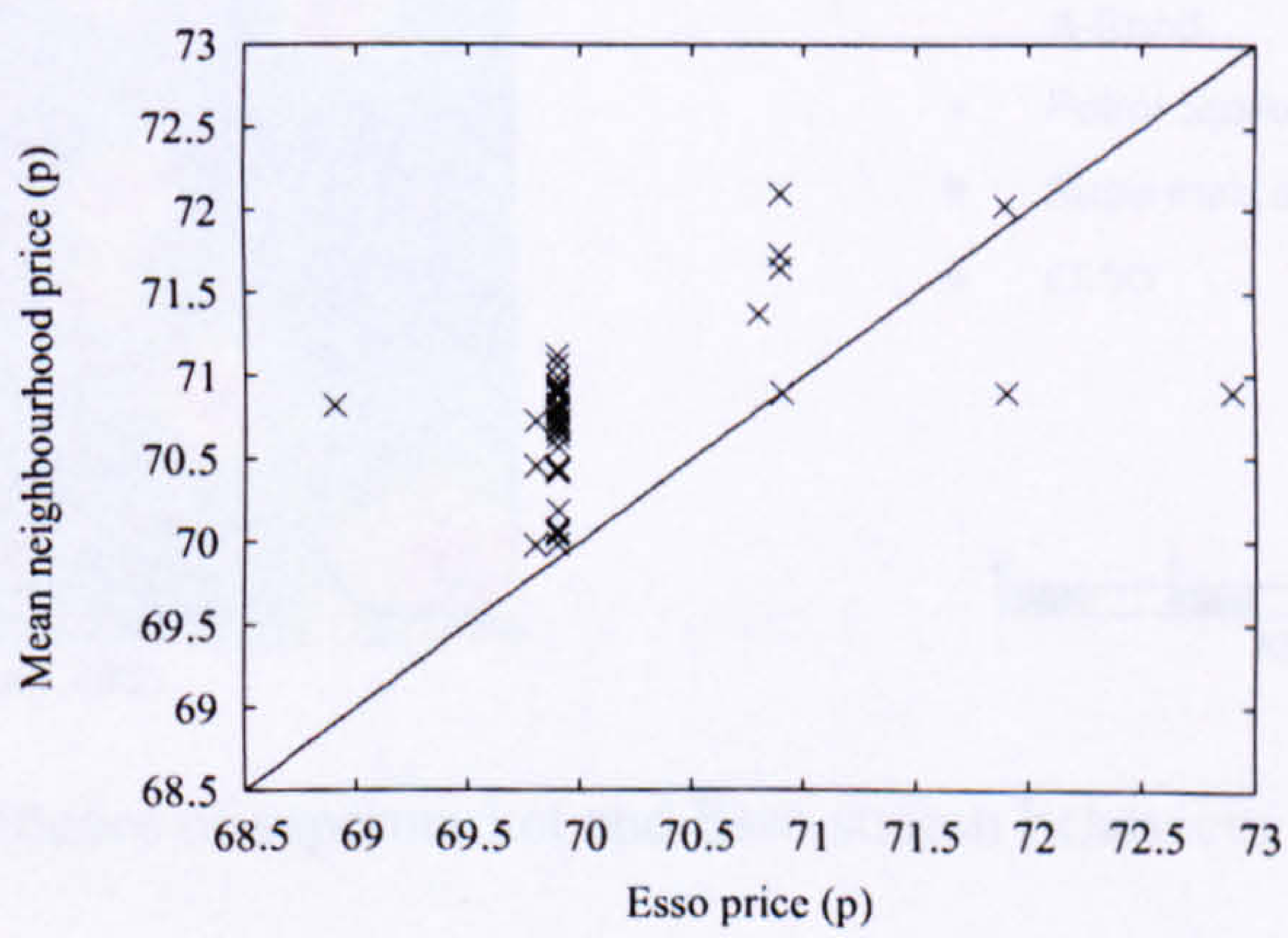


Figure 4.15: Mean price in a 5km (3 mile) neighbourhood of each Esso station plotted against the price at the Esso station for the West Yorkshire data on July 27th.

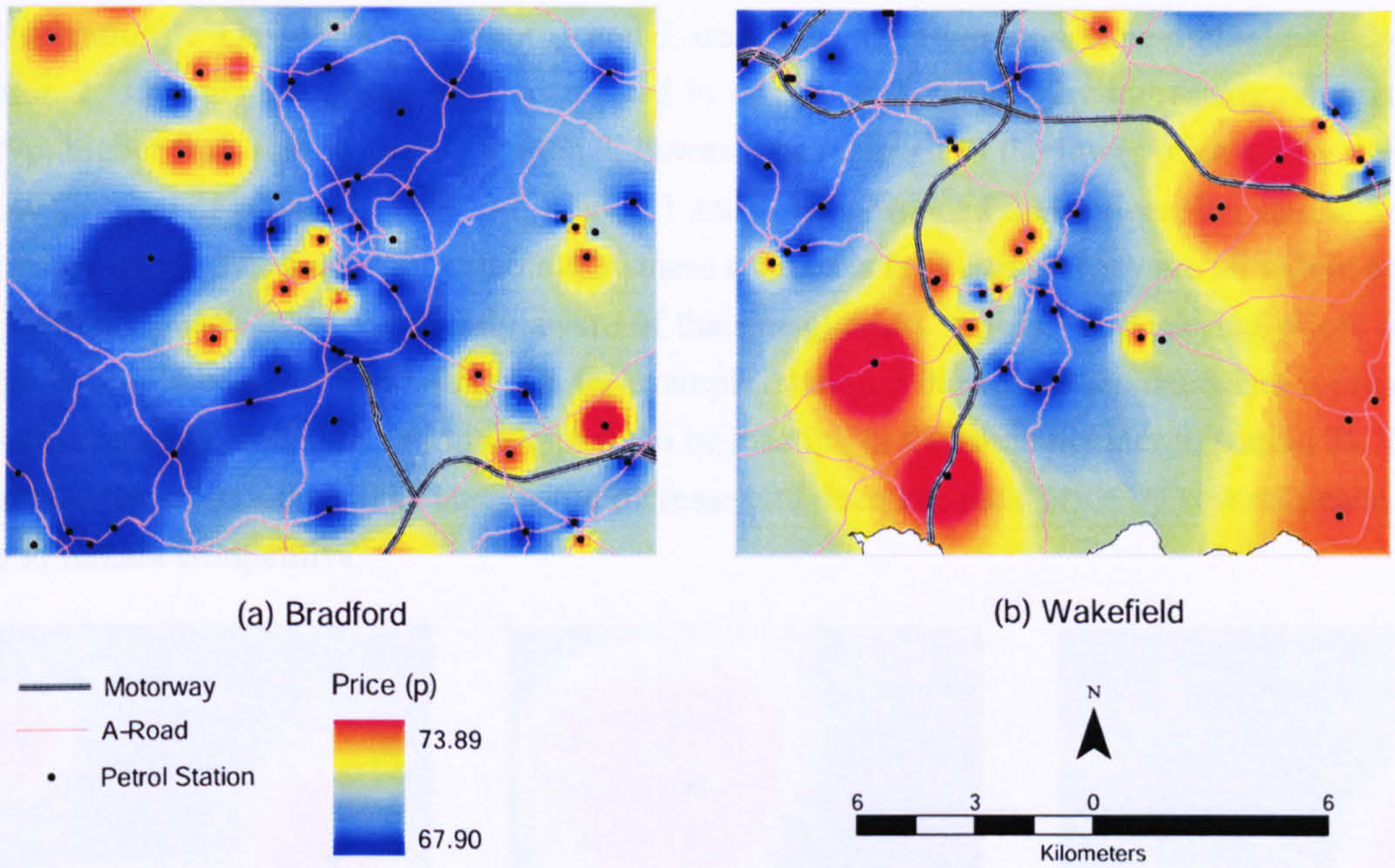


Figure 4.16: Visual evidence of variation within urban areas.

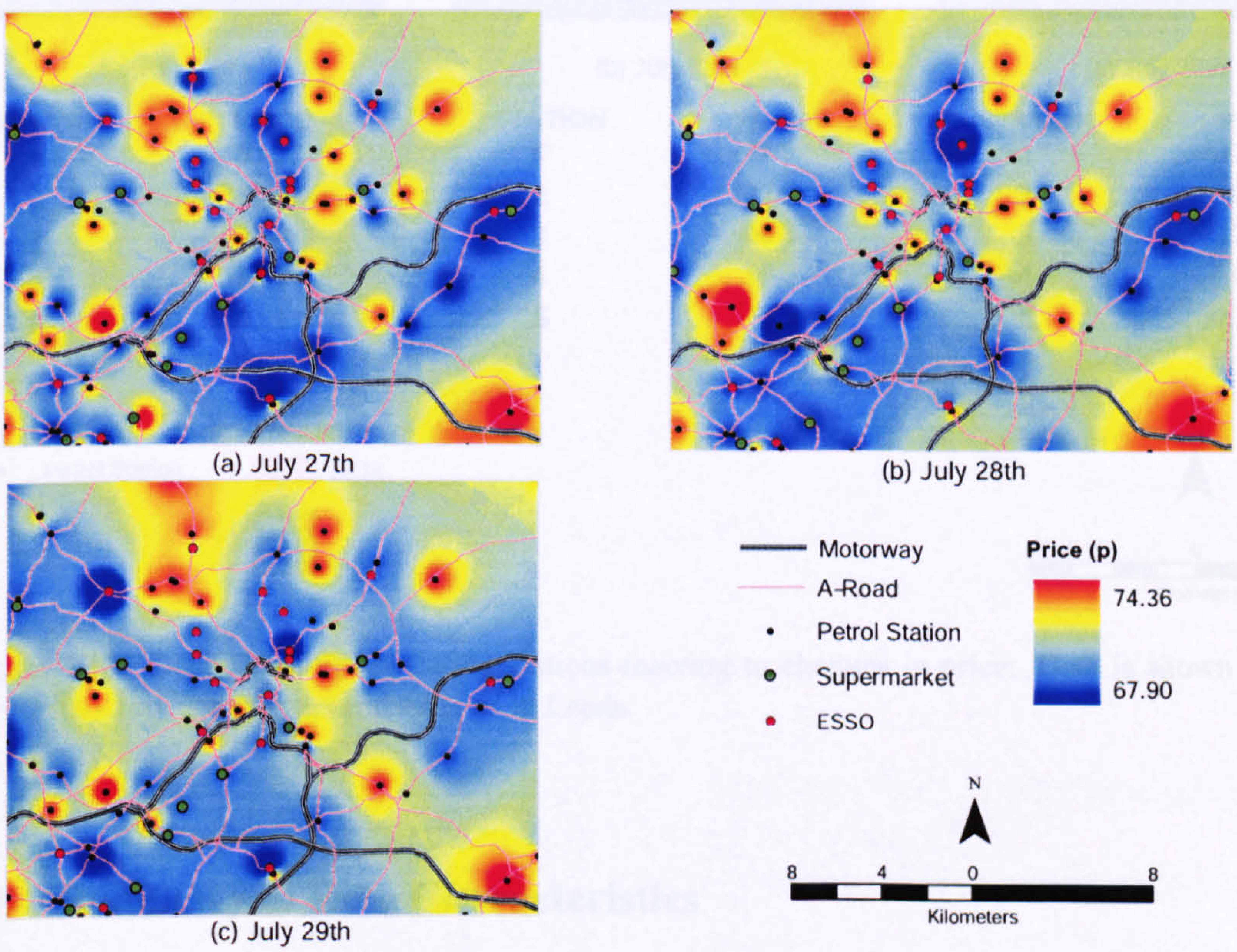
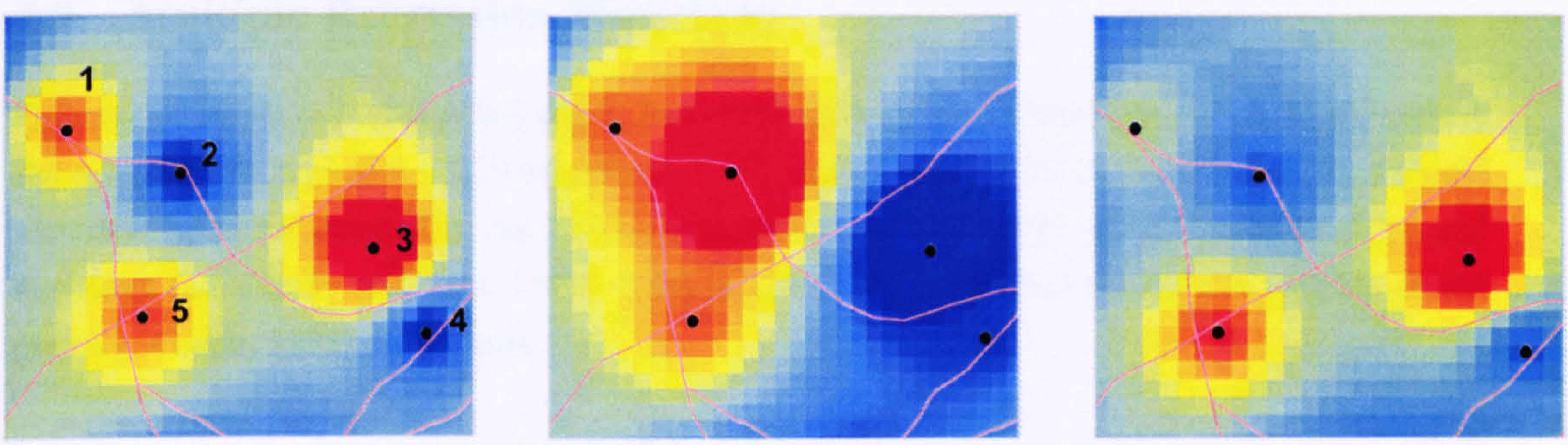


Figure 4.17: Visual evidence of supermarket and Esso station behaviour between July 27th - July 30th in Leeds.

classification)[4]. On July 27th, station 1 and 3 are more expensive than their mutual neighbour, station 2. On July 28th, station 2 has reacted to this price difference by increasing its price to 73.9p, higher than station 1 or 3. Station 3, however has reacted to the lower price set by station 4 and dropped its price. By July 29th, station 1 and 2 have dropped their prices, but station 3 has increased again. The behaviour exhibited by these stations is interesting. They are all within a few kilometres of each other and clearly aware of their neighbours' prices. The varying prices seen each day could be a result of a time lag, for example, station 2 increases then decreases it price in possible reaction to station 3 (which appears to be reacting in the same manner to station 2). This example also shows that some stations do increase and decrease their price by several pence per day to remain competitive.



(a) July 27th

STATION	PRICE
1	71.9
2	69.9
3	72.9
4	69.9
5	71.9

(b) July 28th

STATION	PRICE
1	71.9
2	73.9
3	68.3
4	69.9
5	71.9

(c) July 29th

STATION	PRICE
1	70.9
2	69.9
3	72.7
4	69.9
5	71.9

Figure 4.18: Example of neighbouring stations reacting to changes in price. Data is shown for July 27th - July 30th for several stations in Leeds.

4.8 Summary of Data Characteristics

This study of the characteristics of the petrol price data sets brings up several points which may be of importance in attempting to model the system. One particularly important point is that there is clear evidence of the non-linearity of the system. Changes in petrol price are observed to occur as step changes, often of ± 1 p or of multiples of this. This means that petrol prices do not

[4]Station 1 is Fina; station 2 is SAVE; station 3 is UK; station 4 is Repsol and station 5 is unbranded.

vary continuously which may make them harder to model mathematically. It also means that a very small change in the system may make a larger difference to the price, perhaps the difference between a price increase of 1p or no change in the price.

The findings also show the strong spatial nature of the data. Price variations across the study area are clear, with cheaper prices in urban areas compared to rural areas. Even within a given area there may be significant price variations indicating the importance of other factors. A brief investigation of Esso stations suggests that the Esso Price Watch does have a marked influence on petrol prices on a local scale. Consideration of individual behaviour for different brands and station types may therefore be necessary to accurately represent fuel prices in any model.

4.9 Multiple Regression Modelling

Previous attempts to model petrol prices have been undertaken with the use of empirical models. An example of a typical model was presented and discussed in §2.5.2 (Equation 2.1). This model was developed for examining the relationship between the net retail price, crude oil price and exchange rate (Reilly and Witt, 1998). This is essentially a standard multiple regression model (see e.g. Ebdon, 1985) of the form

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots \tag{4.3}$$

where

- y = dependent variable
- x_1, x_2, x_3 ... = independent variables
- a, b_1, b_2, b_3... = regression coefficients.

The dependent variable will be price and the independent variables will be factors that the price is thought to vary with. The standard method of measuring the goodness of fit of a regression is to calculate the extent to which the regression accounts for the variation in the observed values of the dependent variable (Ebdon, 1985). This is generally calculated using an r^2 test,

$$r^2 = \frac{\Sigma(y_i' - \bar{y})^2}{\Sigma(y_i - \bar{y})^2} \tag{4.4}$$

where r^2 is the coefficient of determination, y_i' is the predicted value of the dependent variable at station i, y_i is the measured value at station i and $\bar{y}$ is the mean value of the y_i.

Performing multiple regression modelling on a data set will produce regression coefficients. The coefficient for each variable is an indication of the effect of the corresponding value, i.e. if this value is large, the variable has a significant impact in determining the price.

The root mean square error (RMSE) could also be used as a method of assessing the performance of various models in comparison with the experimental data. The RMSE is defined as

$$\sqrt{\frac{\Sigma(y_i' - y_i)^2}{n}} \tag{4.5}$$

where n is the total number of stations. One disadvantage of the RMSE is that it can be difficult to compare values from different experiments if the mean values of the y_i are very different in each case. To avoid this problem the standardised root mean square error (SRMSE) will be used. This is

$$\frac{\sqrt{(\Sigma(y_i' - y_i)^2)/n}}{\bar{y}}, \tag{4.6}$$

i.e the RMSE divided by the mean value of the y_i, and is thus a normalised representation of the error between the model and real data. This provides a useful measure of the model performance in predicting the observed values and will be used in this and subsequent chapters as a method of comparing model results.

4.9.1 Choice of Parameters

Variables that were identified as important from the literature and real data analysis were incorporated into the regression model. These will be identified and validated in the following sections.

§2.2.2 listed several external factors that play a part in the setting of petrol prices. These include the cost of crude oil, fuel tax and exchange rates. Each of these factors have a temporal impact on the data, e.g. the increase of crude oil price has a slight time lag before impacting on the petrol price. None of these variables will be included within the model. The results of the regression will be a snapshot of one day and the focus is to examine factors that may be linked to spatial variations, not temporal ones.

The multiple regression analysis was performed using SPSS 11.0 for Windows.

4.9.2 Population Density and Distance to Neighbour

To account for the influence of geography, two surrogate variables were used. These were population density (from each ward) and the proximity to the nearest station[5]. Population density is hypothesised to relate closely to the urban - rural divide. For example, urban areas have a higher population density than rural areas and sustain lower petrol prices (see §2.3.2 and §4.6 for further discussion). §2.3.1 and §4.7.2 identified a station's proximity to its nearest neighbour as a useful indicator of the degree of competition between stations and the resulting price. Esso, for example, set their prices to match the lowest within a 3 mile proximity.

For each petrol station, the population density of its ward was taken and the distance to the nearest neighbour calculated. The real data for West Yorkshire from July 27th was used. The regression equation was taken as

$$P_i = a \times D_i + b \times X_i + c \tag{4.7}$$

where

- P_i = price at station i

[5]Although these two parameters are measuring slightly different features (one is measuring the number of petrol stations and the other is measuring people) there is likely to be some correlation between them. This should be considered when assessing the results. The coefficients maybe smaller than expected because the effect of the population is shared between the two parameters.

- a, b, c = constants
- D_i = population density around station i
- X_i = distance to nearest neighbour of station i (straight-line).

The performance of the regression was poor, the r^2 value was 0.033 with a standard error of 1.04. The SRMSE was 0.140. The coefficients (Table 4.6) show that the constant is the most important factor in the equation. This represents the large proportion of the cost of the petrol that is from factors such as crude oil, taxation and transportation. These are the same for all of the stations. The variation in price between stations caused by geographical and other factors is only a small part of the total price. The population density coefficient is negative, showing that a high population density will reduce the price and vice versa. This supports the evidence for urban stations being cheaper than rural stations. However, this coefficient value is very low and therefore only exerts a small influence on the price. The distance coefficient is even smaller. The value is positive indicating that the greater the distance to the nearest neighbour, the larger the price. Evidence from the real data suggested that stations located close together (urban areas) are more likely to have lower prices.

Coefficient	Value
Constant	71.09
Population density	-0.0001175
Distance	6.921×10^{-5}

Table 4.6: Regression coefficient values from the regression analysis with population density and distance for West Yorkshire on July 27th.

It has been hypothesised that population density and distance between stations are important indicators of the urban - rural divide. The results from this regression would suggest that these alone are not sufficient to explain the observed variations. There could be several reasons for these results. Firstly, the regression takes a snap shot of the data, the stations may not be in equilibrium. Instability could be caused by an increase in crude oil prices, for example the rockets and feathers effect (Bacon, 1991). The analysis of the real data above brought to light not only the price differential between rural and urban areas, but also the large amount of variation within urban areas (see Figure 4.8). There are obviously other factors at play that the model is not taking account of. These other factors could be the pricing strategies operated by different types of stations or decisions taken by individuals based on experience and judgement. This point will be extended in the following section.

4.9.3 Supermarkets, Motorways and Esso

In the absence of detailed and commercially sensitive knowledge, it is impossible to understand how individual stations may determine prices. Evidence from the literature and real data analysis suggested that several groups of stations operated distinct pricing rules. These were Esso, supermarket and motorway stations. Esso has the widely publicised Price Watch; §4.6.2 identified supermarkets as being cheaper (by 1.2p on average) and more competitive than other classifications (Figure 4.11 showed that supermarket stations have consistent pricing across a region). The analysis undertaken in §4.6.1 showed that motorway petrol stations had higher prices and a larger SD than the other categories. Based on these observations and the recent work by McFarland (2003) highlighting that motorway stations respond as a separate system, motorways were given a separate category. The rest of the classifications were not included as either no such published information was available or there was no clear evidence from the real data that individual rules were operated.

The same data for West Yorkshire from July 27th was used but new indicators were added to the regression model,

$$P_i = a \times D_i + b \times X_i + c \times S_i + d \times M_i + e \times E_i + f \tag{4.8}$$

where

- S = supermarkets indicator (1=supermarket, 0=other)
- M = motorways indicator (1=motorway, 0=other)
- E = Esso indicator (1=Esso, 0=other)
- d, e, f = additional regression coefficients.

Adding the supermarkets, motorways and Esso to the population density and distance variables produced a considerable improvement in the performance of the regression model. The r^2 test gave a result of 0.253 with a standard error of 0.926. The SRMSE was 0.0135. Table 4.7 shows that both the population density and distance variables have further decreased in importance. However, the supermarket and Esso coefficients are both large and negative. This indicates that these stations are much cheaper than the other stations included within the regression. The motorway coefficient is smaller and therefore less important. Surprisingly it is negative indicating that motorway stations are cheaper than the other stations. The selection criteria for motorway stations (see Table 4.1) incorporated stations that were located within 0.5km of a motorway. Within West Yorkshire, this includes stations that are located in urban areas. This factor will obviously have impact on the importance of this variable.

Coefficient	Value
Constant	71.337
Population density	-9.30×10^{-5}
Distance	3.599×10^{-5}
Supermarket	-1.281
Motorway	-0.263
Esso	-1.164

Table 4.7: Constant and coefficient results from the regression analysis with population density, distance, supermarkets, motorways and Esso indicators.

The inclusion of the supermarket, motorway and Esso stations as separate indicators obviously made an important contribution to the predicted price. However, the results are not encouraging. An r^2 of 0.253 means that the variables are only explaining 25% of the variation. The multiple regression model gave results that indicated that geography was not a significant factor, despite strong evidence from both the literature and analysis of the real data. Multiple regression does not take into account the complex and non-linear effects of locality, the inclusion of these may improve the results.

4.10 Geographically Weighted Regression (GWR)

Geographically weighted regression (GWR) is an extension to multiple linear regression which adds in a spatial element (Brunsdon *et al.*, 1998a,b). Equation (4.3) shows a typical multiple linear regression model. GWR extends this by making the coefficients, a and b_i, a function of position so

$$y = a(s) + b_1(s)x_1 + b_2(s)x_2 + b_3(s)x_3 + \ldots \tag{4.9}$$

where s is the position at which the variable is being estimated. This spatial variation is included by weighting the contribution, w_{ij}, to the regression at point i from the data at point j according to the distance d_{ij} between the two points. A parameter b, the bandwidth, is used to set a distance scale over which the weighting decreases.

The variation in the regression coefficients means that the regression model can take into account spatial variations in the relationship between the x_i and y caused by other spatial factors which are not included in the regression model through the x_i. This is particularly useful if the system under analysis is thought to be significantly influenced by its locality or other points within it. Use of this technique will hopefully improve upon the results of the multiple regression. The GWR was performed using the package developed by Fotheringham *et al.* (2002) at the University of Newcastle[6]. The reader is directed to Fotheringham *et al.* (2002) for a detailed synopsis.

Price was again selected as the dependent variable. Population density, distance, supermarkets, motorways and Esso were input as the independent variables. The data from July 27th for West Yorkshire was used. The fixed and adaptive bandwidths (number of data points used to perform the regression at any one point) were both experimented with. The adaptive bandwidth was used in the final regression as this gave a slightly better r^2 value because it gave a larger neighbourhood search in areas that contained fewer petrol stations. 166 stations were used within the regression; although this is a large figure, only those stations nearest the regression point will have a large weighting and thereby a significant impact. In addition to this, the standard weighting function was used. This is a bisquare function where the weighting function w_{ij} is given by

$$w_{ij} = \begin{cases} (1 - d_{ij}^2/b^2)^2 & (d_{ij} \leq b) \\ 0 & (d_{ij} > b) \end{cases} . \tag{4.10}$$

The bandwidth b is chosen adaptively by the GWR code to minimise the r^2 value.

GWR does not produce global coefficients. Instead, comparison with the multiple regression

[6]Details on the software can be found at www.ncl.ac.uk/geps/research/geography/gwr/.

model will be achieved through r^2. The r^2 for the GWR was calculated at 0.323 (with a standard error of 0.903). This is an improvement on the results from the multiple regression modelling (an r^2 of 0.253). This translates to a SRMSE of 0.0133, again better than the multiple regression models. This suggests that spatial distribution of the petrol stations does have an effect on price setting within West Yorkshire in addition to the density and nearness metrics, if not as important as originally suggested. Other factors appear to exert more influence. For example, the multiple regression analysis identified both supermarkets and Esso as having a significant impact on the price.

Figure 4.19 shows spatially the results of the multiple regression and GWR on the July 27th data (the real data is included for comparison). The multiple regression (Figure 4.19(b)) has identified some of the lower priced areas, but has failed to detect a large amount of the variation. The GWR (Figure 4.19(c)) has improved on the multiple regression model performance by identifying more of the spatial differentiation in prices.

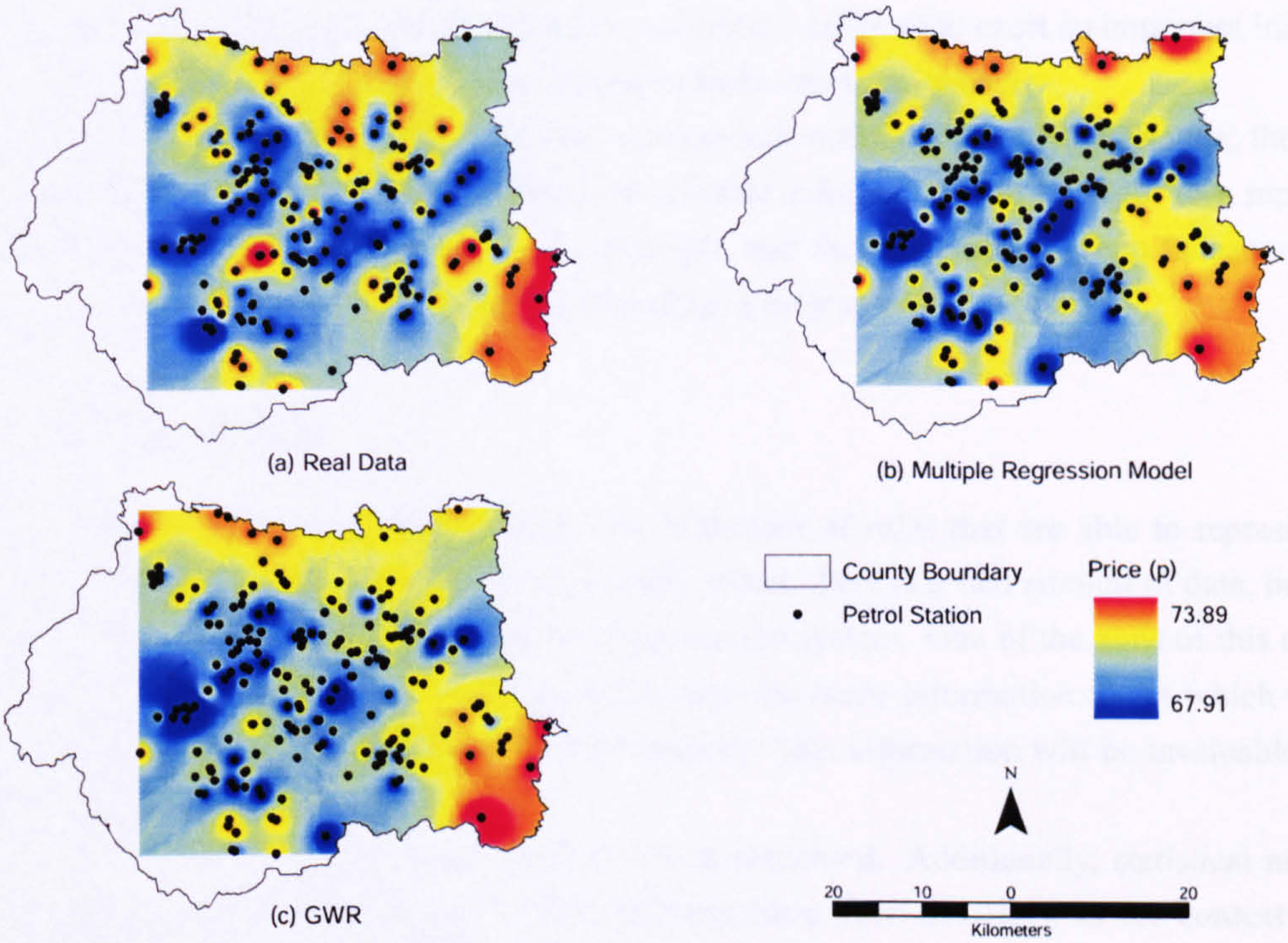


Figure 4.19: Real data from July 27th (a) compared with the results from the multiple regression (b)and GWR (c).

4.11 Summary

The results of both the multiple regression analysis and GWR suggest that the factors controlling price are more complex than can be identified through a regression equation. For example, neither the multiple regression model or the GWR can incorporate potentially important factors such as market strategy or competition. One of the interesting results that came to light was the lack of influence that geography appears to have. The coefficients from the multiple regression model gave little significance to population density and distance to neighbour, but a high significance to the effect that different brands of stations exert. GWR did improve on the results of the multiple regression model. This improvement suggests that locality does play some part in the final price.

One of the largest criticisms of mathematical models is that it is difficult to take into account the actions of individuals and therefore the modifications to the environment which result from their behaviour. This makes it impossible to examine, for example, the effects of the Esso Price Watch policy or the constrained pricing of supermarkets at more than more one spatial level. As Ferber (1999) highlights,

> "If we consider actions only in terms of their measurable consequences at the global level, or of their probability of appearance, it will be difficult to explain phenomena emerging from the interaction of these individual behaviours, in particular all those relating to intra- and inter-specific cooperation."

It is the interaction of these individual behaviours that are believed to exert an important influence on the price and this is therefore a grave failing of these empirical models.

The final criticism that can be addressed at empirical models is that by their nature, they only consider metric parameters. Vast amounts of valuable information can be input to a model by use of behavioural data. This could be for example, that the supermarket group Sainsbury's are offering a discount on petrol to customers who shop at their stores.

4.12 Conclusion

An important focus of this research is the establishment of rules that are able to represent the reality of the petrol pricing market. As previously stated, there is a vast amount of data, but little published information concerning what rules govern the system. One of the aims of this chapter has been to thoroughly analyse the real data to provide more information about which factors control the system and the magnitude of their impact. This information will be invaluable when constructing and testing a model.

Two methods of classifying the data have been presented. Additionally, statistical methods of comparing predicted model prices with real data have been described in the context of the regression models. These methods, along with the classifications, will be used to analyse the performance of the model output in subsequent chapters.

Based on the real data analysis presented within this chapter, the petrol price market has been characterised as strongly non-linear with step changes in price. There is a strong spatial element to the pricing of the petrol with a clear rural-urban divide. Different market strategies implemented by different brands were evident, for example the Esso Price Watch and the aggressive pricing of supermarkets.

An attempt to incorporate these factors into regression models was made. Using multiple and geographically weighted regression to model the prices gave relatively poor results. The conclusion was drawn that for the petrol price market, empirical models are not appropriate. They do not account for all of the spatial variations observed in the real data nor can they represent changes in the system over time. In the following chapter, an agent-based model will be developed which attempts to address some of these failings.

Chapter 5

Agent Model Development

5.1 Introduction

Chapter 2 highlighted the traditional empirical approaches that have been used to model the petrol price market. Two such methods (multiple regression and geographically weighted regression (GWR)) were applied in Chapter 4 to the real data to investigate their suitability. Neither technique performed well and the conclusion was drawn that there were severe limitations in their application to the petrol price market. The focus shifted to using a different type of technique, agent-based modelling. Agents were introduced in Chapter 3, their strengths discussed and through the variety of applications detailed, their flexibility highlighted.

Agents overcome many of the problems associated with more traditional empirical techniques. For example, agent architectures can be decomposed into small blocks with individual tasks that together fulfil their purpose. These tasks could be unique rule sets (i.e. pricing policies) applied to an individual or group of stations. The impact of assigning these policies can be analysed on both a small scale (immediate neighbourhood) and larger scale (regional level). This type of analysis cannot be performed using empirical models.

This chapter describes the conceptual and technical construction of an agent model based on the non-consumer empirical models discussed in Chapter 3. This is followed by the development of a rule set and testing of performance against the real data. Individual rule sets are assigned and the results discussed.

5.2 Agent Model Framework

The agent model is designed to represent as accurately as possible the reality of the petrol price market. The conceptual framework, architecture and structure of the agents has drawn heavily on the available literature and analysis performed on the real data in Chapter 4. The following sections will provide details of the conceptual framework adopted. Details of the technical construction are provided in §5.3.

5.2.1 Conceptual Framework

As outlined in Chapter 3, agent models consist of subsystems, subsystem components, interactions and organisational relationships. For example, in simplistic terms, changes in petrol prices at a local level are thought to be influenced by location of competitors and amount of custom. At a national level, influences such as taxation and crude oil prices become significant. All these factors interact in some way to affect the final price. In this thesis, the focus is more towards local variations which have been much less extensively studied.

The main notion behind the agent model is to emulate as closely as possible the rules which are perceived to govern the way petrol prices are set in reality. This approach contrasts sharply with traditional statistical methods that search for patterns or trends in the data. The results of two such empirical techniques (multiple regression and GWR) were examined in Chapter 4 (see §4.9 and §4.10). The conclusion was drawn that these techniques are largely unsuitable for modelling a complex system such as the petrol price market.

Conceptually, petrol stations can be seen as discrete objects with a degree of control over their own pricing. Evidence from the literature (see Chapter 2) and detailed analysis of the real data (Chapter 4) suggests that there are several important factors that control petrol prices. Of these, rules dictated by corporate policy are perhaps the most important. These rules react to the behaviour of competitors, normally within a local neighbourhood[1] aiming to both retain competitiveness and maximise profit. The Competition Commission (1990) identified several variables that form the basis of these rules, for example, size of neighbourhood, price change per day, undercutting and overpricing amount. Analysis of the real data in Chapter 4 enabled several of these parameters to be assigned a range of potential values.

External factors, for example crude oil prices, exchange rates and fuel tax were also identified in Chapter 2. Unlike the effects of corporate rule sets and local competition, these factors are thought to have an equal effect on all of the stations regardless of geographical location or station type.

Multi-agent systems (MAS) were introduced in §3.3. MAS are a natural metaphor for modelling the petrol price market. Within a MAS, independent agents (petrol stations) can be created and assigned individual rule sets (corporate policy). Co-operation between agents for sharing information can also be built in (reacting to local prices) as required. In addition, an agent methodology is particularly suited to this application for the following reasons:

- Flexibility, i.e. different rule sets can be assigned to different stations.
- Potential to link to other techniques to optimise a solution (e.g. Evolutionary Algorithms).
- It agrees well with the conceptual idea of how the prices are set in real life; theoretically it is a good representation.
- The agent framework can give temporal variations, enabling assessment of the time period required for the system to reach a steady state.

[1]According to Ning and Haining (2003) a local neighbourhood is the distance from a station to one or more "reference stations".

- It is easy to add external influences into the model, e.g. oil prices.

These advantages overcome many of the problems related with traditional, empirical techniques. For example agent models allow examination of the impact of unique behaviours at different geographical scales and over time. Other criticisms with empirical techniques are summarised in §2.6.

5.2.2 Agent Structure

There are numerous definitions of what an agent is (see §3.2 for further details). The "petrol" agents created within this model will be defined by the requirements of the application. This follows the mode of definition proposed by Brenner *et al.* (1998).

The agents are *heterogeneous*; despite having fixed locations, each station can potentially have a different rule set and price; *communicative and cooperative* (though not anti-competitive) with pricing and location information shared between the agents for competition and *reactive*, making decisions and changing their prices based on information supplied to them. The structure is *multi-agent* in the sense that different groups of petrol stations for example, Esso, can be assigned their own unique rule set while the rest of the petrol stations operate an entirely different one. Figure 5.1 presents the mechanism for an individual agent.

```
For each neighbour {
    get price of neighbour()
    get distance to neighbour()
}
Calculate new price based on pre-defined rules;
Repeat until simulation finished.
```

Figure 5.1: Pseudo code for an individual petrol agent

5.2.3 Choice of Architecture

There are an ever increasing number of agent architectures available. Of the architectures outlined in §3.3, multi-agent systems (MAS) are the most appropriate (see also §3.7). Figure 5.2 shows a schematic of the MAS that will be implemented. This architecture was chosen because of the modularity and flexibility that it will bring to the model. This is ideal for modelling complexity and large systems. Different agents (petrol stations) may use different rules depending on their location and corporate policy. For example, allowing the agent to assess prices of other stations within its neighbourhood, under-cutting competition etc. A key requirement in modelling competition in an economic system is that the agents are reactive (i.e. they maintain an ongoing relationship with their environment). Use of a MAS allows the agents to be reactive.

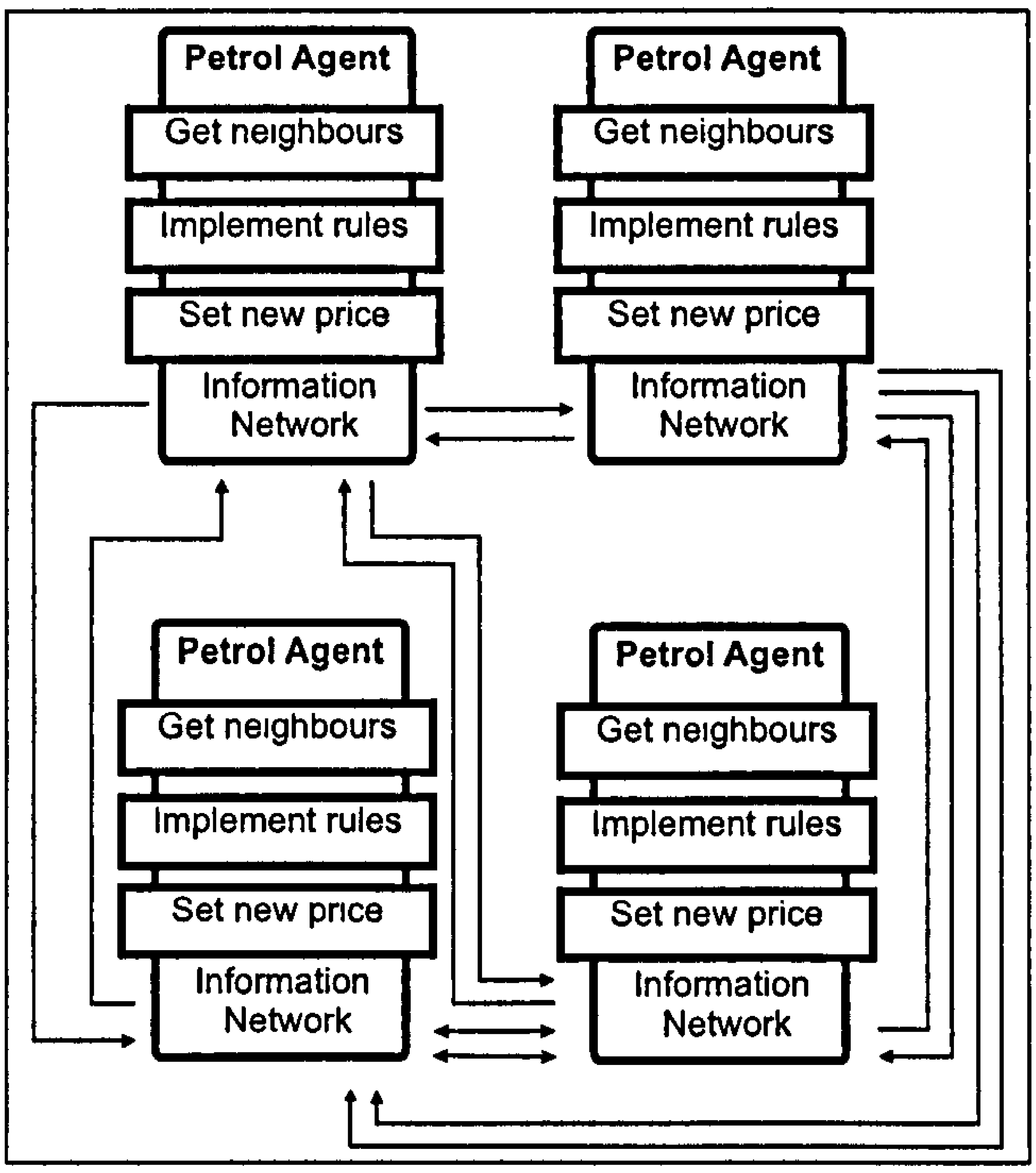


Figure 5.2: A set of petrol agents (following the style of Tsvetovat and Carley (2002))

5.2.4 Knowledge within the System

The petrol agents possess knowledge that enables them to complete their tasks. They are aware of their neighbours' location and fuel prices on the current (t) and previous day ($t-1$), additional data can be fed in as required. Feeding the agents the previous day's price was implemented to provide the system with knowledge of the current trend in price changes as well as the current price. However, two implementation problems arose with this approach. The readings for each petrol station on each day are not complete, i.e. not every station has a reading every day (i.e. at the start of a model run). In cases where there are such absences, the price at t was used as $t-1$. Secondly, the data set itself is not continuous: there are several gaps within the data set limiting the number of days that the simulation can be compared with reality. This problem was overcome by using days at the beginning of a run of consecutive days (e.g. 27th July for the early part of the data set).

5.2.5 Rules

A crucial stage in the development of the model is building an appropriate rule set that models reality. To this end, combinations of rules will be implemented in §5.5 and their performance assessed against the real data in §5.6. The rule sets will be based around the variables detailed in Table 5.1. These variables were chosen on the basis of industry knowledge supplied from the literature[2] and detailed analysis of the real data undertaken in Chapter 4. Figure 5.3 shows how these rules might be implemented by an individual agent while Figure 5.4 illustrates how the model

[2]In particular, the Competition Commission (1990) and research undertaken by Ning and Haining (2003).

would operate including this agent behaviour.

Rule	Explanation
Maximum ***undercutting*** **price**	The amount by which petrol stations can undercut competitors, e.g. 1p, 2p.
Maximum *overprice*	The amount by which petrol stations can be more expensive than their neighbours without becoming uncompetitive and being forced to cut their price.
Maximum *priceChange*	The amount by which petrol stations can change their price per day.
Neighbourhood	The distance that the petrol stations will treat as their neighbourhood e.g. 2km, 3km.
Type of brand	Which group of petrol stations is to be used e.g. supermarkets.
Type of petrol	Which type of petrol is to be assessed.
Region	Selection of particular area, e.g. county, for analysis.

Table 5.1: Summary of the rules that can be potentially implemented by the agent model.

5.3 Construction of the Agent Model

The following work presents the technical construction of the agent model. An overview of the model classes are supplied along with details of other issues, for example reading data in, design of the graphical user interface (GUI) and visualisation of initial results. The section finishes with details of statistical methods used in assessing the performance of the agent model. Experiments are undertaken to assess the impact on the results of the model of using data from various sized geographical regions.

5.3.1 Development Environment

An object orientated language such as C++ or Java is particularly suitable for the development of an agent code because they both possess a similar conceptual basis. It is natural to implement an agent as an object. Java possesses the advantages of being easy to develop, multi-platform and readily available. Furthermore, it is also easy to link to other programs, for example, GIS software, web-based packages and other AI techniques (Heppenstall, 2001). For these reasons, Java was chosen as the programming language for this project.

5.3.2 Overview of Model Implementation

In keeping with the philosophy of agent-based modelling, a modular approach was taken in the design and implementation of the model. Classes were created to contain various distinct parts of the model, for example, the user interface and the agents.

The model can be run in two modes. The first is an interactive mode with a GUI to allow the user to select the input files and specify all the model parameters. The second mode is non-interactive with all the parameters being read from a control file. This mode is run from the

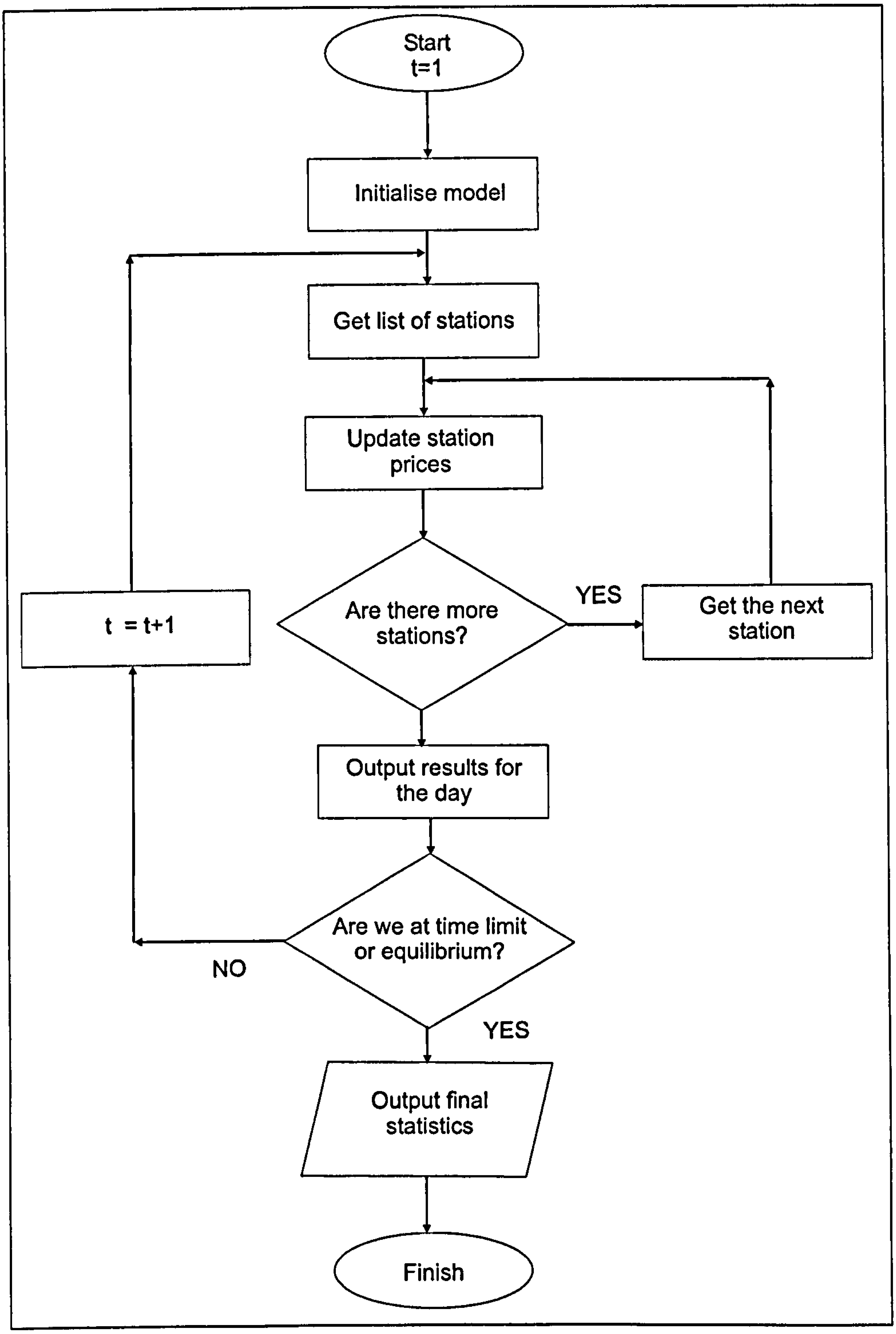


Figure 5.3: Flowchart illustrating operation of rules within the petrol agent during one iteration.

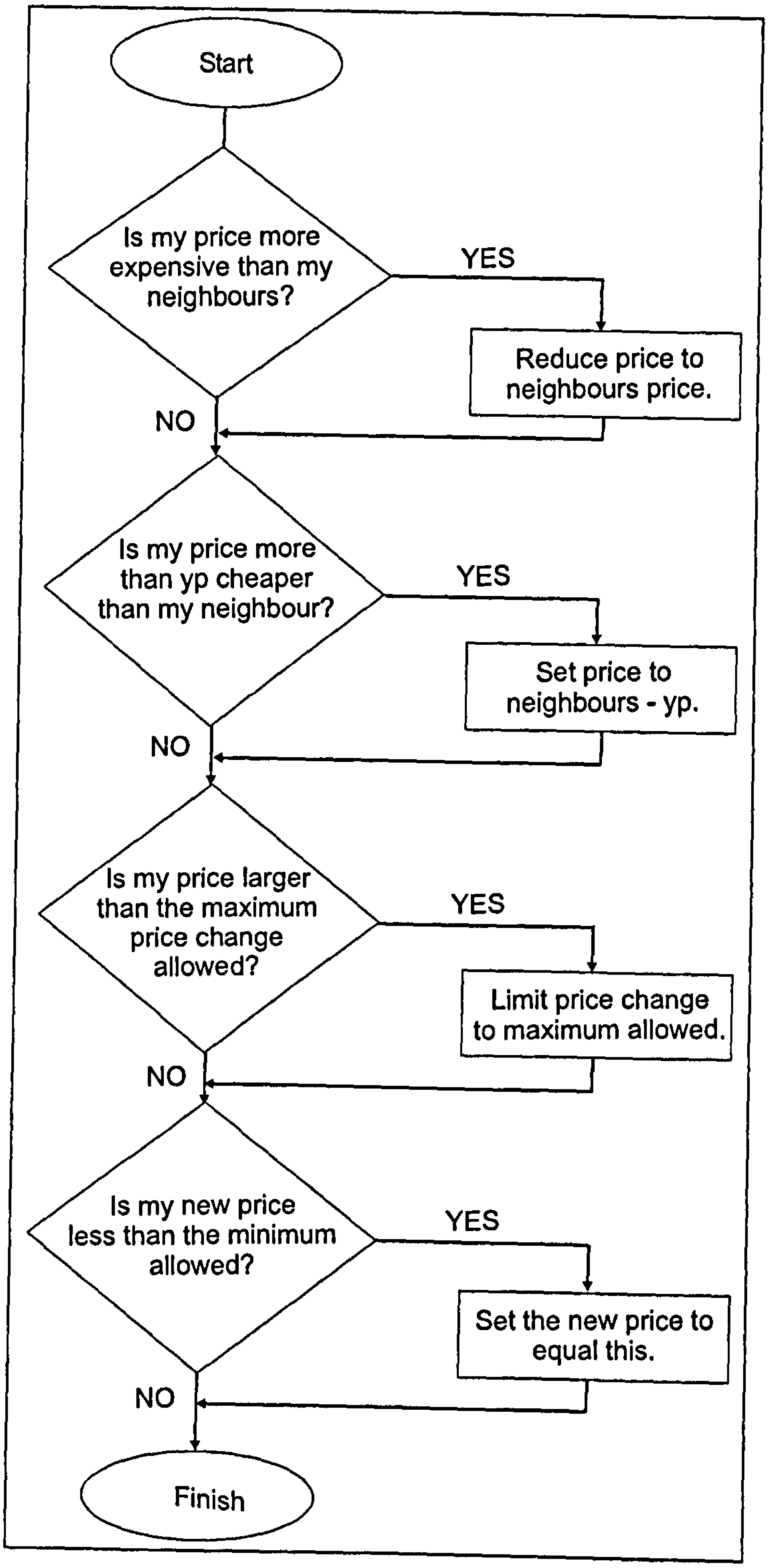


Figure 5.4: Flowchart illustrating how the model runs with a typical rule set.

command line and does not require the GUI so it is suitable for running a batch of pre-defined model runs, running on a remote system or linking with other software (see e.g. Chapter 9).

Figure 5.5 presents a unified modelling language (UML) schematic of the interactions between the methods and classes of the agent model (variables are not included). The two operational modes are implemented through a set of classes. The GUI class creates and controls the windows and dialogs necessary for the graphical user interface. The Control class is an alternative to the GUI class and reads in all the parameters from a specified control file. Both of these classes create an instance of the Simulation class and set the initial conditions and parameters for the model. The Simulation class contains the various data and methods needed to coordinate and run the simulation. The Simulation class also creates an instance of the Petrol class for each agent (petrol station). The Petrol class contains the data and methods pertaining to each station, for example the variable containing the petrol price and the methods to get and set that price. There are other ancillary classes, such as the GlobalRules and Rules classes which contain details of the rule set in operation.

The full source code and a compiled version of the model can be found on the CD included with this thesis.

5.3.3 Graphical User Interface (GUI)

The GUI was designed to enable the user to easily select the required data, the rule sets and simulation parameters. Figure 5.6 (a) shows the menu designed for selection of data. The dialog used for assigning rule sets is shown in Figure 5.6 (b). Parameters can be set as default or specified. Parameters can also be turned off. This provides a very powerful framework for the agent model to work within. Figure 5.6 (c) shows the dialog box used for setting the simulation time and intervals to save data at.

5.3.4 Reading in Initial Data

The data was originally supplied in two separate files in .mbd (Microsoft Access) format. The first file contained the day of the month with petrol prices and ID number, whilst the second contained additional information such as the location and type of each petrol retailer. These files were exported as .csv files (comma separated values) which could be read into the agent model.

The data from the two files had to be matched up by ID to create a complete record for processing. To ensure that only complete data was fed into the model, data from one file was not used if it did not match up with the second file.

Once the data had been read into the model, a subset of the data can be selected based on a variety of conditions, for example, petrol station type (e.g. Esso) or geographical area (e.g. West Yorkshire). This filtering is done purely based on matching the relevant field in the second input file containing the petrol stations' details. Several lots of data can be read in; for example all the Esso stations could be read in first and then all the Shell stations. The agents are created and initialised using this filtered data.

Some additional consistency checks are made to ensure the petrol prices are reasonable. A visual inspection of the database showed that there were some errors in the recorded petrol prices

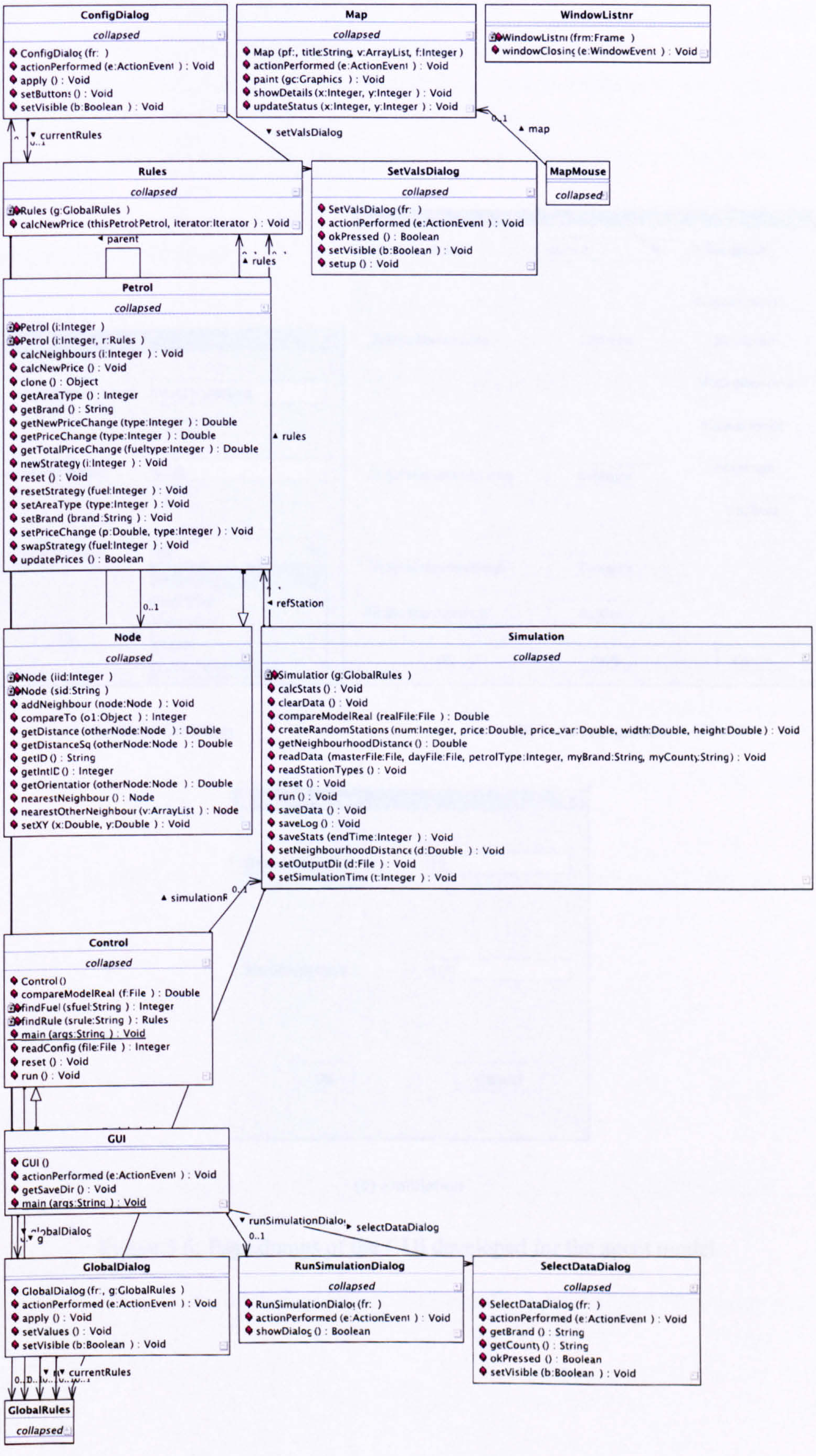
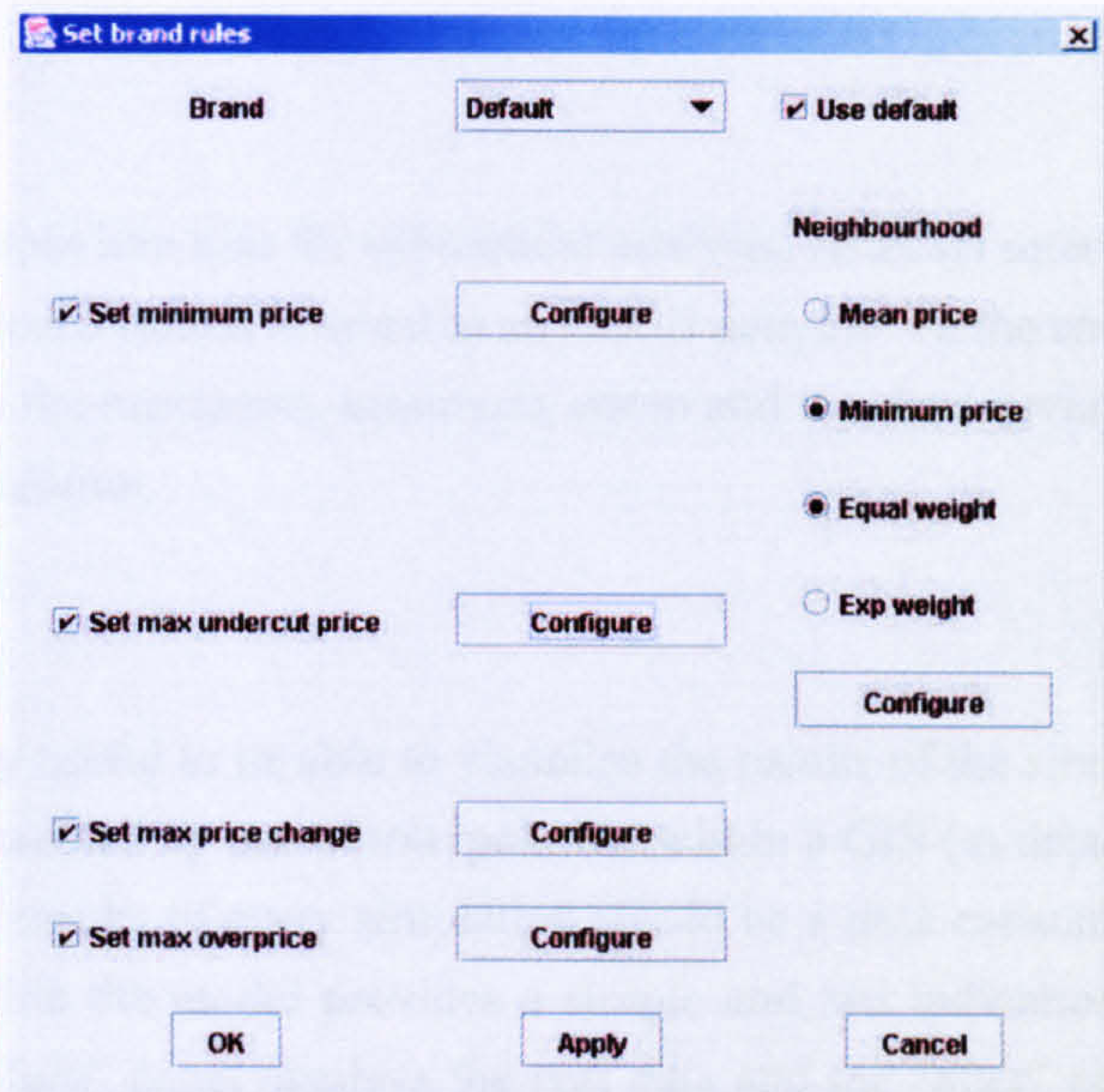
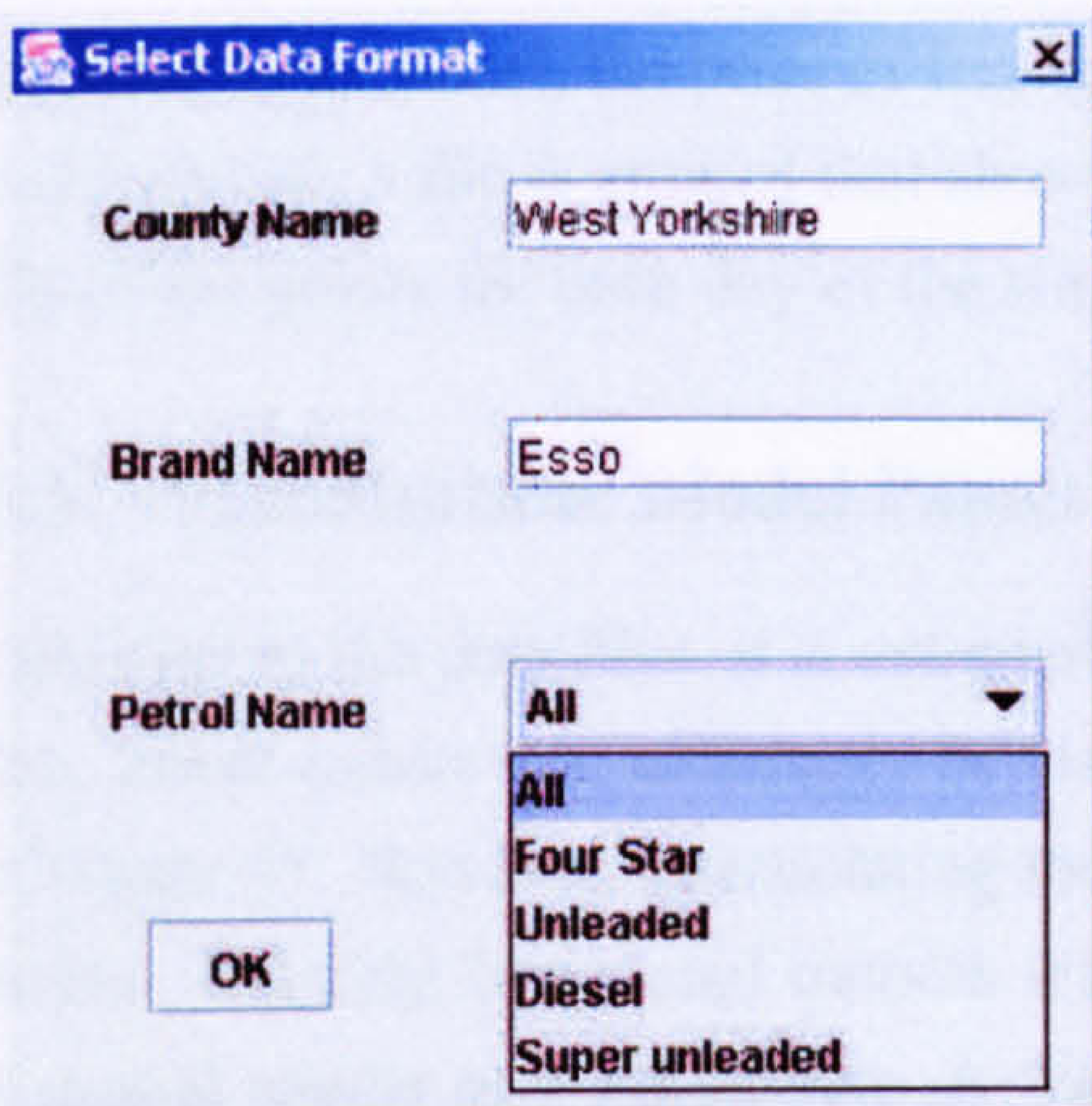


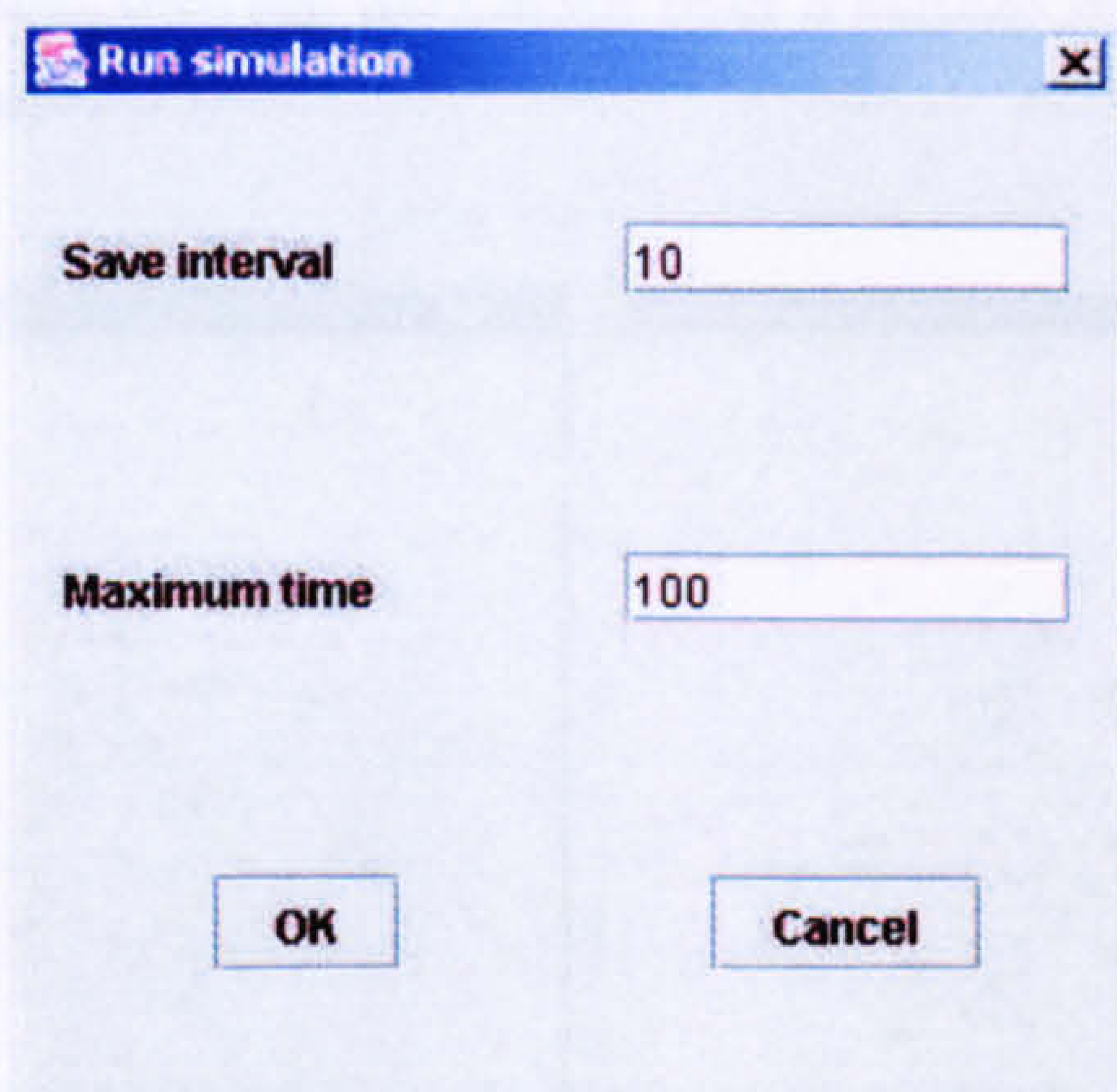
Figure 5.5: UML of the classes of the agent model.





(a) Data Selection

(b) Rule Assignment



(c) Simulation

Figure 5.6: Page dumps of the GUI developed for the agent model.

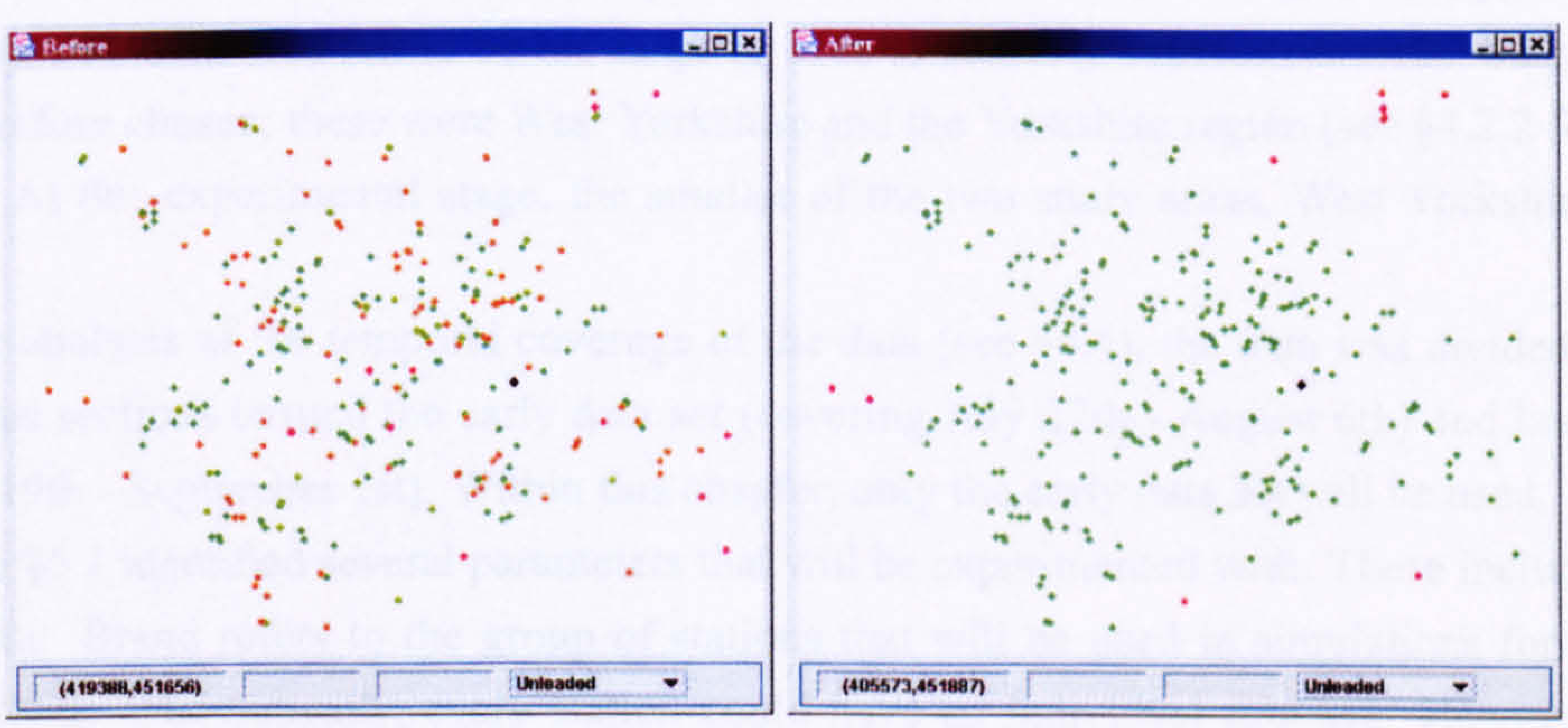
at a few stations. These would be characterised by a very high or low petrol price, often only on one day with sensible values on the previous and subsequent days. To eliminate these points only unleaded fuel prices in the range 65-75p were accepted when reading in data. Values outside this range were ignored and the station was treated as if there was no recorded price for that day. This range was taken so that it fully included the observed range of prices, whilst eliminating the suspect values.

5.3.5 Model Output Files

During and after the simulation, data is output into files for subsequent analysis. After set intervals (specified by the user), the price of fuel at each station is saved to an ASCII data file. At the end of the simulation, a file is created that shows the maximum, minimum, mean and standard deviation (SD) of the prices for each day of the simulation.

5.3.6 Visualisation: Model Panels

In addition to the data files, it is extremely useful to be able to visualise the results of the simulations. These results will ultimately be visualised by use of interpolation within a GIS (as detailed in Chapter 4). However, interpolating the results of every simulation would be a time consuming process. Creating two visual outputs within the model provides a simple and fast indication of the spatial results of a simulation. A "before" panel displays the real data and the "after" panel displays the last day of the simulation (Figure 5.7). A colour scale is used to differentiate between high and low prices (green:low - red:high). Black points represent the stations that do not sell that brand of petrol. In the software, moving the cursor over a station will display the coordinates and the fuel price in the status bar.



(a) Before (b) After

Figure 5.7: Example of "before" (a) and "after" (b) model panels displaying the location of petrol stations and their prices.

5.3.7 Statistical Analysis of Model Data

The model can output prices at each time step. Statistical methods will be used to assess the performance of the model in comparison to the real data. The mean, standard deviation (SD) (see §4.4.1) and the standardised root mean square error (SRMSE) (see §4.9) will be used. A combination of these techniques will provide a measure of how the model is performing.

Two scripts were written to assess the differences in the real and model data at the same geographical point, these are:

- **DIFFS.SH:** To calculate the difference in price between the real and model data at each petrol station on a given day.
- **STATS.AWK:** Using the output from the DIFFS.SH, this script calculates the mean, SD, Root Mean Square Error (RMSE) error and the SRMSE of the differences between the real and model data at each petrol station. This can also be used for the prices as well as price differences.

These scripts can be found in Appendix C.

5.4 Details of Model Simulations

This section will briefly provide details of the set up of the simulations. This will include a description of the data that will be used and the initial conditions.

5.4.1 Model Inputs

In Chapter 4, the data was described in spatial and temporal terms. Geographically, the data set covers the UK. This was felt to be too large an area to sensibly experiment with. Smaller areas were therefore chosen; these were West Yorkshire and the Yorkshire region (see §4.2.2 for further details). At this experimental stage, the smaller of the two study areas, West Yorkshire will be used.

After analysis of the temporal coverage of the data (see §4.4), the data was divided into two continuous sections termed the early data set (covering July 27th - August 6th) and late data set (August 19th - September 1st). Within this chapter, only the early data set will be used.

Table §5.1 identified several parameters that will be experimented with. These included brand and region. Brand refers to the group of stations that will be used in simulations for example, supermarkets. Theoretically, every station could have an individual rule set. However, at this stage running the model with all the agents operating separate rule sets would be unnecessarily complicated. Assigning the same rule set to all the agents whilst varying parameters will enable a clear indication of the impact of each variable. Individual rule sets will be assigned in §5.8.

The largest amount of data available is for unleaded petrol, simulations will therefore be run using this data. Each model experiment will be run to equilibrium. As the real data is daily, the model is run with one price change equalling one day. For comparison with the real data, the 10th day of the model run will be compared with day 10 of the real data (6th August).

In summary, unleaded fuel prices from West Yorkshire covering July 27th - August 6th will be used. The agents will all operate the same rule set.

5.4.2 Study Area Size

§4.3 and §5.4.1 stated that only those stations contained within the West Yorkshire boundary would be used during experimentation. This section will assess how much information could be lost by using this technique. The most obvious place for loss of data is at the edge of the study area. However, will the overall pricing patterns and those in specific locations, e.g. the cities located next to the boundary be affected? The purpose of this section will therefore be to assess the impact of varying the number of points in the interpolation, not an examination of how well the model is performing.

The agent model was initialised with the real data from July 27th and simulations ran with the default parameters. These were *neighbourhood*: 5km; *undercutting*: 1p; *overpricing*: 1p and maximum *priceChange*: 1p (experimentation with these parameters is undertaken in §5.5). Table 5.2 summarises the geographical areas included within each model run:

Model Run	Geographical Areas Used
a	West Yorkshire
b	Yorkshire Region
c	Surrounding Counties

Table 5.2: Summary of the geographical extent of data included within each model run.

For comparative reasons, each model run was interpolated after 10 days.

Figure 5.8 shows that inclusion of all the petrol station data surrounding West Yorkshire (Figure 5.8(c)) has had the greatest impact on the highly priced stations on the edges of the study area. With the presence of competitive neighbours, these stations are now approximately 3-4p cheaper. The impact of the Yorkshire region data (Figure 5.8(c)) is not as great with the increase in price of several of the edge stations by approximately 1p.

Away from the edges, there are no significant differences in the price of petrol between the three simulations. The only discernible difference between the surrounding counties (Figure 5.8 (c) and West Yorkshire (Figure 5.8 (a)) can be seen in the eastern petrol stations (marked with an arrow) where the stations in the surrounding counties have a slightly higher price (of a magnitude of approximately 0.5p).

In summary, using data from all the surrounding counties instead of just West Yorkshire or the Yorkshire region has had the effect of eliminating some of the highly priced edge stations. However, there are no significant differences in the spatial pricing trends. West Yorkshire will be continued to be used.

5.5 Building a Rule Set

One of the main objectives of this research is to model reality of the petrol price market as we understand it, i.e. the main processes and drivers of a system as identified through the literature

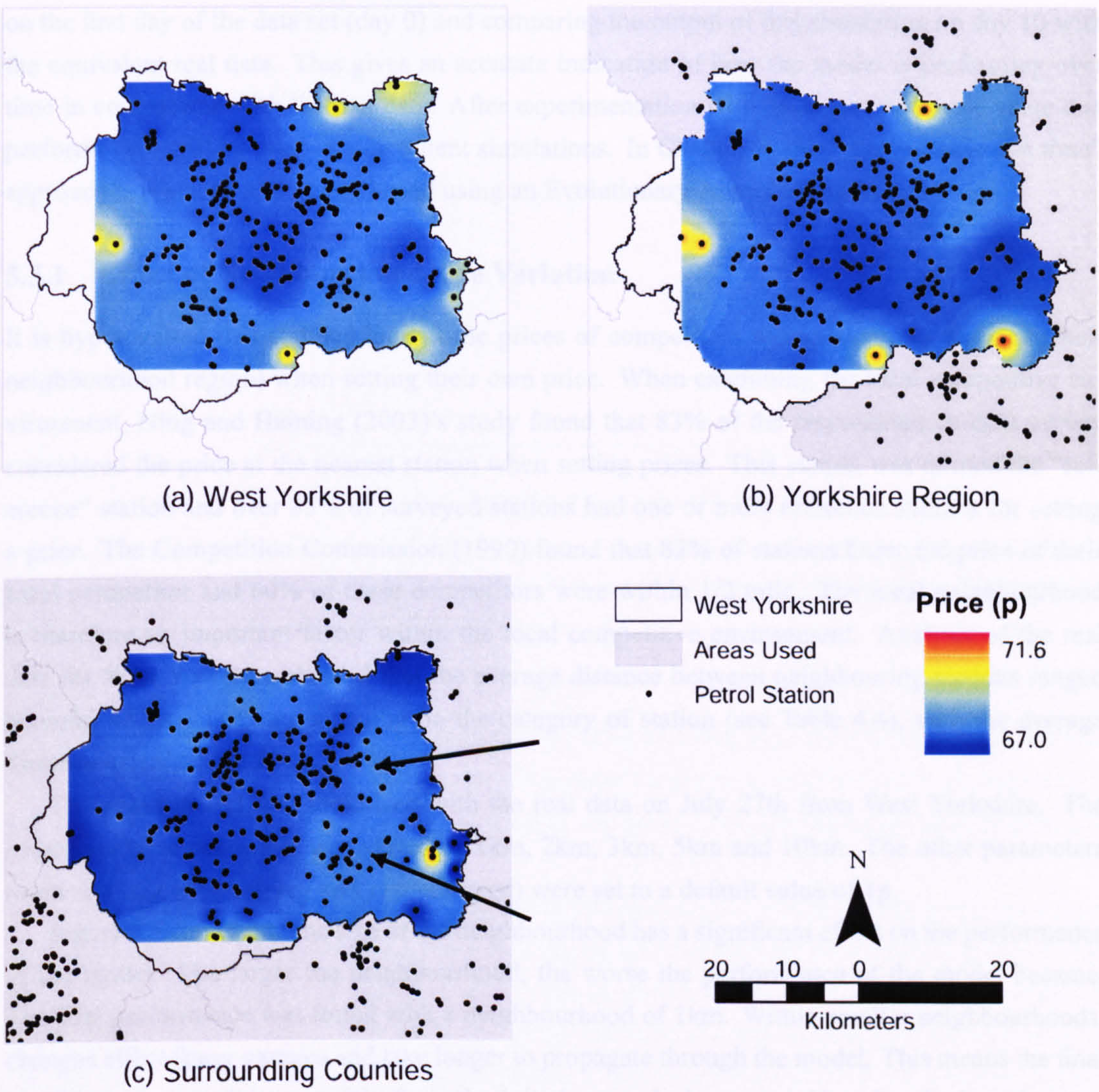


Figure 5.8: Interpolation of model output after 60 days for West Yorkshire using (a) points within West Yorkshire (b) Yorkshire Region and (c) all the surrounding counties. The arrows indicate regions of interest discussed within the text.

and detailed analysis undertaken in Chapter 4. The precise rule set used and the parameter values within the rules will be adjusted to best represent the real data. A crucial element of this is testing that the model is performing sensibly. Adjustment of parameters to sensible values can be achieved via industry knowledge and experimentation by drawing on available industry knowledge and real data analysis.

The following sections will experiment with parameters that can be adjusted to reflect the characteristics of the data. These parameters are: *neighbourhood* size, price *undercutting* amount, maximum *overpricing* amount and maximum *priceChange*. Basic definitions of these parameters can be found in §5.2.5; an explanation of the significance of each parameter will be supplied as they are adjusted. Examining the impact of parameters will be achieved by initialising the model on the first day of the data set (day 0) and comparing the output of this simulation on day 10 with the equivalent real data. This gives an accurate indication of how the model is performing over time in comparison with the real data. After experimentation with each parameter, the value that performs best will be used in subsequent simulations. In Chapter 9, this "one variable at a time" approach is replaced and strengthened using an Evolutionary Algorithm approach.

5.5.1 Effect of Neighbourhood Size Variation

It is hypothesised that stations look at the prices of competitors within a certain distance (their neighbourhood region) when setting their own price. When examining the local competitive environment, Ning and Haining (2003)'s study found that 83% of the respondents to their survey considered the price at the nearest station when setting prices. This station was termed the "reference" station and over 85% of surveyed stations had one or more reference stations for setting a price. The Competition Commission (1990) found that 82% of stations knew the price of their local competitor and 60% of these competitors were within 1/2 mile. The local neighbourhood is therefore an important factor within the local competitive environment. Analysis of the real data for West Yorkshire showed that the average distance between neighbouring stations ranged between 482m - 1122m depending on the category of station (see Table 4.4), with the average distance calculated at 701m.

The agent model was initialised with the real data on July 27th from West Yorkshire. The *neighbourhood* parameter was varied by 1km, 2km, 3km, 5km and 10km. The other parameters (*undercutting*, *overpricing* and *priceChange*) were set to a default value of 1p.

Figure 5.9 shows that the size of the neighbourhood has a significant effect on the performance of the model. The larger the neighbourhood, the worse the performance of the model became. The best performance was found with a neighbourhood of 1km. Within smaller neighbourhoods, changes affect fewer garages and take longer to propagate through the model. This means the final conditions do not change greatly from the initial ones. In larger neighbourhoods the prices are being over-predicted. Stations are reacting to competitors located at a distance. These competitors may be, for example in different geographical areas and subject to different natural variations in prices. As the neighbourhood size increases the price differences get closer to the mean price difference presumably because the variations in the price differences are averaged out. This leads to a smaller standard deviation (SD) even though the mean price difference actually increases. There is a relatively large error with little variation.

5.5.2 Effect of Price Undercutting

The *undercutting* parameter sets the maximum amount by which a station should attempt to undercut the other stations in its neighbourhood. If it is more than this amount cheaper than its competitors it can raise its price to the undercut level to increase its profit while still remaining cheaper. This introduces realistic behaviour into the model by allowing stations to remain both competitive and to increase their profit. The experimentation within this section will determine how a neighbourhood reacts if a competitor lowers its prices. Will price changes diffuse throughout the area, or will they remain static? To test this, the *undercutting* parameter was set to 0.5p,

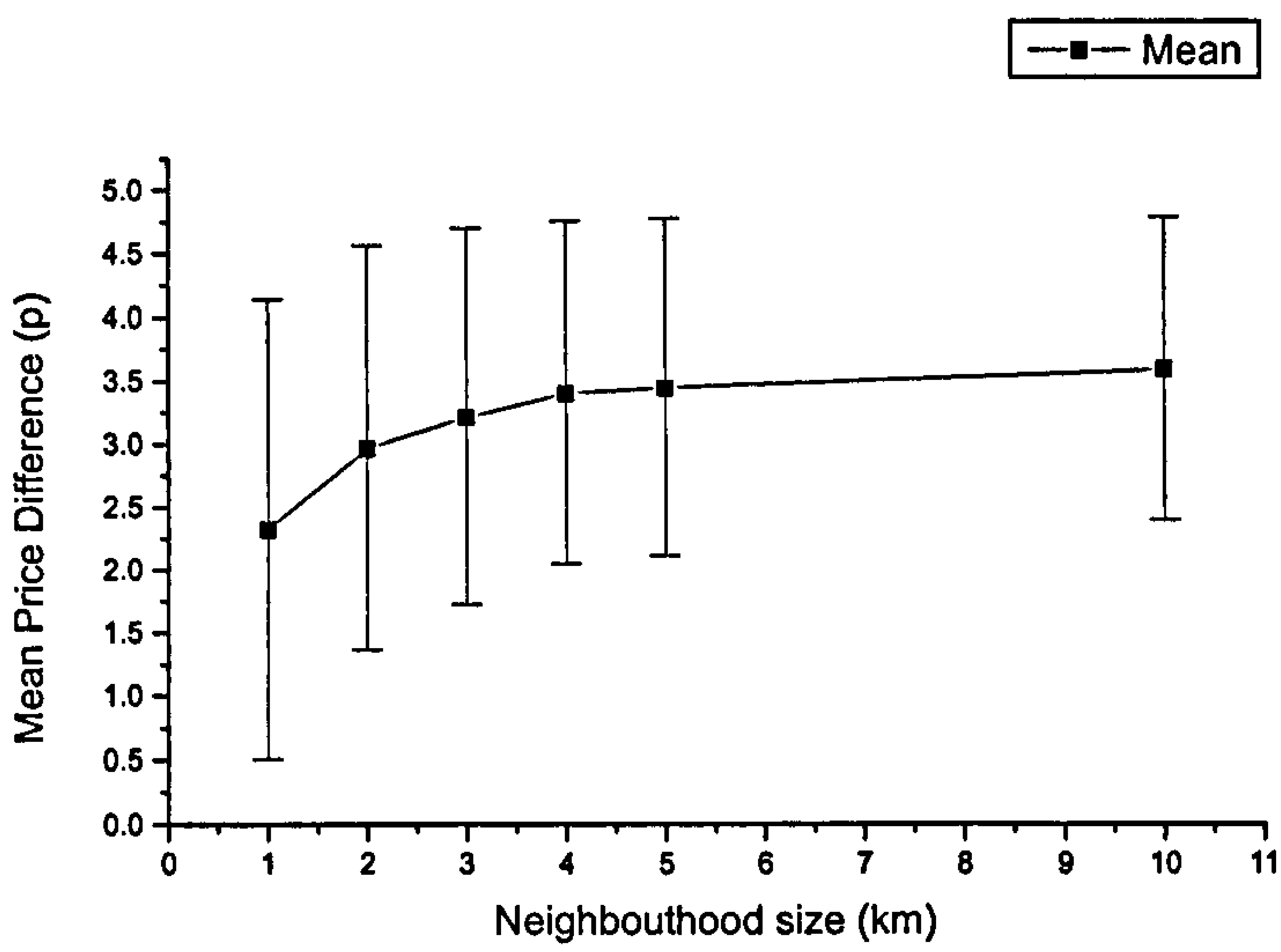


Figure 5.9: Effect of different neighbourhood sizes on the mean price difference and SD of the price differences (represented by vertical bars). Maximum *undercutting* price set to 1p; maximum *overprice* 1p and maximum *priceChange* 1p.

1p, 2p, 3p and 5p. These values were chosen after Ning and Haining (2003)'s research found that in over 66% of cases, price differences were less than 1p per litre, and in 82.7% of cases differences were less than 2p. The *neighbourhood* size was set to 1km; *undercutting* value to 0.5p and maximum *priceChange* to 1p.

Figure 5.10 shows that unlike the neighbourhood differences, where clearly a smaller neighbourhood improved the model performance, changing the undercutting price appears to make little difference. The mean price difference and SD remain almost identical throughout all the different model runs. This shows that the price undercutting rule is not having an effect, most likely because the real data is already competitive, i.e. no station is significantly cheaper than the others. Unless there are stations which are significantly cheaper this rule will not have a great impact.

5.5.3 Effect of Maximum Overprice

The maximum *overprice* is the maximum amount by which a station is can be more expensive than its competitors before it has to drop its prices. By operating this behaviour, stations are forced to remain competitive. This is necessary for them to maintain petrol sales. The research of Ning and Haining (2003) found that stations are never more than 1p - 2p different in price. This provides a range of likely *overprice* values.

The results of varying the maximum *overprice* were interesting (Figure 5.11). Altering the amount by 0.5p did not make any difference, but increasing the amount by 1p had the effect of improving the results; this correlates with the observations of Ning and Haining (2003). However, a maximum *overprice* value of 5p produced the best results. These results can be explained by the fact that the greater the maximum *overprice*, the fewer petrol stations are changing price. The higher *overprice* is only affecting the petrol stations with higher prices, such as those found on motorways and in rural areas. These results show that the model is operating in a sensible manner; it is not expected that larger price differences will have a great effect in urban areas. Most petrol

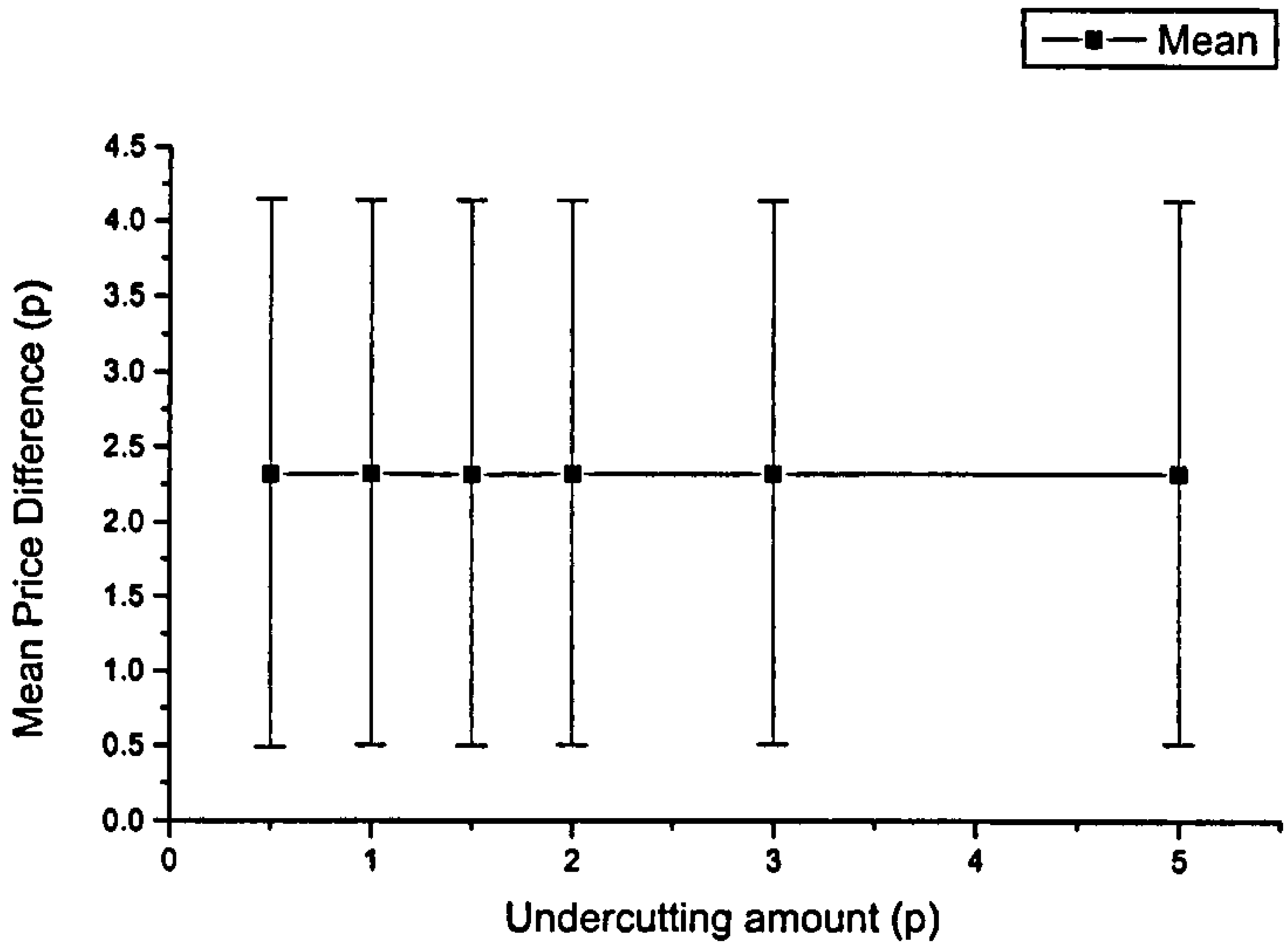


Figure 5.10: Effect of different *undercutting* values on the mean price difference and SD in the price difference (represented by the vertical bars) . The *neighbourhood* size was set to 1km; *undercutting* value to 0.5p and maximum *priceChange* 1p.

stations will, on average, have a maximum *overprice* of 1 or 2p. For further simulations, this value will be kept at 1.5p. This is because there was a significant improvement in the model between 1 - 1.5p. After this, the improvement was marginal.

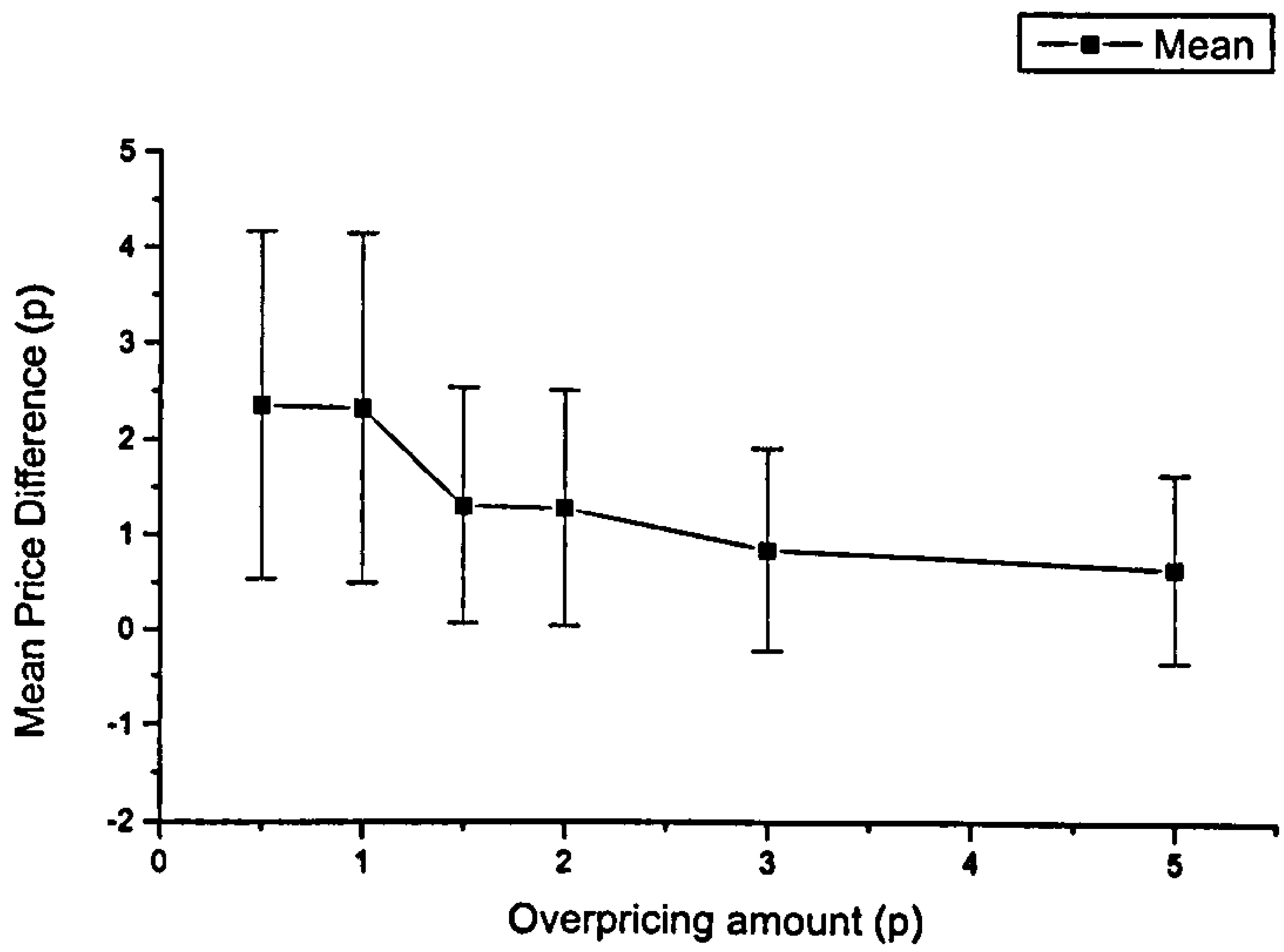


Figure 5.11: Effect of different *overprice* values on the mean price difference and SD in the price difference (represented by vertical bars). The *undercutting* price is set to 0.5p; *neighbourhood* to 1km and maximum *priceChange* 1p.

5.5.4 Effect of Maximum PriceChange

The maximum *priceChange* is the amount in pence that the station may change its price by each day. By setting this constraint, the model cannot create huge price differences. For example, if one station has a starting price of 75p and the others in its neighbourhood have a price of 68p, the station with the price of 75p will not drop to 68p in one day. Instead the price change will be gradual, calming fluctuations. Large price changes are generally not observed in real petrol prices, possibly because of the adverse reaction of consumers. The price difference observed between stations within the literature is 1 - 2p.

Figure 5.12 shows that increasing the maximum *priceChange* increases the mean price difference and standard deviation. The best results came from a *priceChange* set at 0.1p, the worst at 10p. From these results, it can be concluded that prices change slowly and none of the stations are changing their prices more than 0.1p per day. This is not surprising; the changes within the model are continuous and not as abrupt as found within the real data where prices can, in rare circumstances, change by a few pence per day. Additionally, it is possible that there are other factors that could dampen price changes in reality, e.g. aggressive marketing scheme by a competitor, effect of crude oil prices etc.

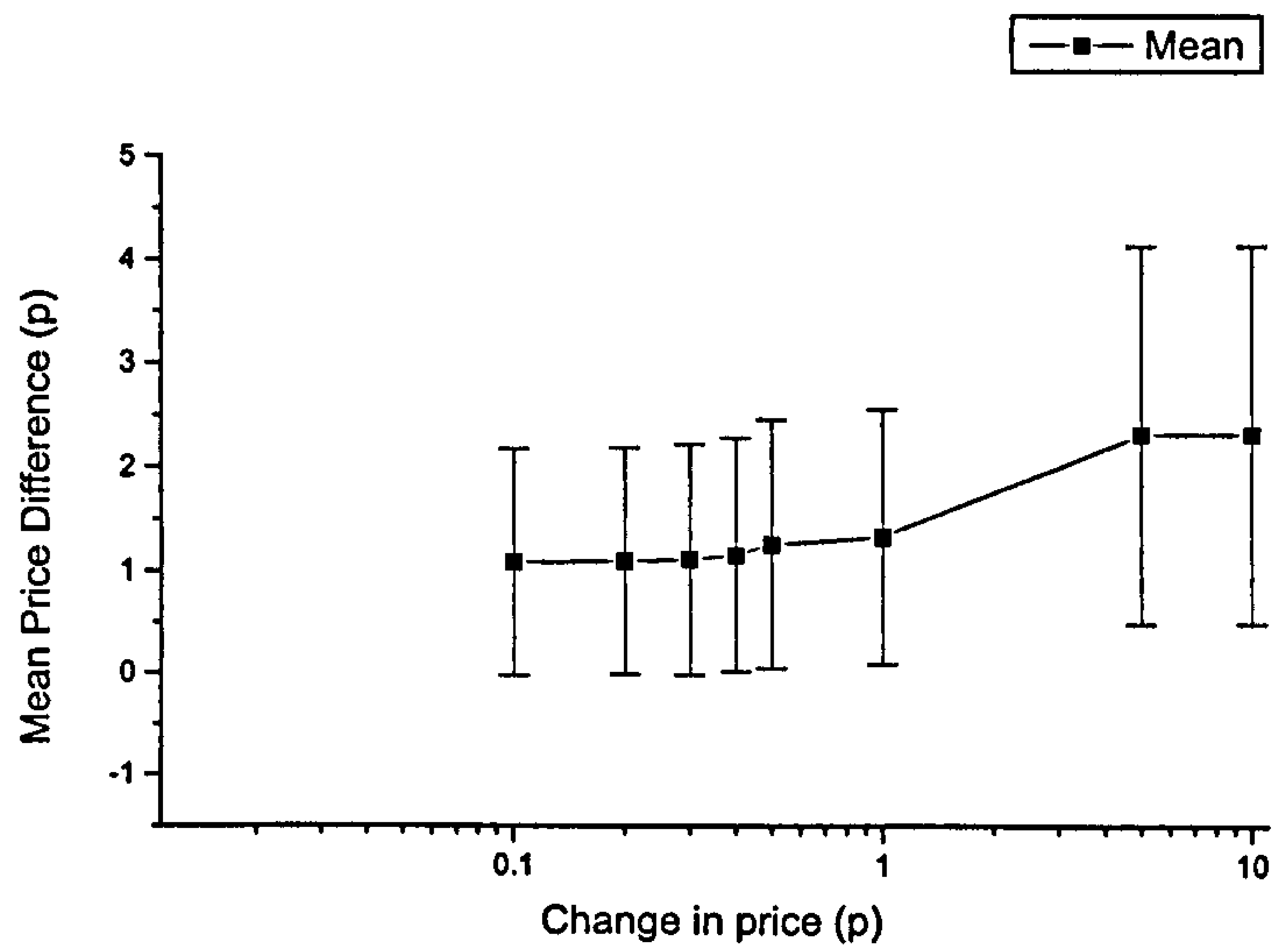


Figure 5.12: Log graph showing the effect of the maximum *priceChange* on the mean price difference and SD in price difference (represented by vertical bars). The *undercutting* parameter is set to 0.5p; maximum *overprice* 1.5p and *neighbourhood* size 1km.

5.6 Comparison of the Agent Model with Real Data

Work has so far concentrated on building up a set of rules that will enable the petrol price market to be modelled. Based on the parameters in Table 5.3 and the results of experimentation, the rules that each agent will operate can be summarised as:

- Assess my neighbours price within a 1km radius.
- If they are more expensive than I am, raise my price up to 0.5p of theirs.
- If I am more than 1.5p expensive than they are, lower my price to match theirs.
- Do not change my price by more than 0.1p per day.

In the analysis of the real data, variations over time and in different geographical regions were identified. Will the agent model reproduce rural-urban variations? How accurate will the results be? The geographical and petrol station type classifications introduced in Chapter 4 will be applied here (see §4.5 for explanations). This will allow a comprehensive examination of the performance of the model; an overall poor performance of the model could be due to an under or over-prediction within one of the classifications.

Parameter	Value
Neighbourhood	1km
Undercutting	0.5p
Overprice	1.5p
PriceChange	0.1p

Table 5.3: Summary of optimal parameter values for the agent model.

5.6.1 Spatial and Temporal Robustness

Figure 5.13 shows that the agent model does not follow the trend of the real data accurately. Over the course of the simulation, the average price remained firmly around 71p ($\pm$0.01p) whilst the average price of the real data decreased to 70.5p before steadily increasing to just over 71p by August 6th. However, the variation of the real data was preserved (as indicated by the SD) throughout the simulation.

Despite these results, it is difficult to assess whether the model is spatially and temporally robust. The model hit equilibrium after only a few days which is clearly not enough time to build a price distribution based on the rule set. The variations within the real data were maintained suggesting that the agent model was not modifying the real prices, merely preserving its variations.

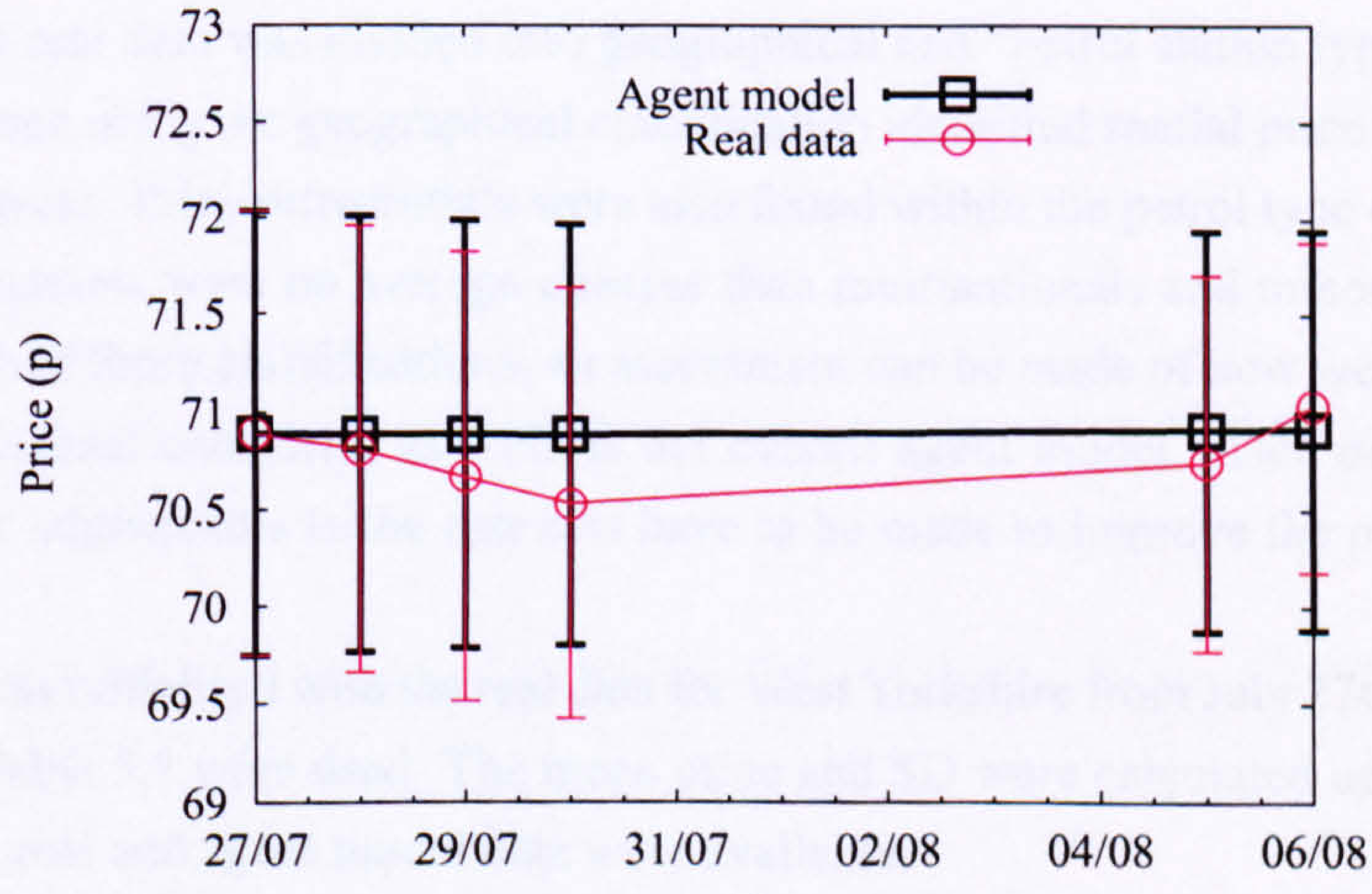


Figure 5.13: Comparison of mean price per day between the agent model and real data over time within West Yorkshire. The SD is indicated by the vertical bars.

5.6.2 Comparison of Performance with Regression Models

The standardised root mean square error (SRMSE) is a standardised method of comparing the agreements between a model run and real data. A low value indicates a good fit, a value of zero would be a perfect match. The SRMSE for the multiple regression, geographically weighted regression (GWR) and agent models were calculated by comparing the predicted value with the real data on day 10 (August 5th). The results are presented in Table 5.4.

Model	SRMSE
Constant Data	0.0131
Multiple Regression	0.0135
GWR	0.0133
Agent	0.0128

Table 5.4: Comparison of the SRMSE for each of the models. For comparative purposes, the SRMSE of the data if nothing had changed (constant data), was also calculated.

Table 5.4 shows that the agent model produced the best SRMSE results and was an improvement on not doing anything (i.e. the real data was kept constant). The GWR slightly outperformed the multiple regression. Unlike multiple regression, GWR takes into account geographical factors weighting neighbouring stations more highly than those located at a distance.

The results reflect the different ways in which the regression and agent models work. The regression model attempts to fit a pattern to all of the data. This would allow the price of an extra station to be predicted. It does not attempt to model the temporal evolution or local competitive behaviour.

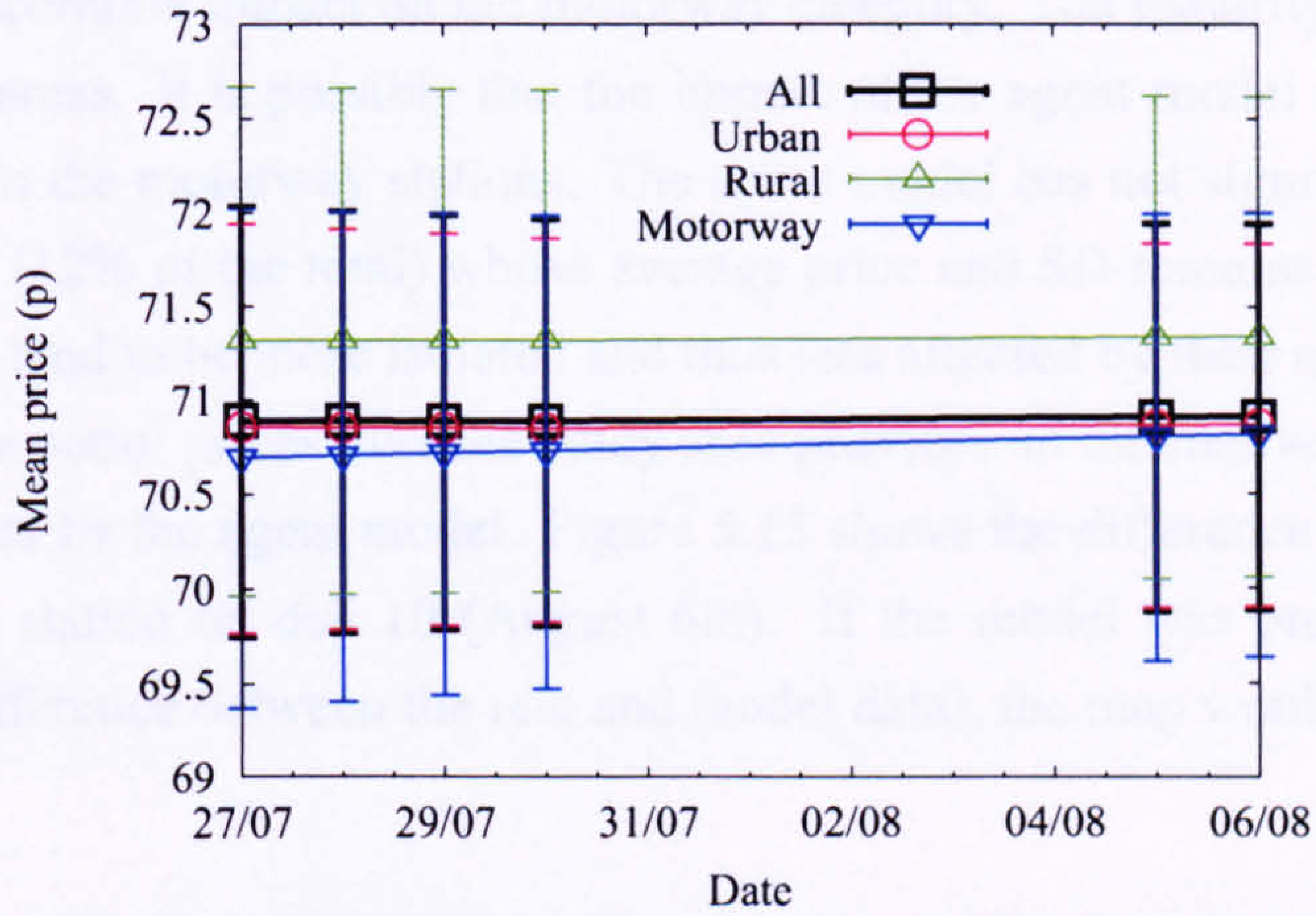
5.7 Analysis with Classifications

In Chapter 4, the real data was divided into geographical and "petrol station type" classifications. Analysis performed using the geographical classification identified spatial price variation between rural and urban areas. Price differentials were also found within the petrol type classifications, for example, supermarkets were on average cheaper than multinationals and minors. By calculating statistics for each of these classifications, an assessment can be made of how well the agent model performs in individual categories as well as the overall agent model. This may suggest where improvements or adjustments in the rule sets have to be made to improve the performance of the model.

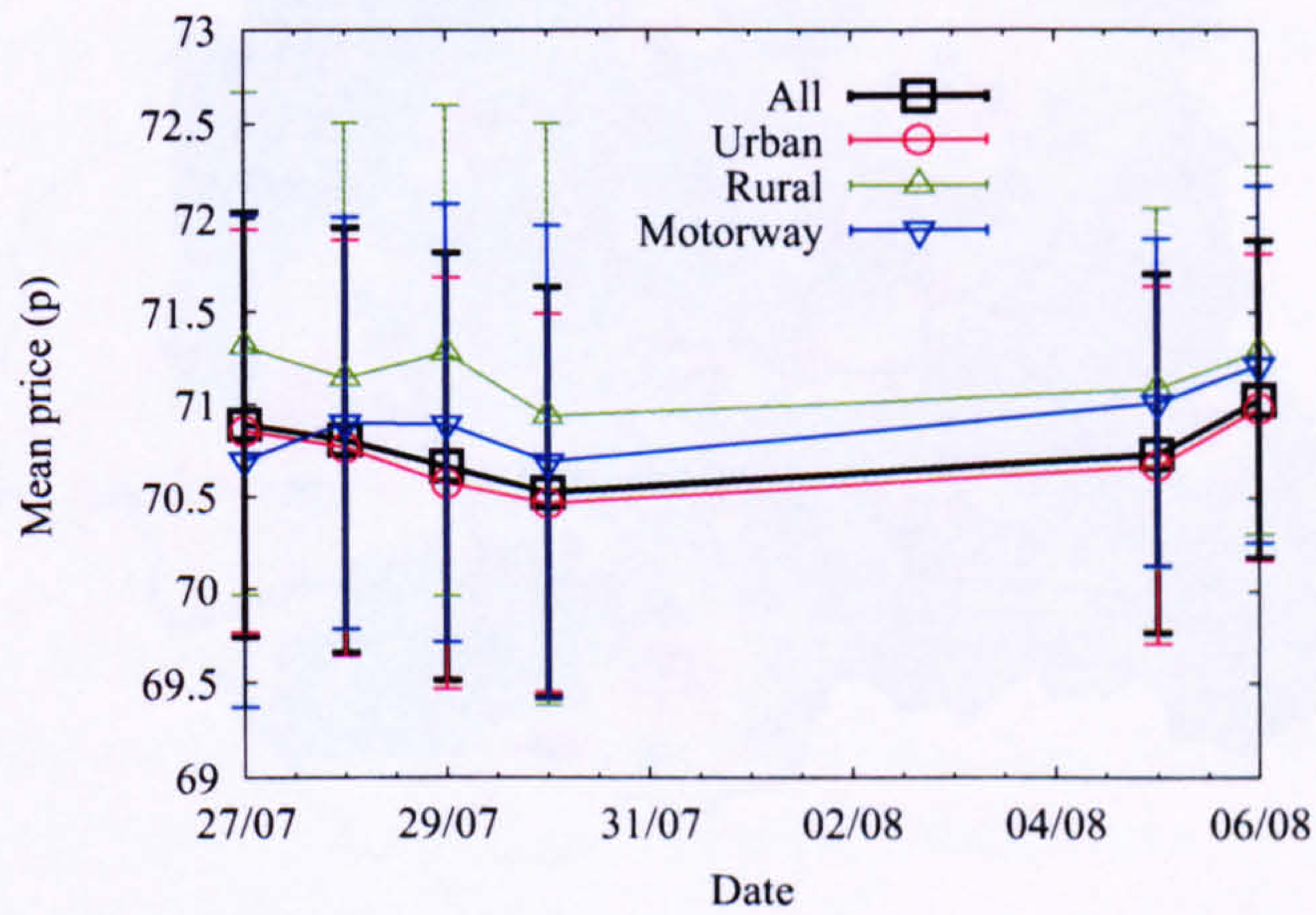
The model was initialised with the real data for West Yorkshire from July 27th. The parameters summarised in Table 5.3 were used. The mean price and SD were calculated up to August 6th on days where both real and agent model data were available.

5.7.1 Geographical Analysis

Figure 5.14 (a) shows quite clearly that the agent model is only modifying the initial data a small amount. Pricing structures and variation apparent within the real data (Figure 5.14 (b)) are preserved. The rural category, for example, sustains the highest average prices, the urban classification the lowest. The SD, represented by vertical bars, in all categories remains almost identical to the initialisation day (July 27th).



(a) West Yorkshire: Model Data



(b) West Yorkshire: Real Data

Figure 5.14: Graph showing a comparison of the trends of the mean price within the different geographical classifications of the agent model (a) and real data (b). Vertical bars show the SD. Results are shown from West Yorkshire for the data initialised on July 27th.

Akin to the real data, the agent model predicts an increase in price over the time difference, however the magnitude of the increase is much smaller in the model (0.05p in comparison to the real increase of 0.4p). The best results come from the urban category. The average price decreases slightly before a gradual increase between July 30th - August 6th, a pattern found within the real data. Results from the motorway and rural areas were both poor. The average price and SD of the rural classification remained constant throughout the simulation. Despite an increase in price (and slight decrease in SD), the prices at the motorway stations were consistently lower than the real data.

These results suggest that the model rules have had the most impact on the motorway and urban stations and the least on the rural category. The urban stations are the largest category (84% of the total). It is reasonable to suppose that any modifications that the agent model has made has fallen on stations within this category. Despite being the smallest (4% of the total), the agent model has had a discernible impact on the motorway category. The majority of these stations are located near urban areas. It is possible that the impact of the agent model on the urban stations has been passed onto the motorway stations. The agent model has not significantly modified any of the rural stations (12% of the total) whose average price and SD remains constant throughout. These rural stations tend to be more isolated and thus less affected by their neighbours.

Interpolating the petrol prices over the study area provides an alternative method of assessing the patterns produced by the agent model. Figure 5.15 shows the difference between the real and model data at each station on day 10 (August 6th). If the model was predicting the real data perfectly (i.e. no difference between the real and model data), the map would be completely light green-blue.

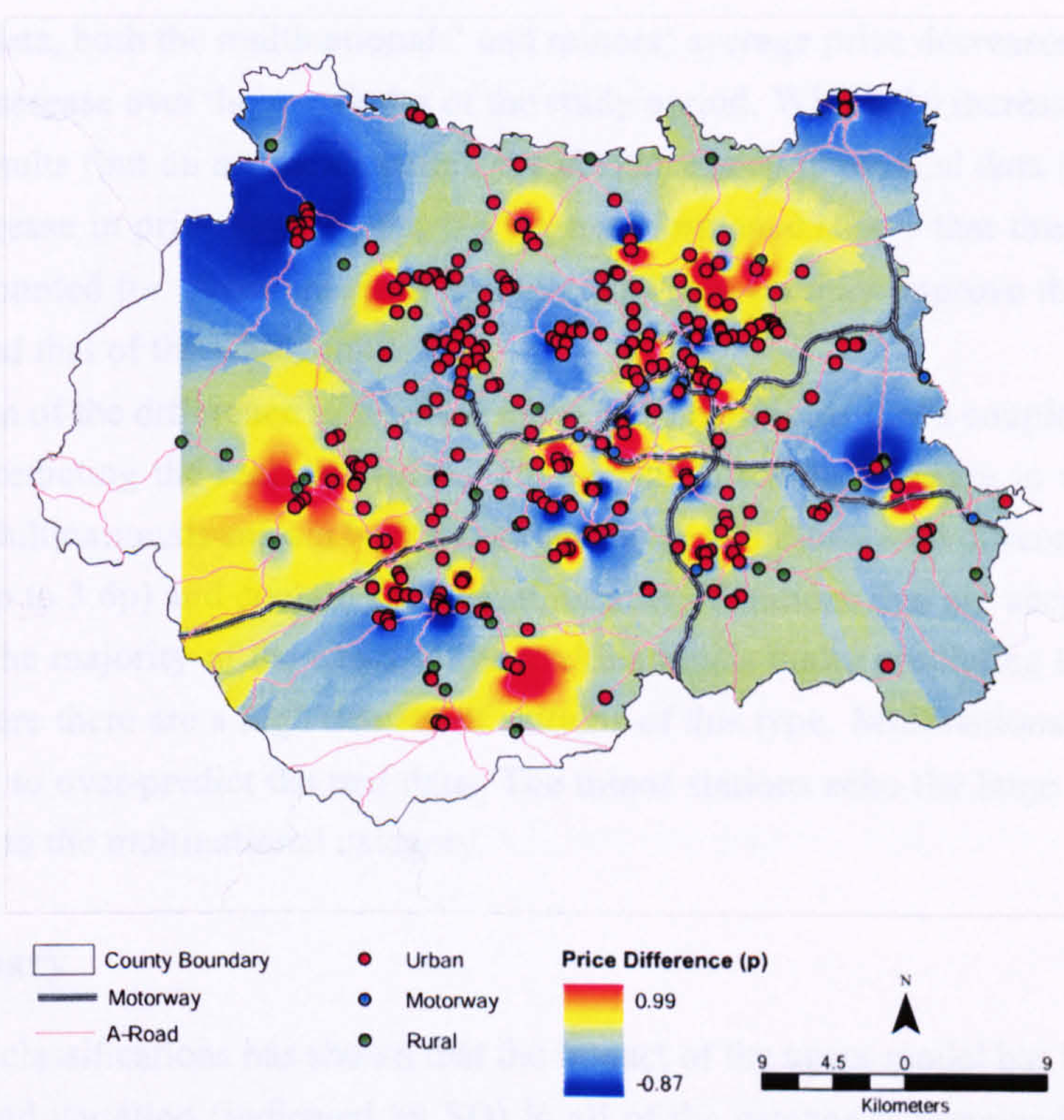


Figure 5.15: Interpolated surface of the price difference between the agent model prediction and the real data on day 10 (August 6th). The geographical classification of each petrol station is overlaid.

Figure 5.15 shows that petrol stations in the urban areas are generally over predicting the real data (indicated by red areas), a pattern echoed by the motorway stations. The majority of rural stations are located in light green-blue areas. A few are situated in areas that are under-predicted, such variations will have been masked when calculating the mean price per day in Figure 5.14. Figure 5.15 shows quite clearly the large amount of spatial price variation that has been preserved by the agent model.

5.7.2 Petrol Station Type Analysis

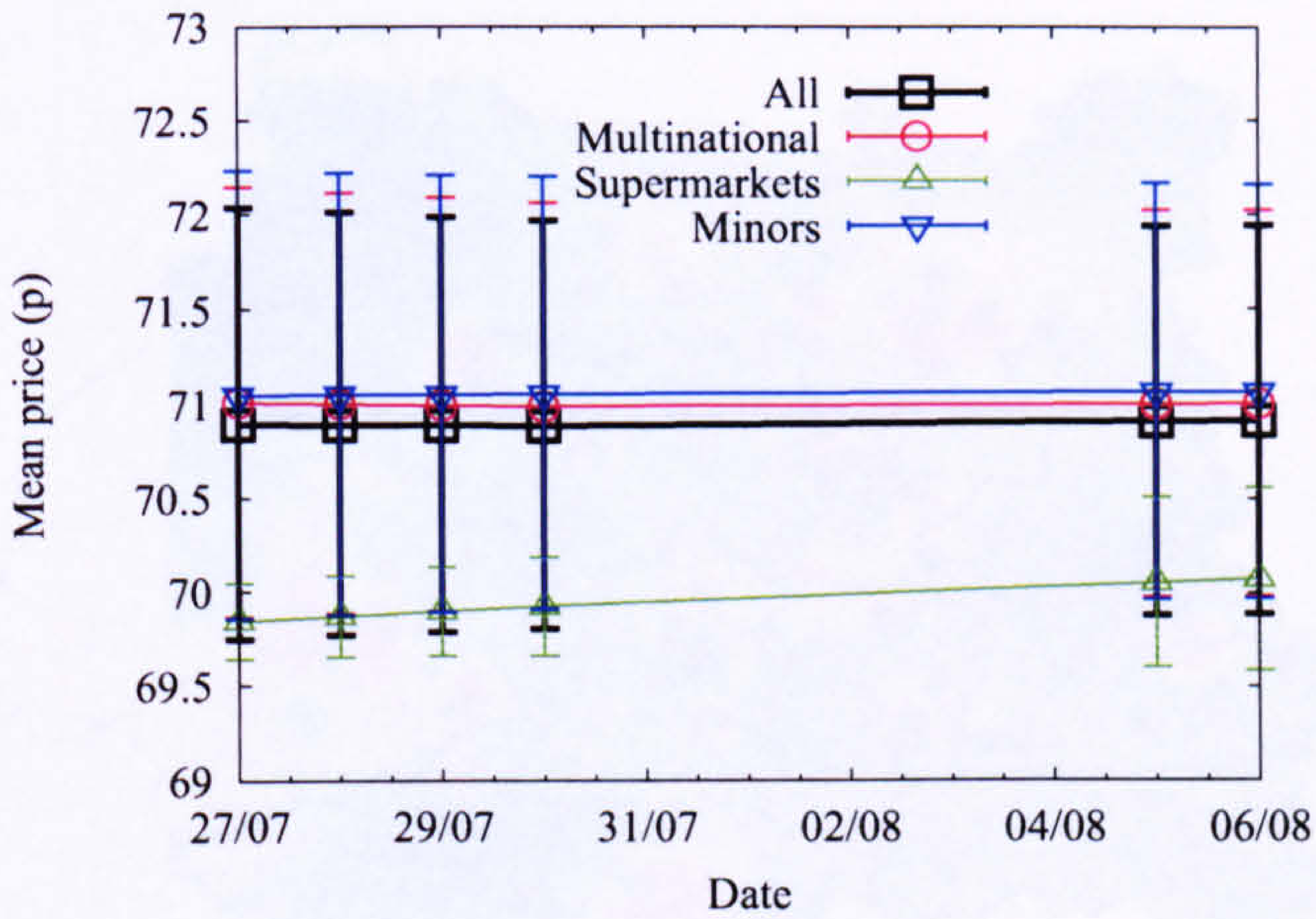
Within the real data (Figure 5.16 (b)), the supermarkets and minors were the lowest and highest priced stations respectively. These pricing patterns have been preserved by the agent model (Figure 5.16 (a)). Whilst the variation within the supermarkets (indicated by the vertical bars) has increased, both the multinationals and minors have experienced a reduction. Together, the multinationals and minors comprise the largest category (94% of the total). The smaller category of supermarkets (6% of the total) appear to be the most altered by the agent model. Both the average price and SD have increased over the time period. This could be a reflection of the large price increase that is evident within the real data (a rise of over 1p). The supermarkets also tend to have much lower prices because of their aggressive pricing strategies (see §4.6.2), but this is not taken into account by the agent model which is purely interested in price rather than profit or market share.

In the real data, both the multinationals' and minors' average price decreases immediately before a gradual increase over the remainder of the study period. Whilst the increase is distinguished in the model results (but on a smaller scale), the decrease seen in the real data is not. The supermarkets do increase in price slightly, but the poor performance shows that this behaviour is not being fully accounted for. Assignment of an individual rule set may improve the performance of this category and that of the whole model.

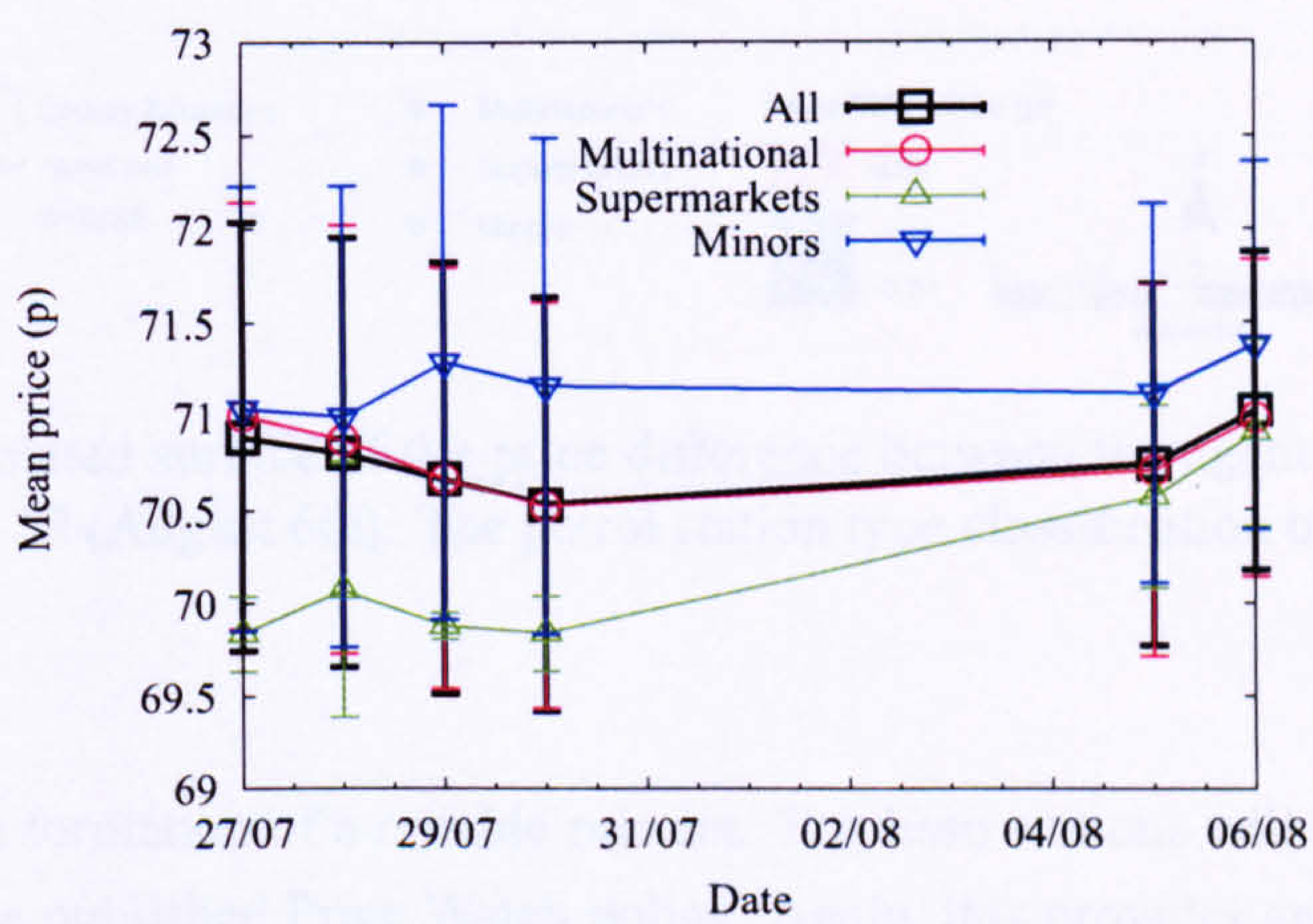
Interpolation of the difference in prices (Figure 5.17) highlights that a couple of the supermarkets are over predicting the real data price. This accounts for the increase in variation over the study period. Multinationals sustain both the largest price variations with differences covering the full range (-1.9p to 3.6p) and contain the largest number of stations that are accurately predicting the real data. The majority of the cases of the multinationals under predicting the real data price are in areas where there are a high density of stations of this type. Multinational stations in more rural areas tend to over-predict the real data. The minor stations echo the large variation in price difference seen in the multinational category.

5.7.3 Summary

Analysis of the classifications has shown that the impact of the agent model has been limited. The average price and variation (indicated by SD) in all of the categories remained almost constant, the largest impact was evident within the supermarkets class. This trend was found in §5.6.1 with an examination of the spatial and temporal robustness of the model performance. The model does not significantly alter the data from its initial values and it is this rather than the agent rules which lead to the spatial variations in the end results. The real data shows temporal variations in price which are not captured by the agent model.



(a) West Yorkshire: Model Data



(b) West Yorkshire: Real Data

Figure 5.16: Graph showing a comparison of the trends of the mean price within the different petrol station classifications of the agent model (a) and real data (b). Vertical bars show the SD.

5.8 Case-Studies

To date, the agent model has only operated with an identical rule set applied to all the agents. The purpose of this has been to test the stability of the model and to build an operational rule set based on reality. However, this does not fully exploit one aspect of the functionality of a MAS, i.e. the ability to specify individual rule sets for different "categories". Use of a separate assignment of rules for different types of stations may enable the system to better model reality by accounting for the differences in behaviour. This could improve the performance of categories, for example supermarkets, thereby improving the global model performance.

The following sections examine the results of application of different rule sets to the supermarket category and Esso stations. From the real data analysis performed in Chapter 4 and evidence in the literature, supermarkets were identified as being competitive and tightly priced. These char-

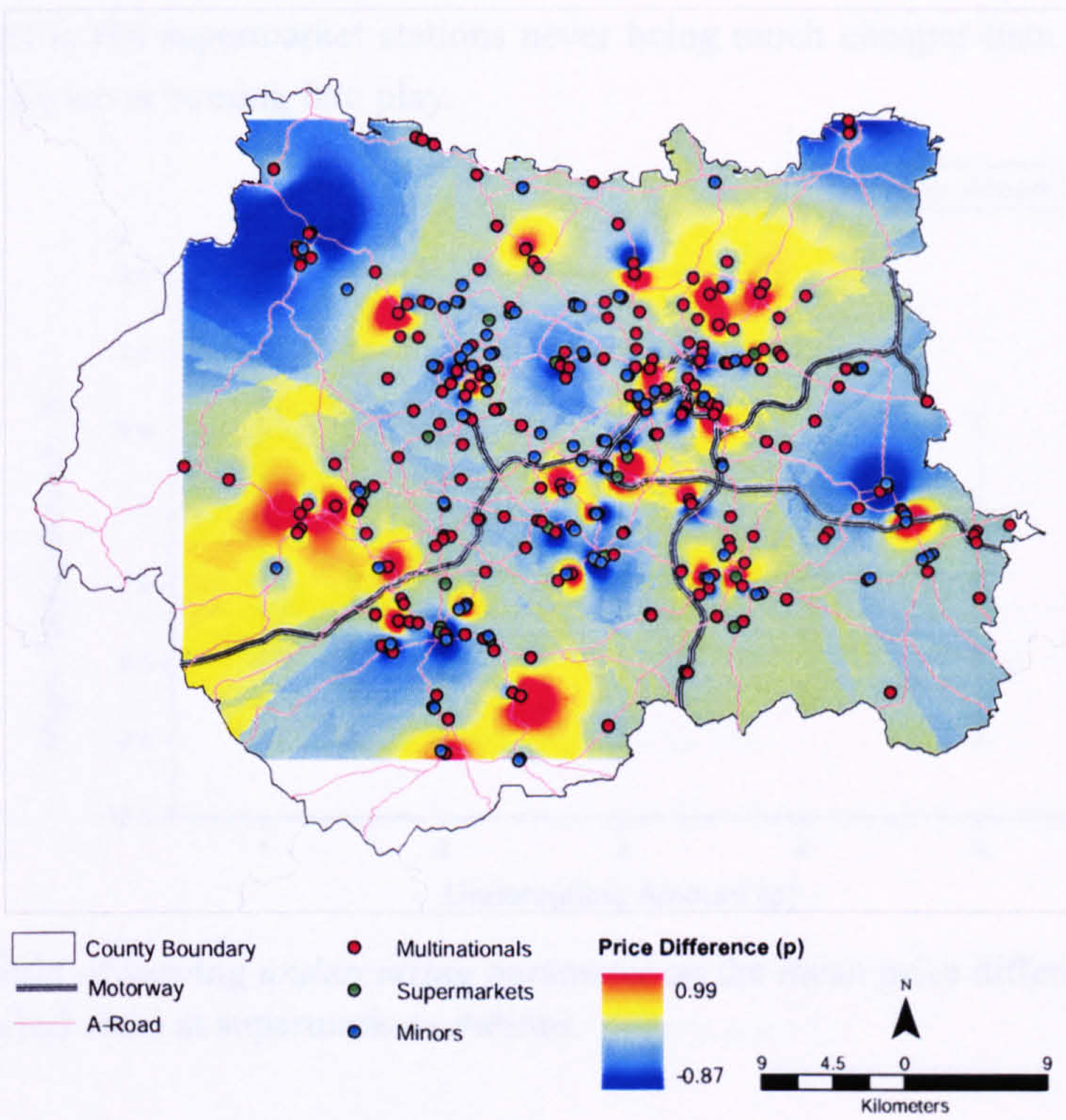


Figure 5.17: Interpolated surface of the price difference between the agent model prediction and the real data on day 10 (August 6th). The petrol station type classification of each petrol station is overlaid.

acteristics allow the formation of a reliable rule set. The Esso stations will be assigned a separate rule set based on the published Price Watch policy. Again, this provides an opportunity to assign a reliable rule set and assess the impacts on the agent model performance.

5.8.1 Supermarket Competitiveness

Supermarkets are generally more aggressive on price than other petrol stations. This was illustrated in §4.6.2, where it was found that supermarkets are the most cheaply priced and competitive (as indicated by their small SD). This trend was also recently seen in May 2002 when both Morrisons and Asda lowered their prices after earlier rises due to conflicts in the Middle East. From this example of behaviour, it can be assumed that these stations will not let themselves be overpriced. To investigate the impact of this behaviour, the *undercutting* parameter will be set to 2p on supermarkets and 1p on all others. The *neighbourhood* is set to 1km and *overprice* parameter to 1.5p. All the other stations were run on the optimal parameters outlined in Table 5.3.

Figure 5.18 shows that the mean price difference decreases with an increase in the undercutting amount. However, once the *undercutting* parameter reaches 1.5p there is little further change in the mean or SD. Since supermarkets tend to operate an aggressive price policy and undercut other stations in the area, they might be expected to have a price difference of a penny or more. Making the maximum *undercutting* parameter too small restricts this aggressive price behaviour and leads to the model over predicting the supermarket prices. Increasing the *undercutting* parameter results in the supermarket stations never being much cheaper than the other stations leading to this rule never coming into play.

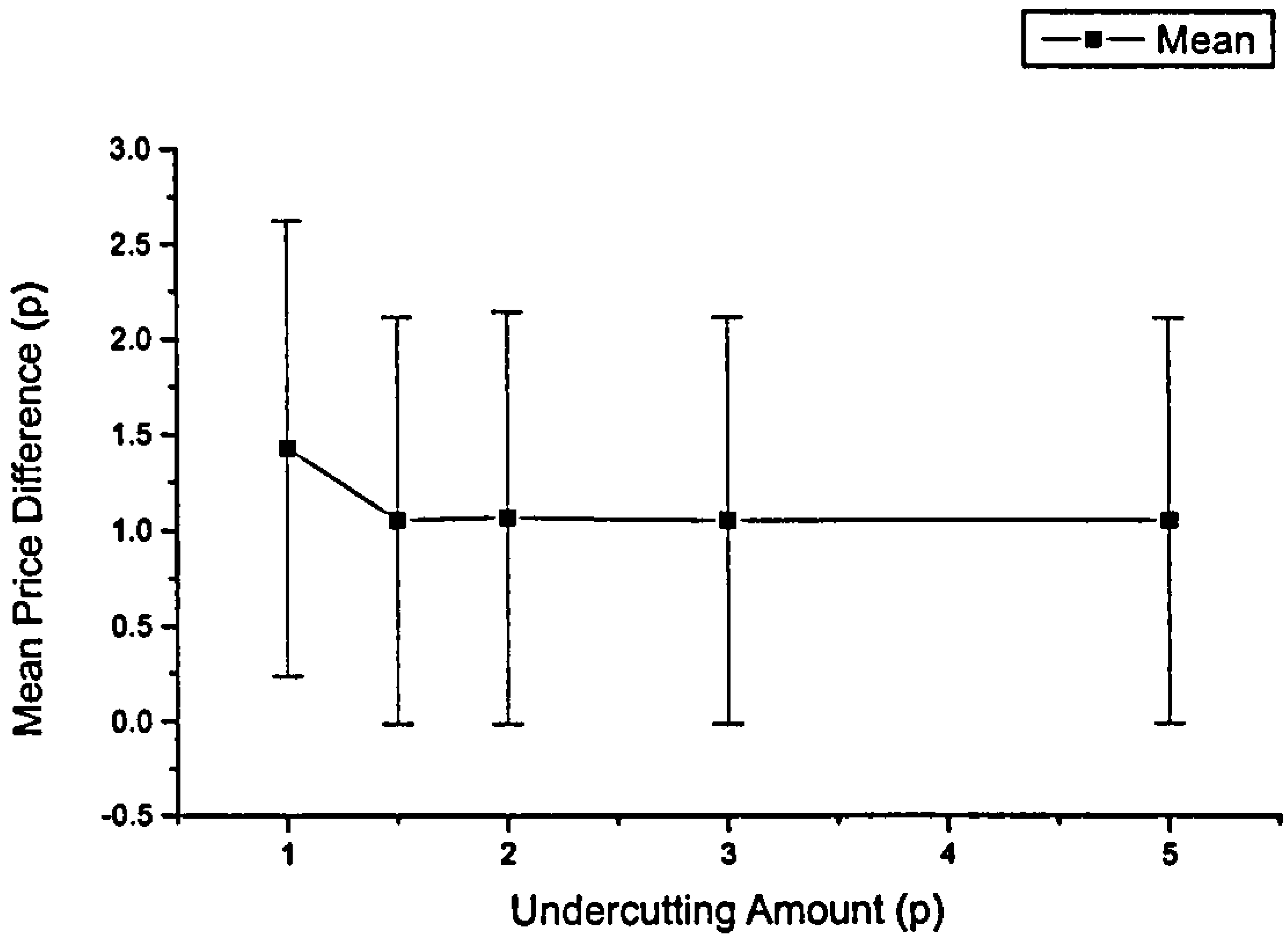


Figure 5.18: Effect of varying *undercutting* parameter on the mean price difference and SD (represented by vertical bars) at supermarkets stations.

5.8.2 Esso Price Watch: Tiger or Pussy Cat?

The Esso Price Watch pledge states that they will match any price within the surrounding 3 miles. To attempt to test whether this is the case, the Esso brand was selected and given its own rules. The *neighbourhood* size varied by 1km, 3km, 5km[3] and 10km and the maximum *overprice* was set to 0p. This option enables the user to specify a threshold difference at which the price will change in line with other prices; i.e. if it is set to 0p, Esso will drop its prices to the lowest around it. This is in line with the Price Watch policy. The *undercutting* and *priceChange* parameters for the Esso stations were given the optimal values of 0.5p and 0.1p respectively. All the other stations were run on the optimal parameters outlined in Table 5.3.

Figure 5.19 shows that changing the neighbourhood size of the Esso stations does not make much difference as a whole to the mean and SD of the error of the data set. The error is highest for the 2km and 3km neighbourhoods, with the values for 1km, 5km and 10km almost identical. Comparing the SRMSE from the model with all stations using the same rules (0.91p) and with Esso stations operating different Price Watch rules on a 5km neighbourhood (1.00p) shows no clear sign of improvement with the inclusion of the more complicated Esso rules. From these results, it cannot really be determined whether the Price Watch is working (i.e. including it in the agent model does not significantly improve the agreement with the real data). The parameter changed (maximum *overprice*) is only affecting the Esso stations and these may only be a playing a small part in the system as a whole.

The mean price difference and SD were calculated for all the stations with and without the Esso rules turned on. The rules showed that using the Esso Price Watch did not improve the

[3] 5km is equivalent to 3 miles.

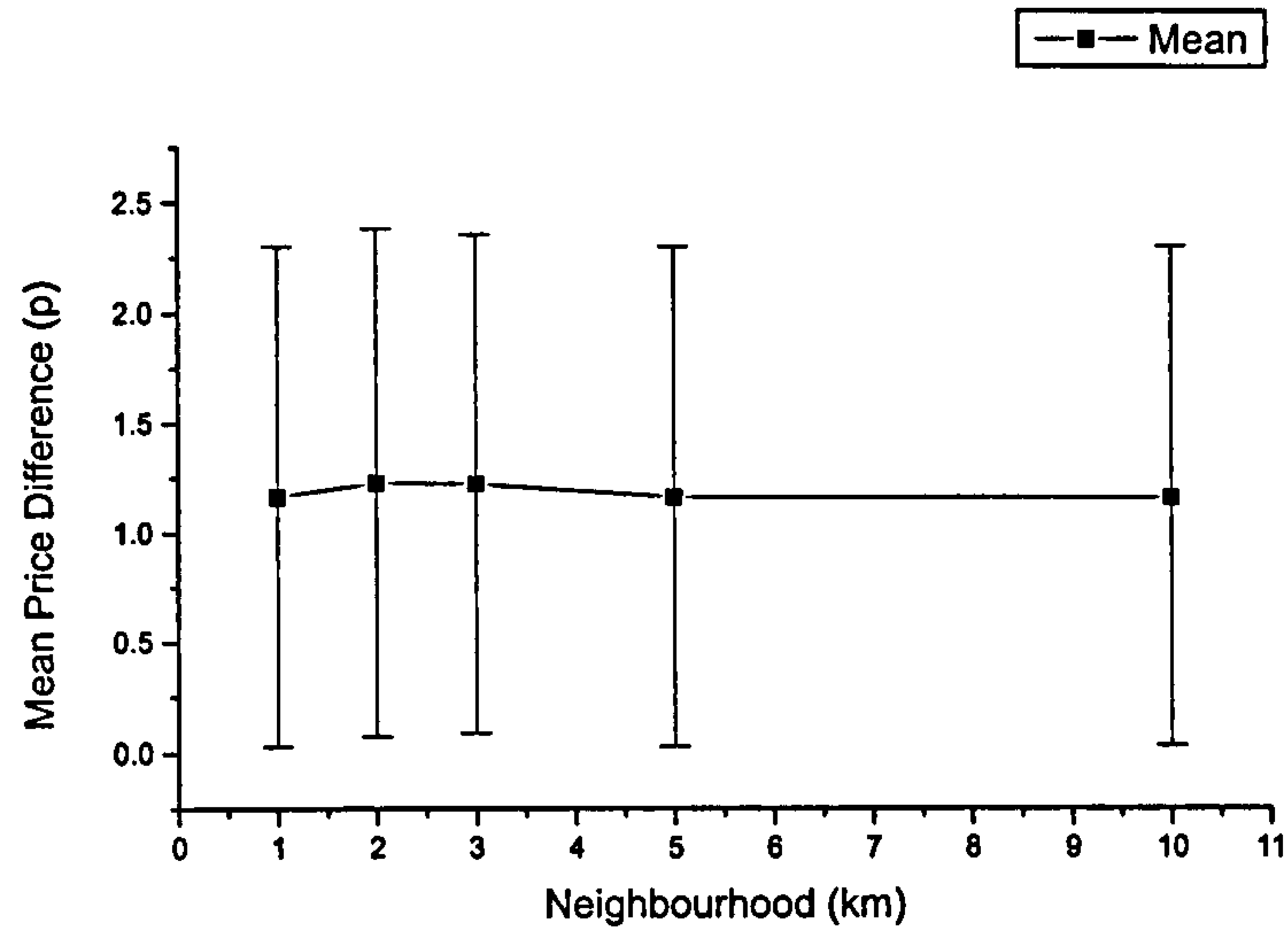


Figure 5.19: Effect of varying *neighbourhood* size on the mean price difference and SD at Esso stations.

performance of the overall results. The results are in Appendix D.

5.9 Recreation of Spatial Patterns

Experimentation undertaken has led to the conclusion that the agent model is merely preserving variations within the data and not implementing the rule sets. This is partly backed up by the short time that the agent model took to reach equilibrium (between 2-3 days). By initialising the model with constant prices, the ability to generate rather than preserve variations within the data is tested.

The agent model was run with all the petrol stations assigned the same initial prices. The average price on July 27th (71p) was chosen and the model run to equilibrium. Figure 5.20 shows that no changes were observed. The agent model has assumed that the situation was stable and hit equilibrium immediately.

5.10 Conclusion

This chapter has focused on the construction of a MAS to model the behaviour (patterns and trends) of the petrol price market. This model incorporated no consumer behaviour, instead only operating by stations comparing their price to others in the neighbourhood. "Petrol agents" have been designed to share information for mutual benefit and react in a competitive manner to the actions of the stations around them. A rule set based on industrial information was built and applied to all of the agents (petrol stations).

Comparison of the results spatially, temporally and within different classifications was undertaken. The results showed that the model was stable; there was no irrational behaviour. It was clear through the analysis performed on the results that instead of modifying the real data, the model was largely preserving its variations. Analysis using the classifications showed that the model was having an impact on the categories containing the largest number of stations. Attempts were made

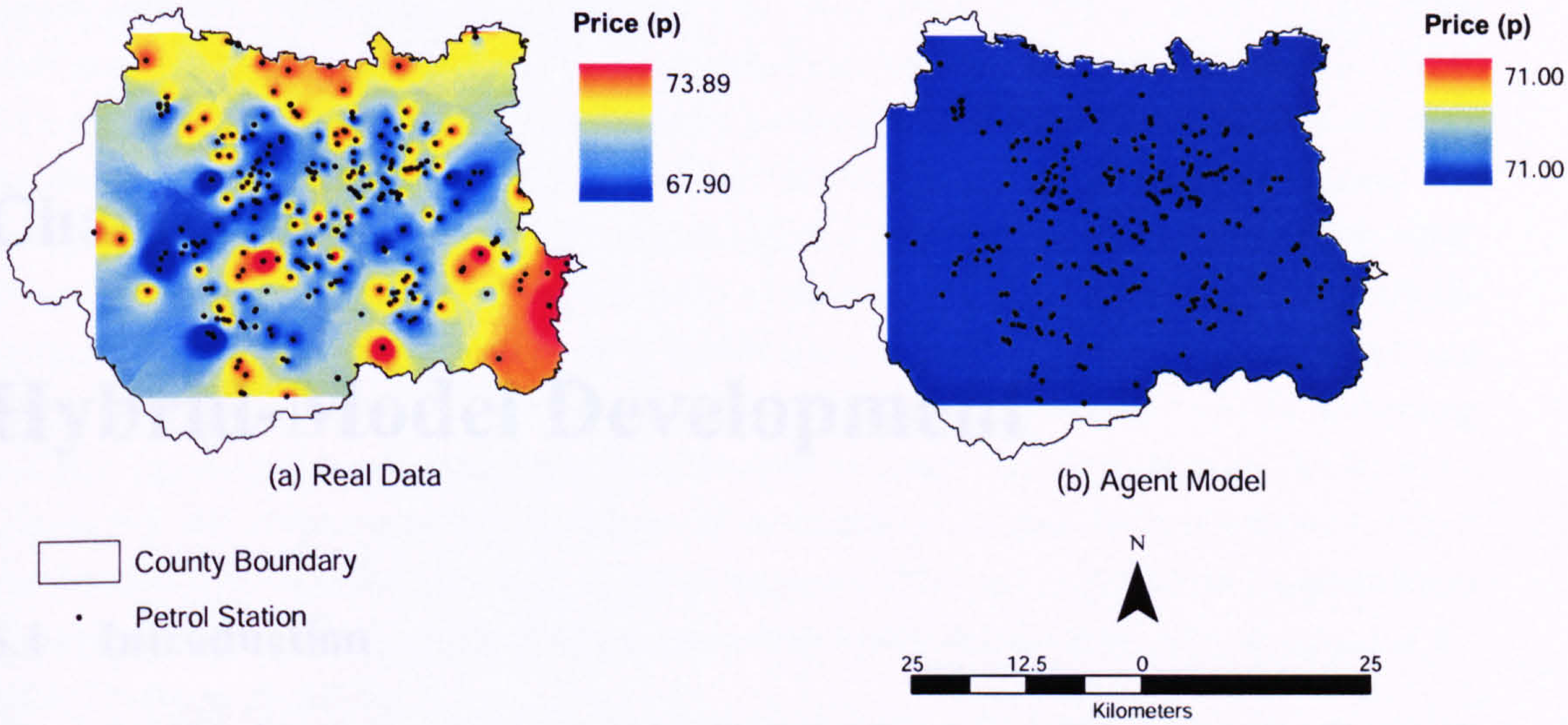


Figure 5.20: Comparison of the price distributions of (a) the real data and (b) agent model 10 days initialised with all stations at 71p.

to improve the results by assigning separate rule sets to the supermarket and Esso stations. No improvement was noted.

The model presented within this was chapter is simply an agent version of the non-consumer models reviewed in Chapter 2. The goal of the system was to simulate real petrol stations by being competitive, i.e. assessing the price of the nearest neighbours and undercutting them. This mode of application has not been entirely successful. However, the use of a MAS has brought valuable functionality that is not available using traditional empirical techniques. Regression models do not attempt to model the temporal evolution or local competitive behaviour. Neither can they model the impact of assigning individual rule sets to stations at a local or regional level.

An avenue of research that could be investigated is the implementation of alternative modes of behaviour. This approach would make use of the agent methodology, for example, the use of different strategies at individual petrol stations depending on what is occurring in the immediate environment. These strategies need to be driven towards a common goal. The assumption can be made that the petroleum industry, akin to any business, is driven by profit; one potential goal of each petrol station could be maximisation of profit. This does assume that profit maximisation and not other strategies, for example, attainment of market share is dominant. However, a model driven by profit maximisation would be more realistic than a model just based on competition.

The agent model does not account for the behaviour of customers. Consumers are obviously vital for any business. Petrol stations must attract customers in order to sell petrol. Pricing is an important part of a consumer's decision on where to buy their fuel. Any realistic model of petrol pricing must account for this. A technique that would be suited to calculating the flows of people to a service is spatial interaction modelling. The potential of this technique will be explored in Chapter 6.

Chapter 6

Hybrid-Model Development

6.1 Introduction

The work within this chapter aims to extend and improve the agent model presented in Chapter 5. One of the main criticisms levelled at the agent model was that it did not take the influence of consumers into account. In this respect, it was very similar to the business centred regression models presented in Chapter 3. Each station operated a strategy that enabled it to be more competitive than its neighbours. This was felt to be too simple and the incorporation of new strategies were needed.

Spatial interaction modelling was suggested as a possible method to enable consumers to be modelled and thereby allow the inclusion of more sophisticated strategies. Use of a MAS for modelling consumers was rejected due to computational requirements. The typical ward in West Yorkshire has a population of 20,000 and so modelling the consumers at the level of the individual rather than at the ward level would, at a rough guess, take at least 10^4 times more storage and computational time than the approach described here. This is not currently realistic. Additionally, the data requirements for constructing a knowledge set for the consumers would be immense. For example, detailed information would be required on buying patterns such as frequency of visit, amount purchased and impact of marketing strategies etc.

This chapter presents a brief review of the main developments and approaches within spatial interaction modelling. This is followed by a discussion centred on which is the most suitable model for this application. The specifics of the spatial interaction model and strategies are presented along with an alternative method of distributing the sales. Testing and evaluation of this new hybrid model is undertaken in Chapter 7 and Chapter 8.

6.2 Spatial Interaction Modelling

Spatial interaction models are used to facilitate the explanation and prediction of human and economic interactions over geographical space. These interactions can be, for example, the movement of goods, information or people. These models have been used extensively within geographical applications (see Birkin *et al.*, 2002, for a discussion).

The first spatial interaction models can be grouped under the generic heading of gravity models. These models were characterised by attempting to represent the behaviour of demand or supply segments, rather than that of individual firms (Roy and Thill, 2004). The 1960s and 1970s were marked by a huge outpouring of both theoretical and empirical work (see Fischer *et al.*, 2003, for a review). The gravity analogy was replaced by the general concepts of entropy (a statistical framework borrowed from physics) and work by Stouffer, Isard and Wilson filled regional science journals with their theoretical and methodological contributions (Isard, 1960; Stouffer, 1960; Wilson, 1971, 1974, 1981). With the exception of the work by Fotheringham (1983b), very little further theoretical progress was made in spatial interaction modelling between the 1970's and 1990's. However, in recent years with the rise of technological innovations, powerful computing and data rich environments, researchers have returned to spatial interaction theory. The literature reflects this with articles published on the use of evolutionary computation to breed new forms of spatial interaction models (Openshaw, 1988; Turton *et al.*, 1997) and network-based approaches to spatial interaction leading to neural spatial interaction models (Openshaw, 1993; Reggiani *et al.*, 2001; Fischer *et al.*, 2003).

The following sections will present a brief review on spatial interaction modelling with a view to identifying a suitable model for the research requirements. This is not meant to be a comprehensive review, rather a concise overview. For a more detailed review of the evolution of this field, the reader is directed to Sen and Smith (1995) and Roy and Thill (2004).

6.3 Early Development,Wilson and Family

One of the main theories that dominated studies in the latter half of the nineteenth century was Newton's Theory of Universal Gravitation. Many of the ideas put forward to explain patterns in human activity between separate entities drew inspiration from Newton's theories. The first applications of Newton's theory arose in the mid-1850s. These applications saw gravitational force replaced with intensity of interaction between two areas (expressed as the number of trips) with the purpose of attempting to explain migration choices (see Roy and Thill, 2004, for discussion).

It was not until the 1960's with the work of Isard (1960) that gravity models were recommended as a tool within regional science. This was a reaction to the growth of large regional shopping centres where competition between spatial entities needed to be accounted for. Huff (1963) reacted to this need and developed a probabilistic retail model which evaluated the choices of alternative shopping centres by sets of shoppers. This model moved away from the rigidity of the Newtonian model by allowing the gravity exponent to be calibrated and treating time rather than distance as the separation component.

These early attempts at spatial interaction modelling coincided with the early years of the quantitative revolution. This period saw the rigorous mathematical formalisation of spatial interaction models culminating in the development of the family of models proposed by Alan Wilson.

The work of Wilson (1967, 1970, 1971) was a major step in the development of the spatial interaction model. His "family" is designed around three basic assumptions drawn from Newtonian laws of gravitational attraction. These state that the magnitude of any flow between an origin and destination is:

1. directly proportional to the trip-generation capacity of the origin O_i.

2. directly proportional to the trip-attraction capacity of the destination D_j.

3. inversely proportional to the distance or travel cost C_{ij} between the two zones, also referred to as the distance decay effect.

From these assumptions, the basic unconstrained model was developed as

$$T_{ij} = kO_iD_jC_{ij}^{\beta};\ k = \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij}^{pred} / \sum_{i=1}^{n}\sum_{j=1}^{m} T_{ij}^{obs}, \tag{6.1}$$

where

- T_{ij} is the flow or trip intensity.
- $i(i = 1,n)$ is an origin zone.
- $j(j = 1,m)$ is a destination zone.
- β is an exponent which controls the rate at which the flow decreases with increased distance or travel cost.
- k is an optional scaling set to constant to ensure that the total number of predicted interactions is equal to the total number of observed flows.

However, in many applications, the need for consistency between observed and predicted trips became apparent. For example, work by Huff (1970) and Wilson (1974) identified the need to equate either origin and destination totals (or both) across observed and predicted trip volumes. Consequently, this led to the family of spatial interaction models being extended to include some form of constraint on trip totals at the origin, destination or both. These new models can be written as

$$T_{ij} = A_iO_iD_jC_{ij}^{-\beta};\ A_i = \left[\sum_{j=1}^{n} D_jC_{ij}^{-\beta}\right]^{-1}, \tag{6.2}$$

$$T_{ij} = B_jO_iD_jC_{ij}^{-\beta};\ B_j = \left[\sum_{i=1}^{m} O_jC_{ij}^{-\beta}\right]^{-1} \tag{6.3}$$

and

$$T_{ij} = A_iO_iB_jD_jC_{ij}^{-\beta};\ A_i = \left[\sum_{j=1}^{n} B_jD_jC_{ij}^{-\beta}\right]^{-1}\ B_j = \left[\sum_{i=1}^{m} A_iO_iC_{ij}^{-\beta}\right]^{-1}. \tag{6.4}$$

Equation (6.2) represents the model that has come to dominate the literature, the origin- or production-constrained model (Fotheringham *et al.*, 2001). This model can estimate the ability of destinations to attract a share of known total flows from an origin. The constraint A_i serves to equate the origin totals across both observed and predicted trip matrices. The destination or attraction constrained model is shown in Equation (6.3). The constraint B_j ensures that the actual and predicted destination totals are equal, such as in a residential location model where the number of workers living in each origin zone is fixed. The doubly-constrained model, Equation (6.4), incorporates the constraints of both and is most widely used in transportation studies. After the model is chosen for a particular application, it must then be calibrated on a data set to determine the appropriate value of the distance decay parameter β.

6.4 Further Developments

A great deal of work has been undertaken to improve the basic specifications of gravity models either by modifying the mathematical structure or by the addition of further variables. This section of work will concentrate on detailing entropy maximisation, intervening opportunities and the competing destinations models. It is acknowledged that this omits work carried out using statistical models, for example, log linear (Willekins, 1983), Poisson (Flowerdew, 1991; Yano, 1993) and multinomial logit (Ben-Akiva, 1985) modelling approaches. The reader is referred to Sen and Smith (1995) for more detailed specifics.

6.4.1 Entropy Maximisation

One of the initial modifications to the early spatial interaction models was the incorporation of the statistical concept of entropy maximisation (Wilson, 1971). This analogy with statistical models answered previous criticisms that individual behaviour was not accounted for (Huff, 1961), and the models had no sound theoretical basis (Foot, 1981). This was achieved by making it possible to calculate the most probable interaction matrix that will arise from a maximisation of individual trips without having to explain the behaviour of individuals at the micro level. The results of applying entropy maximisation to the derivation of spatial interaction models was to change the form of the cost function from a negative power in the travel cost within the system. An origin-constrained entropy maximising spatial interaction model can therefore be written as

$$T_{ij} = A_i O_i W_j \exp(-\beta C_{ij}); \; A_i = \left[\sum_{j=1}^{n} W_j \exp(-\beta C_{ij}) \right]^{-1}. \qquad (6.5)$$

Each member of the spatial interaction family can be modified in this way to reflect entropy maximisation.

6.4.2 Intervening Opportunities

First developed by Stouffer (1940) and refined by Schneider (1960) for modelling journey-to-work in the Chicago Area Transportation Study, the intervening opportunities model was an early alternative to gravity models. This model formalisation took into account the existence of multiple origins or destinations. For each origin zone i, the possible destination zones j as j_{im} are ranked in order m of their proximity to zone i, where the closest zone to i is denoted by $m = 1$. T_{ijm} is the number of trips from zone i stopping at the mth ordered destination j out from i. O_i is the number of commuters available at zone i. Approximating the difference equations by a differential equation, T_{ijm} the intervening opportunities model can be given as

$$T_{ijm} = k_i O_i [\exp(-(pW_{ijm} - 1)) - \exp(-pW_{ijm})], \qquad (6.6)$$

in which W_{ijm} is the total number of jobs passed up to and including j_{im}. However, this model has two key limitations: (i) the inability to satisfy destination constraints and (ii) the fact that destinations can be distributed over 360 degrees surrounding the origin zone i, implying that the opportunities in successive destinations at increasing distance out from i may not be truly intervening (Guldmann, 1999)[1].

6.4.3 Competing Destinations

A major contributor to the development of spatial interaction modelling is the notion of the choice set formation (Thill, 1992)[2]. Fotheringham (1983a,b, 1984, 1985, 1986) and Fotheringham *et al.* (2001) successfully used the notion of choice sets within the development of the competing destinations model. An origin-constrained spatial interaction model with a competing destinations component can be written as:

$$T_{ij} = A_i O_i D_j Q_{ij}^{\beta_2} \exp(-\beta_1 C_{ij}) \qquad A_i = \left[\sum_{j=1}^{n} Q_{ij}^{\beta_2} \exp(-\beta_1 C_{ij})\right]^{-1}$$

$$Q_{ij} = \sum_{k=1, k \neq j}^{n} D_k / C_{jk} \tag{6.7}$$

where Q_{ij} is an extra variable for accessibility of a destination as perceived by individuals, and it reflects the competition between destinations for interactions. D_k is the attractiveness of other destinations in the system and β_1 and β_2 are parameters which are determined through calibration. The parameter β_2 has special theoretical significance in that when β_2 is greater than 0.0, competitive effects dominate. When β_2 is less than 0.0, agglomerative effects dominate and when β_2 equals 0.0, the competing destinations model becomes a conventional entropy maximising origin-constrained equation (Fotheringham, 1984).

6.4.4 Artificial Intelligence Approaches

With the increase in availability of interaction data and powerful computing resources, it has been possible to apply inductive AI tools (in particular evolutionary computing and artificial life) to spatial interaction modelling. These techniques have arrived at a convenient time with the difficulties associated with further theoretical model development and traditional deductive approaches limiting the advances made to spatial interaction modelling.

Diplock (1996) built on work by Openshaw (1994) demonstrating that genetic algorithms (GA) can provide very good solutions to the spatial interaction model calibration problem, especially when calibrating complex multi-parameter models. Diplock and Openshaw (1996) also demonstrated that it is possible to breed spatial interaction models that outperform conventional models using GAs.

Artificial neural networks (ANNs) have been applied extensively to the problem of modelling flows. They have an advantage over conventional spatial interaction methods because they make no assumptions on the form or distributional properties of interaction data and predictors. This

[1] This can be overcome by introducing an extra route index along which opportunities "directly" intervene (Roy, 1999).

[2] This represents the decision-making processes within the model specification. Conventional models assume that the individual considers every possible opportunity when deciding where to travel.

allows spatial interactions to be modelled even when the only data available is noisy or statistically ill-conditioned (Roy and Thill, 2004). Empirical evidence reported by Openshaw (1993); Fischer and Gopal (1994); Reggiani and Tritapepe (2002) showed that they can outperform conventional spatial interaction methods. Neural network techniques however still work by recognising patterns in the supplied data, whereas the intention in this thesis is to model the processes which produce these patterns. Neural networks will therefore not be discussed any further.

Fuzzy logic and fuzzy hybrid techniques have also been used to improve the performance of spatial interaction models. Initial work was described by Openshaw and Openshaw (1997); Openshaw *et al.* (1998) with research by See (1999) demonstrating the vast potential of fuzzy hybrid spatial interaction models. Here, fuzzy logic was used to optimise the spatial interaction model instead of traditional calibration methods. The focus here is not to improve the spatial interaction model, rather to use the spatial interaction model to feed into another model, in this case, the multi-agent system (MAS).

However, no evidence could be found within the literature at the time of writing for the hybridisation of spatial interaction models and MAS. Perhaps this is not surprising as much of the recent research into spatial interaction modelling with AI techniques has been focused on improving the performance of the spatial interaction model. The research within this thesis does not seek to improve the theoretical basis of spatial interaction models, rather to use these models to improve the performance of a MAS. By taking this approach, this thesis creates a new avenue of potential research for both areas.

6.5 Discussion

The main purpose of the spatial interaction model is to account for customer behaviour, thereby providing local controls within the system. In short, the spatial interaction model will alter the fuel sales at each station as customers react to changes in fuel prices at different locations. This lends a more realistic behaviour pattern to the overall model. However, which of the models reviewed above is the most appropriate for use within this research? This decision will be based on the requirements of the system and ease of implementation.

The market interdependencies between petrol stations are hypothesised to be dependent upon the daily activities of consumers, such as price awareness, the spatial configuration of the market, company pricing policies and external influences such as crude oil prices. For example, the study area (West Yorkshire) is divided into numerous wards i containing N petrol stations at fixed locations. Within their ward[3], consumers must decide where to make their purchases based on knowledge of a subset of petrol stations. Consumers residing in any market i are conceptualised as having familiarity with a subset of petrol stations and are limited in the volume that they can buy each day. The sale of petrol within ward i is shared by several stations j. There is no constraint on the amount of petrol that each station can sell per day. The market share per station depends on factors such as fuel price and proximity to i.

The constraints on the system are at the origin; consumers can only buy x amount of petrol per day. The natural approach would be to use Wilson's origin-production constrained model

[3]Consumers are not necessarily resident in a ward; they maybe passing through to their ultimate destination.

(Equation 6.2). However, it is desirable for stations located closer to the ward centroid to have a higher attractiveness than petrol stations located at a distance. A cost function is required that allows this. One choice would be Wilson's entropy maximisation model. Use of this model would fit well into the model formulation.

6.6 Construction of the Spatial Interaction Model

The following sections outline the construction of the spatial interaction model, specifically detailing the formulae and strategies used.

6.6.1 Components of the Spatial Interaction Model

The spatial interaction model that will be used is based on Wilson's origin-production constrained model (see §6.3) and entropy maximisation model (see §6.4.1). The amount of fuel required by each ward, O_i, is assumed to be H_iF, where H_i is the number of households in ward i and F is the average amount of fuel required per household. The assumption is made that prices do not rise sufficiently to reduce consumption in the short term. W_j is modelled as $e^{-\lambda p_j}$ where p_j is the price of the fuel at station j, so that stations with a high price are less attractive to consumers. The cost function C_{ij} is the distance d_{ij} between ward i and station j. In the entropy formulation, the cost function is used through the exponential $e^{-\beta d_{ij}}$ so that the amount of fuel sold decays with distance from the station. β and λ are coefficients that determine how the fuel sales vary with distance from the station and the fuel price (further details of their choice can be found in Chapter 7). This gives the formula

$$S_{ij} = \frac{(H_iF)(e^{-\lambda p_j})(e^{-\beta d_{ij}})}{\sum_j e^{-\lambda p_j} e^{-\beta d_{ij}}} \tag{6.8}$$

for the fuel, S_{ij}, sold by station j to ward i. This is for a single fuel type; however, it can be applied to different fuel types with potentially different values of p_j, F, β and λ. The sum on the bottom of the fraction represents the normalising process to ensure that the total volume of each fuel sold to each ward equals the demand H_iF.

The probability that a consumer will purchase petrol from any of the stations is hypothesised to depend on the price of petrol in that sub-market relative to the prices charged at the other sites. There are two elements to petrol prices, the cost of the petrol and the transportation costs incurred by the consumer going to the garage. This model assumes that these transportation costs are negligible. This can be justified where journeys are short and other purchases are being made. Plummer *et al.* (1998) hypothesise that this is the case for most petrol purchases because they are made in the consumer's immediate neighbourhood during the course of other activities rather than as the result of special trips. Evidence from Ning and Haining (2003) supports this assertion.

6.7 Potential Problems

Spatial interaction models have been thoroughly formalised from a theoretical perspective. However, for successful application of these models, there are a series of important design and implementation issues. The following sections outline the main considerations and actions taken to overcome them.

6.7.1 Data

Lack of sufficient data can have great consequences for the model (Openshaw, 1976). To avoid this, data of a suitable precision and accuracy were entered into the model (Table 6.1). The only data that had to be estimated was the different petrol types bought per household per week; no breakdown was given. The figures utilised were; 15% of sales are leaded petrol; 10% are super unleaded and 75% unleaded[4]. Due to these approximations, stations may make less profit, but the overall patterns should not be affected.

Data	Source
Ward Boundaries	Taken from UK Borders (accessed 2002)
Population per enumeration district	Derived from the UK census (Office for National Statistics, 1991) [a].
Number of households per ward	Derived from the UK census (Office for National Statistics, 1991).
Number of cars per ward	Derived from the UK census (Office for National Statistics, 1991).
Total amount of petrol sold per week	Derived from household expenditure survey (Office for National Statistics, 1998).

[a]The 2001 census data was not available at the time of development and experimentation.

Table 6.1: Data sources used in the spatial interaction model.

Using household data makes an implicit assumption that a consumer's journey started at their home. The distance to the petrol station was calculated from this point. This ignores the fact that people may buy petrol during the course of other journeys (see §6.6.1). This will be investigated further in §6.9.

6.7.2 Size of Areal Unit

The model is based on UK ward geography. The choice of size and configuration of these units will affect model performance (Thomas and Huggett, 1980). This is due to the modifiable areal unit problem (Openshaw, 1983; Fotheringham, 1981). This problem describes the incidence of different model results occurring for alternative spatial configurations of zones and also at different levels of spatial aggregation. Areal zones should ideally be as small and homogeneous as possible to avoid this problem. This is particularly important for the calculation concerning the spatial interaction model as consumers are assumed to live at the geographical centroid of the unit. Use of a small geographical region will reduce the distance between the consumer and petrol station (a factor that is important to the calculation of β and λ in Equation (6.8)). This should provide a better resolution of data. Figure 6.1 illustrates the two choices available, wards and enumeration

[4]The total fuel sales of petrol and diesel are known. The estimates for leaded, unleaded and super unleaded are based on consumer information (Office for National Statistics, 1998).

districts (ED). Ideally, the smaller of the two, EDs would be used. These are composed of about 200 households (approximately 450 people). Wards are considerably larger; the average number of people within wards in West Yorkshire is 20,000. This figure varies across England. The main benefit of using wards is that there are fewer of them and therefore the number of calculations to be performed would be smaller than if EDs were used.

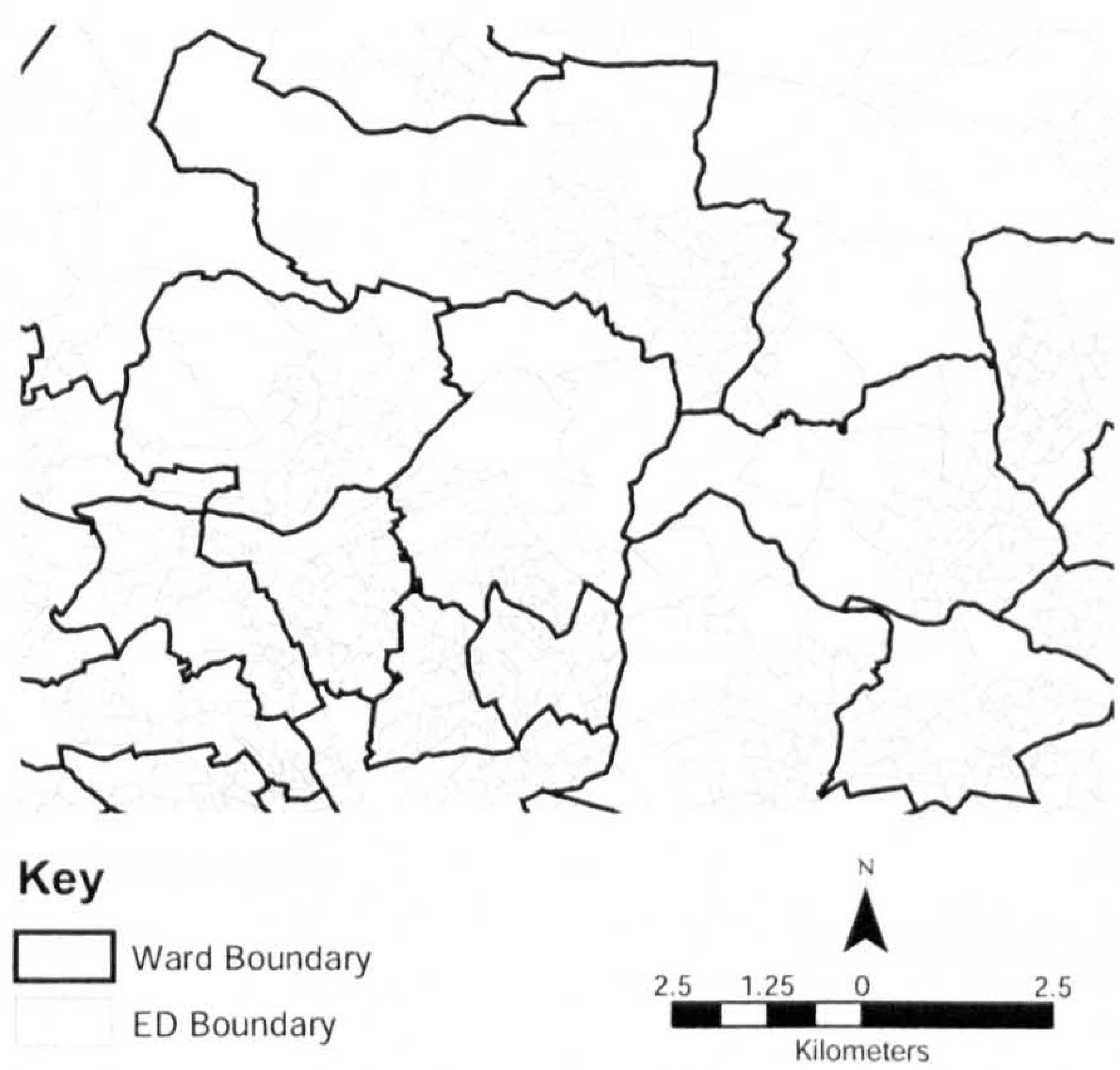


Figure 6.1: Illustrative diagram comparing the geographical extent of wards and EDs within a section of West Yorkshire.

A compromise that draws on some of the benefits of using the smaller EDs while retaining the computational efficiency of using wards is a population weighted centroid. Each ward is divided into its constituent EDs (Figure 6.1). The population within these EDs can be used to calculate a new location for the ward centroid.

Comparing the locations of the population weighted and geographical centroids, the greatest differences can be seen within the rural wards (Figure 6.2). These areas are characterised by being both geographically large and possessing low population densities. Using the population method has pulled the ward centroid nearer to the populated areas of the ward. Within the urban areas, there is little change in the position of the population and geographical ward centroids. This is expected as urban wards are geographically small and more uniformly densely populated.

6.8 Linking the Spatial Interaction and Agent Model

Linking the spatial interaction model to the agent model was undertaken by building two new classes in the model code; Wards and SIM. A Ward object was created for each ward of the study area; this contains information on the number of cars, ward population etc. One SIM object was created; this contains Equation (6.8). The SIM object collects all the information from the Ward and Petrol classes (containing price and location of the other petrol stations) objects and calculates the flow matrix containing the fuel sales for each station. These sales figures are passed back to the Petrol objects. New rules can be inserted into the Petrol object to make use of this information.

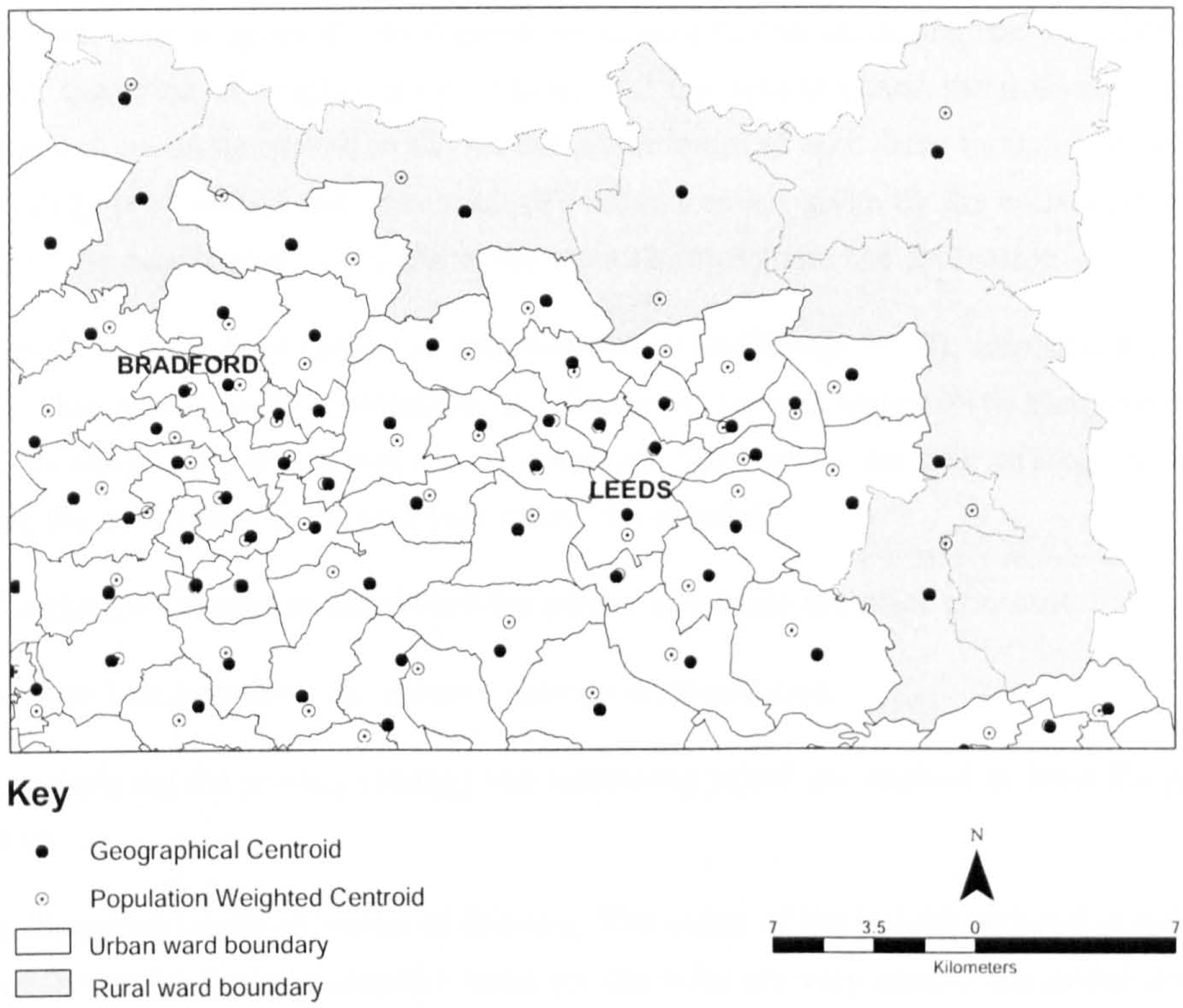


Figure 6.2: Location of the geographical and population weighted centroids within a selection of wards within the study area.

6.8.1 Profit/Strategy

The lack of evidence surrounding the strategies employed by petrol companies has been highlighted on several occasions throughout this thesis. It is important to reiterate this point as one of the central objectives of this thesis is to attempt to recreate the reality of the petrol price market. Given the absence of this information (with the exception of Esso's Price Watch), assumptions have to based on inferences from analysis of the real data and evidence from the literature. A basic assumption that can be made is that the dominant strategy employed by petrol stations is to maintain and maximise their profit. Taking this approach does however discount other potential motivations that could be of importance, for example attainment of market share or implementation of a pricing policy such as Esso's Price Watch[5].

A simple guide would therefore be to state that companies can either increase or decrease their prices to maintain their profit levels. Which of these strategies will give the best results? If stations increase their prices, they will increase their profit levels at the potential expense of their sales. Decreasing their price may initially reduce profit, but in the long-term this may increase sales and thus profit. An indicator of the likely behaviour of stations can be seen through the "rockets and feathers" effect. Although this refers to the reaction of stations to an external increase in crude oil prices, the rapid response of stations to higher prices may also be evident within a local neighbourhood.

[5]The Price Watch policy may not be the optimal policy for maximisation of profit levels.

Which behaviour is optimal may depend on several factors including the availability of local competition, the price of neighbouring stations and the sensitivity of the market to price. The strategy that will be implemented in the model will attempt to take these factors into account.

The strategy is to change the price each day by an amount given by the variable *priceChange*. To determine the new strategy (and price) the stations implement the following:

1. If (profit is falling) or (profit is negative and *priceChange* == 0), implement a new strategy. This new strategy is obtained as follows. If (*priceChange* != 0) then *priceChange* = -*priceChange*, i.e. the strategies are swapped. Otherwise, the new strategy is determined using the getStrategy() method (see below for details).

2. If the change in profit is small then the policy is to keep the price constant.

3. If neither 1 or 2 applies, the current strategy is maintained.

4. After applying the pricing strategy the remaining rules[6] are applied to limit the price at the station.

The getStrategy() method works as follows. The mean of the neighbourhood is calculated. If the price of the station is more than the mean (or the sales are very small), the initial strategy is to decrease the price. If this is not the case, the initial strategy is to increase the price. The rationale behind this is that if a station's price is already high (in relation to its surrounding area), it cannot afford to further increase its price without risking loss of sales. Similarly, if the sales are very low, the station cannot risk losing even more sales. In these cases, the best strategy is to decrease the price. However, if the price is low, the station is already competitive and decreasing the price further may only serve to deplete profits even if there is an increase in sales[7].

To account for the strategies, three equations were built into the model. The income for station j from fuel type m is given by

$$D_j^m = \sum_i S_{ij}^m p_j^m \tag{6.9}$$

where

- D_j^m is the amount of money that station j takes from fuel m.
- S_{ij}^m is the amount of fuel bought by ward i from station j.
- p_j^m is the price of fuel m at station j.

The total cost of production is

$$C_j^m = \sum_i S_{ij}^m P_j^m(0) + Q_j^m \tag{6.10}$$

where

- C_j^m is the amount that it costs the station j to produce the petrol.

[6]For example, the amount that the stations can overprice and undercut by. These are the rules developed for the pure agent model (see Chapter 5).

[7]A simpler strategy was not used because experimentation with a fixed initial strategy (either to increase or decrease prices) was not effective.

- $P_j^m(0)$ is the cost per litre to produce and sell fuel m at station j.
- Q_j^m is the fixed cost to the station of selling fuel m at station j. This might include for example the cost of staffing the station and maintaining the buildings which do not depend on the amount of fuel sold.

The amount of profit of fuel m at station j is

$$M_j^m = (D_j^m - C_j^m) \tag{6.11}$$

where

- M_j^m is the profit made by station j selling fuel m.

6.8.2 Interaction Between the Agent and Spatial Interaction Model

The interaction between the agents and the SI model can be summarised as:

1. The pure agent model is initialised with the real data (Day 0).
2. Additional data e.g. ward data is read in.
3. The fuel price and station location data is passed to the SI model.
4. The flow matrix is calculated.
5. The flow matrix values are used to "inform" the agents how much fuel has been sold each day. Stations can use this information to calculate their profit and set their strategy (see §6.8.1 for a detailed outlined of strategies implemented).
6. The SI model gets the new prices from each station and the simulation returns to step 4 until completion.

Where a particular fuel type is not sold at a station, the SI model ignores it and the value is set to zero. Figure 6.3 diagrammatically represents the procedure of the SI model and agent interaction.

The full source code and a compiled version of this hybrid model can be found on the CD accompanying this thesis.

Given the addition of profit, several new strategy variables could be added over and above those of the pure agent model (Table 6.2). Values for these new parameters will be calculated in Chapters 7 and 8. To allow selection of these additional parameters at the same time as the agent rules, the graphical users interface (GUI) was enhanced (Figure 6.4). This creates a very powerful and flexible framework within which detailed rule sets can be assigned to, for example all stations or particular brands or categories.

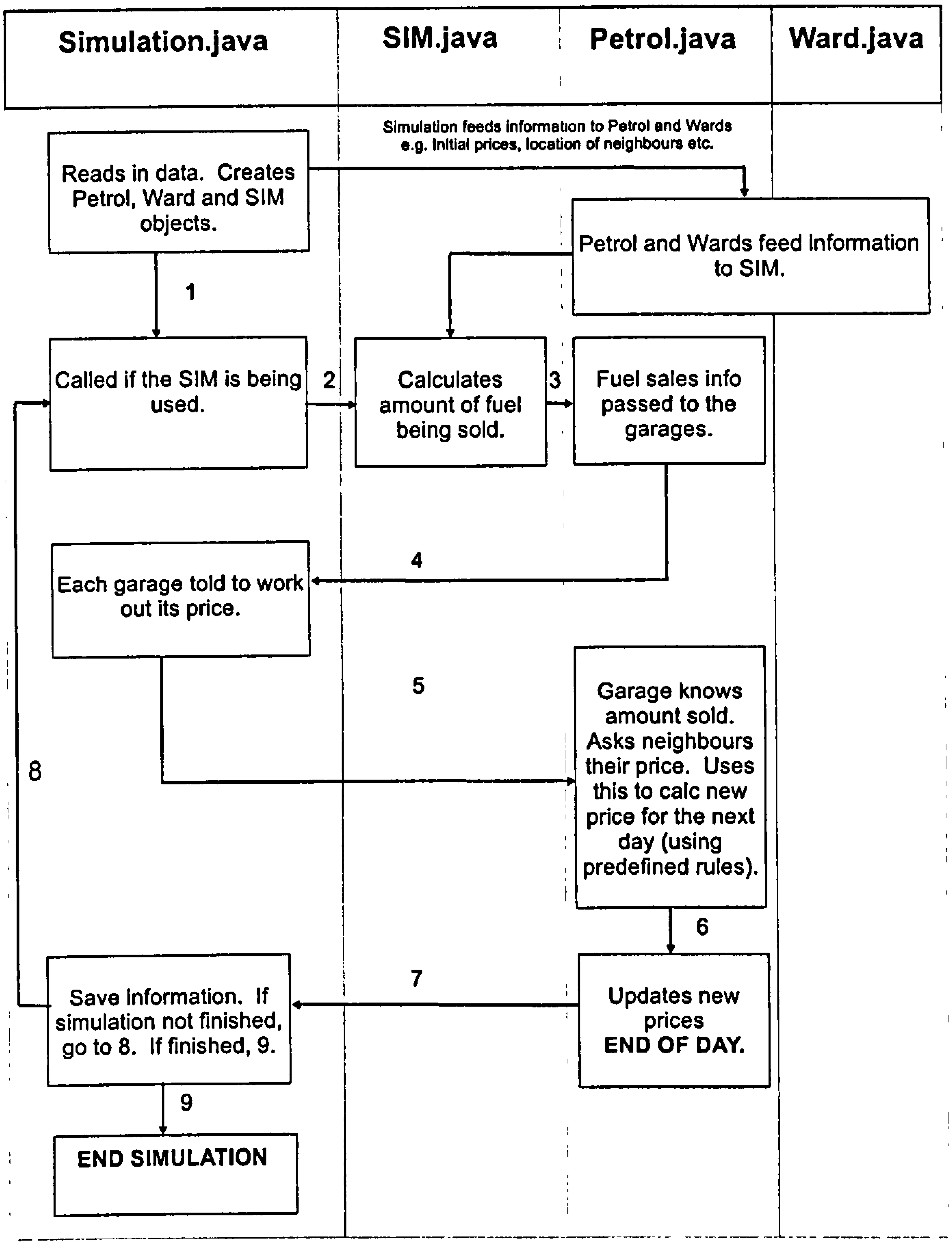


Figure 6.3: Integration of the spatial interaction model into the Java classes.

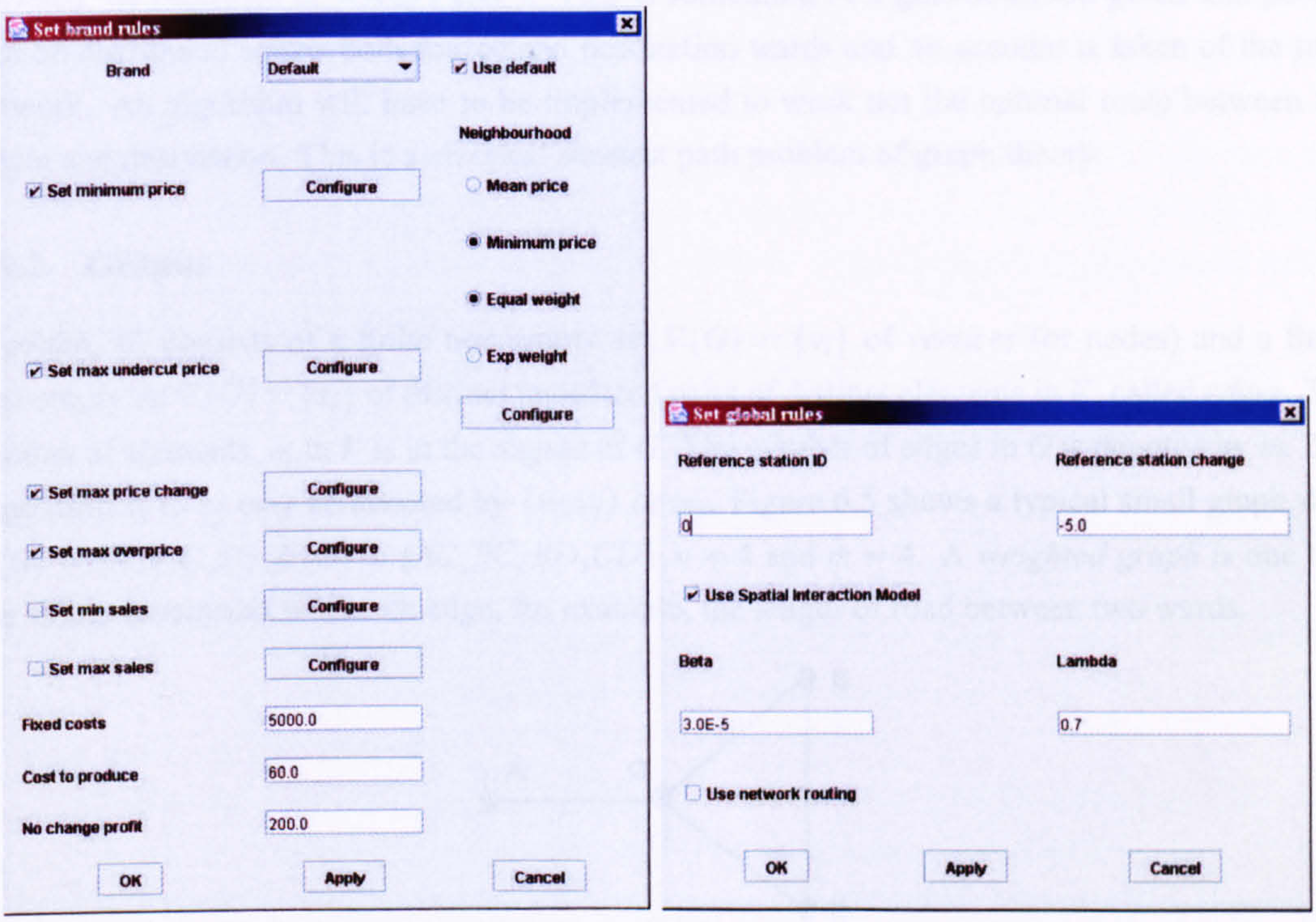
6.9 Redistribution of the Population

One criticism that could be levelled at this approach is the assumption that consumers only buy petrol from the ward that they live in. Research undertaken by Ning and Haining (2003) suggests that this is an unrealistic premise. Part of this research surveyed households about their petrol buying habits. Responses showed that petrol was most frequently bought as part of a trip to work. Shopping trips and social or recreational trips also accounted for a large proportion of trips where petrol was bought. Very few respondents made special trips to purchase petrol.

To account for this behaviour within the system, a method is required that will allow the population to be redistributed around the study area based on journeys made. An intervening opportunities or competing destinations model would require detailed data on the travel paths for

Parameter	Explanation
β	Coefficient controlling the influence of distance.
λ	Coefficient controlling the influence of price.
$costToProduce$	The amount per litre that it costs the station to produce and sell the petrol.
$fixedCosts$	The amount the petrol station has to pay per day to keep running.
$changeInProfit$	The level of profit under which the station will not change its strategy.

Table 6.2: New parameters incorporated within the hybrid model.



(a) Parameter Setting (b) Global Parameter Setting

Figure 6.4: Screenshots of the enhanced dialogs in the GUI allowing the specification of rules for the SI model

the whole population within the study area. Unfortunately, data of this resolution is not available (see below).

An approach that would fit in the overall framework of the model and data availability is a network model. By implementing this technique, an alternative method of distributing the sales is created. The theory and construction of this model are detailed in the following sections [8].

[8]This work was undertaken with David O'Sullivan and James MacGill within the Geovista Centre, Pennsylvania State University. The work formed part of a exchange program funded by the World University Network (WUN).

6.9.1 Data

The data used for redistributing the population was a database of travel to work data derived from WICID (1991) based on the 1991 UK census. It is drawn from a 10% sample of the population and is non-cascading (i.e. only the origin i and destination j of the consumers are known, the route taken is unknown). The data is based on journeys to work including movement between wards. The flows represent the number of cars travelling to work, not the number of people. There are two problems with data of this nature. Firstly, the data for each ward is not complete, but is just a 10% sample of the population. However, assuming the sample is representative, this should give a fair indication of population movements. Secondly, and potentially more seriously, the assumption has to be made that, in the absence of any other information, consumers will take the shortest path between the centroids of wards i and j. This is something of a generalisation given that people will be distributed across both source and destination wards and no account is taken of the road network. An algorithm will have to be implemented to work out the optimal route between the origin and destination. This is a classical shortest path problem of graph theory.

6.9.2 Graphs

A *graph*, G, consists of a finite non-empty set $V(G) = \{v_i\}$ of *vertices* (or nodes) and a finite non-empty set $E(G) = \{e_i\}$ of distinct unordered pairs of distinct elements in V, called *edges*. The number of elements, n, in V is in the *degree* of G. The number of edges in G is denoted by m. The edge from v_i to v_j may be denoted by $\{v_i, v_j\}$ or e_{ij}. Figure 6.5 shows a typical small graph with $V(G) = \{A,B,C,D\}, E(G) = \{AC,BC,BD,CD\}, n = 4$ and $m = 4$. A *weighted graph* is one that has values associated with each edge, for example, the length of road between two wards.

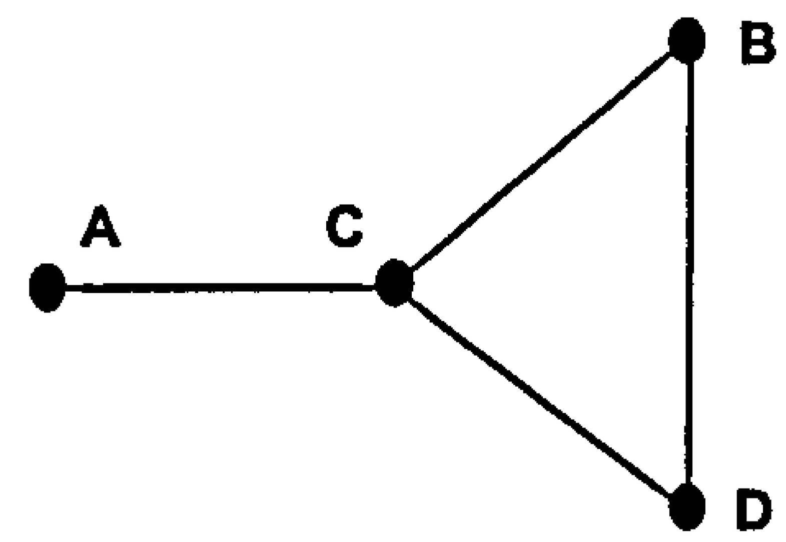


Figure 6.5: A typical graph (after O'Sullivan, 2000)

In graph theory, the shortest path problem is the following: Given a weighted graph, that is a set N of nodes, a set E of edges and a real-valued function ($f : E \longrightarrow \mathbb{R}$), and given further two elements n_1, n_2 of N, find a path P from n_1 to n_2, so that

$$\sum_{p \in P} f(p) \tag{6.12}$$

is minimal among all paths connecting n_1 to n_2. One of the most important algorithms for solving this problem is Dijkstra's algorithm (Dijkstra, 1959).

6.9.3 Dijkstra's Algorithm

Dijkstra's algorithm solves the shortest path problem for the case where the weighting of the graph edges are all non-negative. For example, if the vertices of the graph represent the centre of a ward and the edge weights represent driving distances between wards, Dijkstra's algorithm can be used to find the shortest route between two wards. The condition on the graph weighings being non-negative has to be satisfied as the algorithm assumes that path lengths increase as the number of nodes in the path increases. This is obviously the case when the weights are the distances between wards.

The algorithm (see Wikipedia, 2003) works by keeping for each vertex v the length $d[v]$ of the shortest path found so far. Initially, this value is 0 for the source vertex s and infinity for all other vertices, representing the fact that we do not know any path leading to those vertices. When the algorithm finishes, $d[v]$ will be the length of the shortest path from s to v or infinity, if no such path exists.

The basic operation of Dijkstra's algorithm is edge *relaxation*: if there is an edge from u to v, then the shortest known path from s to u can be extended to a path from s to v by adding edge (u,v) at the end. This path will have length $d[u]+w(u,v)$. If this is less than $d[v]$, we can replace current value of $d[v]$ with the new value.

Edge relaxation is applied until all values $d[v]$ represent the length of the shortest path from s to v. The algorithm is organised so that each edge (u,v) is relaxed only once, when $d[u]$ has reached its final value.

The algorithm maintains two sets of vertices S and Q. Set S contains all vertices for which we know that the value $d[v]$ is already the length of the shortest path and set Q contains all other vertices. Set S starts empty, and in each step one vertex is moved from Q to S. This vertex is chosen to be the one with the lowest value of $d[u]$. When a vertex u is moved to S, the algorithm relaxes every outgoing edge (u,v).

There are limitations with using this approach. The algorithm used for determining the shortest path between i and j is based on the calculation of distance and direction. Movement of consumers between i and j assumes that the journeys are made without detours and the consumers are aware of the shortest route and take it. Overall, this may not produce an appropriate distance. Ideally, if the appropriate data was available, the edges would be journey time, not distance. This approach also assumes that the population lives at the centre of the ward, in reality they would be distributed throughout the ward. Furthermore, in "real life" movements would normally be made along roads, and while the road network is dense in the UK, there will be some variation from the direct distance.

6.9.4 Nodes and Edges

To re-distribute the population around West Yorkshire using Dijkstra's algorithm, nodes and edges have to be established. The nodes can be represented by the population weighted centroids developed in §6.7.2 and connections between neighbouring wards can be represented by weighted edges (Figure 6.6). The weights represent the distance between neighbouring nodes.

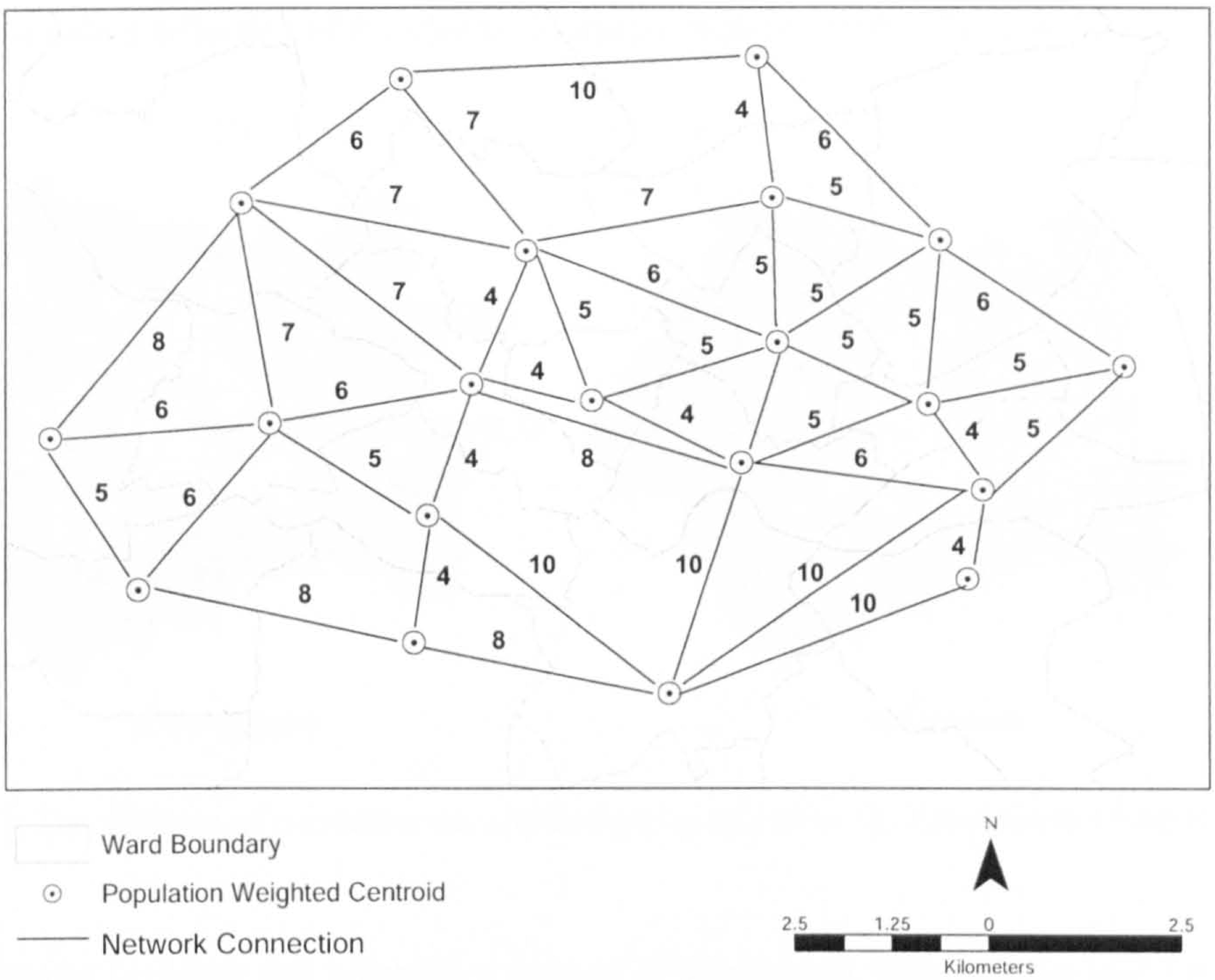


Figure 6.6: Example of a weighted graph network.

6.10 Running the Network Model

The basic network model is run before any of the interactions between the agent and spatial interaction model take place to calculate the customer's location data. The shortest path between each pair of wards is calculated. Each route in the journey to work data is taken as the shortest path between the start and end wards. The model takes each route in the journey to work data and redistributes the people from the origin ward taking this route equally amongst each ward the route passes through. This creates a redistributed population surface that is fed to the spatial interaction model in Stage 3 (see the interactions detailed between the agent and spatial interaction model in §6.8). The network model is not run again during the course of the simulation.

The full source code and a compiled version of the hybrid model (including the network extensions) can be found on the CD accompanying this thesis.

6.10.1 Re-distribution of Consumers

What effect has running this network model had on the performance of the hybrid spatial interaction-agent model? The distribution of consumers before and after the application of the networking was calculated (Figure 6.7). Figure 6.7 (a) clearly shows that before the networking was applied, high densities of consumers were found predominately in the centre of the study area. This area covers the West Yorkshire conurbation of Leeds, Bradford, Wakefield, Huddersfield and Halifax. In the suburban areas of Leeds and Bradford, the density of consumers is at its highest. After the networking has been applied (Figure 6.7 (b)), all of the city areas, not just the suburbs are characterised by having an even higher density of consumers. This increase in population within the city centres is simply a reflection of the number of people who live within the suburbs, but work in the cities.

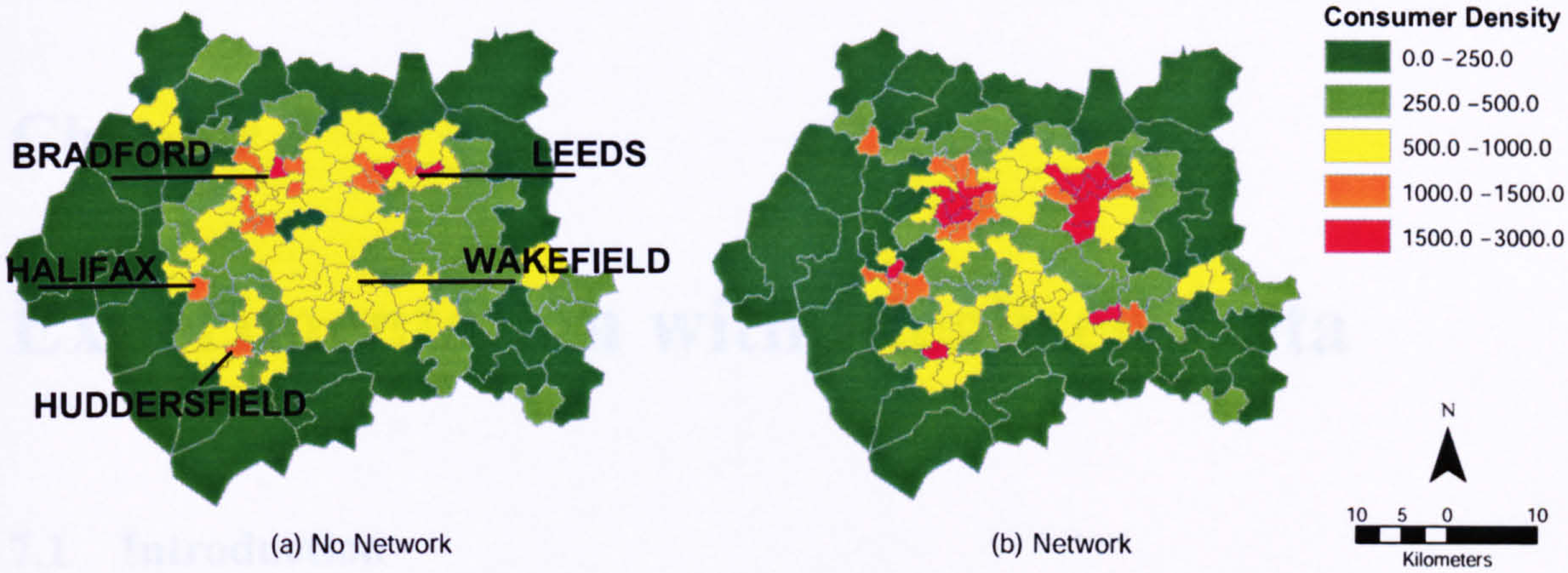


Figure 6.7: Distribution of consumer density before (a) and after (b) application of the networking.

The impact of using this alternative method of population redistribution will be tested out using the real data for West Yorkshire in Chapter 8.

6.11 Conclusion

One of the main criticisms levelled at the pure agent model presented in Chapter 5 was the lack of customer behaviour. This chapter has seen the development of a spatial interaction model to address this problem. Wilson's origin-production constrained model was selected as the most appropriate of the models reviewed and linked to the pure agent model. The resulting hybrid model is not only entirely novel, but provides a powerful framework for experimentation. The addition of the spatial interaction model has also allowed a new set of sophisticated strategies based on stations' profit to be implemented.

However, this "hybrid" approach is not without its criticisms. Consumers were assumed to buy petrol from the ward that they lived in. Research by Ning and Haining (2003) showed this not to be the case. A network model grounded in graph theory was constructed to provide alternative method of redistributing the population based on journey to work data. The robustness and performance of this new hybrid model will be tested out in Chapter 7. Application to the real data and comparison against the agent model will take place in Chapter 8.

Chapter 7

Experimentation with Idealised Data

7.1 Introduction

Chapter 6 presented the extension of the agent framework to incorporate consumer behaviour. This was achieved by developing and linking a spatial interaction model to the agent system. A simple network model was developed to provide an alternative method of redistributing the sales. The aim of this chapter is to examine the behaviour of this hybrid model. However, as seen in the analysis of the real data in Chapter 4, population density and petrol station distribution are not uniform across the study area. This presents an added complication when attempting to understand the behaviour of the model and the effect of individual parameters. By standardising the geography and population through idealised data, the behaviour of the model can be more easily understood. A thorough understanding of the behaviour and sensitivity of the hybrid model in idealised situations is essential before application to the more complex real system.

The work within this chapter is divided into four sections. Each section is linked by the common aim of improving knowledge of the model behaviour. The first section uses simulations to investigate the sensitivity of the spatial interaction model parameters. Following this, a series of diffusion experiments will be presented. These simple experiments have a dual purpose, they provide a test of the model behaviour whilst allowing the influence of individual parameters to be noted. Experiments investigating the effect of varying population and petrol station distribution will follow this. Finally, the sensitivity of the model to perturbing parameters will be examined. Conclusions drawn from these experiments will be used in Chapter 8 with the investigation of the real data.

7.2 Sensitivity Analysis with Spatial Interaction Model Parameters

Chapter 6 provided an overview of spatial interaction models and detailed the construction of one such model. This was linked to the agent model to account for consumer behaviour. This section uses numerical experimentation to evaluate the contribution of the spatial interaction model parameters of β, λ, *fixedCosts* and *costToProduce*. To isolate the effect of the spatial interaction model from the rest of the hybrid model a version of the spatial interaction model was constructed within an Excel spreadsheet using a 20km x 20km grid[1] (Figure 7.1). Petrol stations were located at nodes equally spaced at 2.0km. Each square represents a ward (the ward centre is located at the centre of the square) and contains a population of 2000 and 1000 cars. Each station is at the junction of four wards.

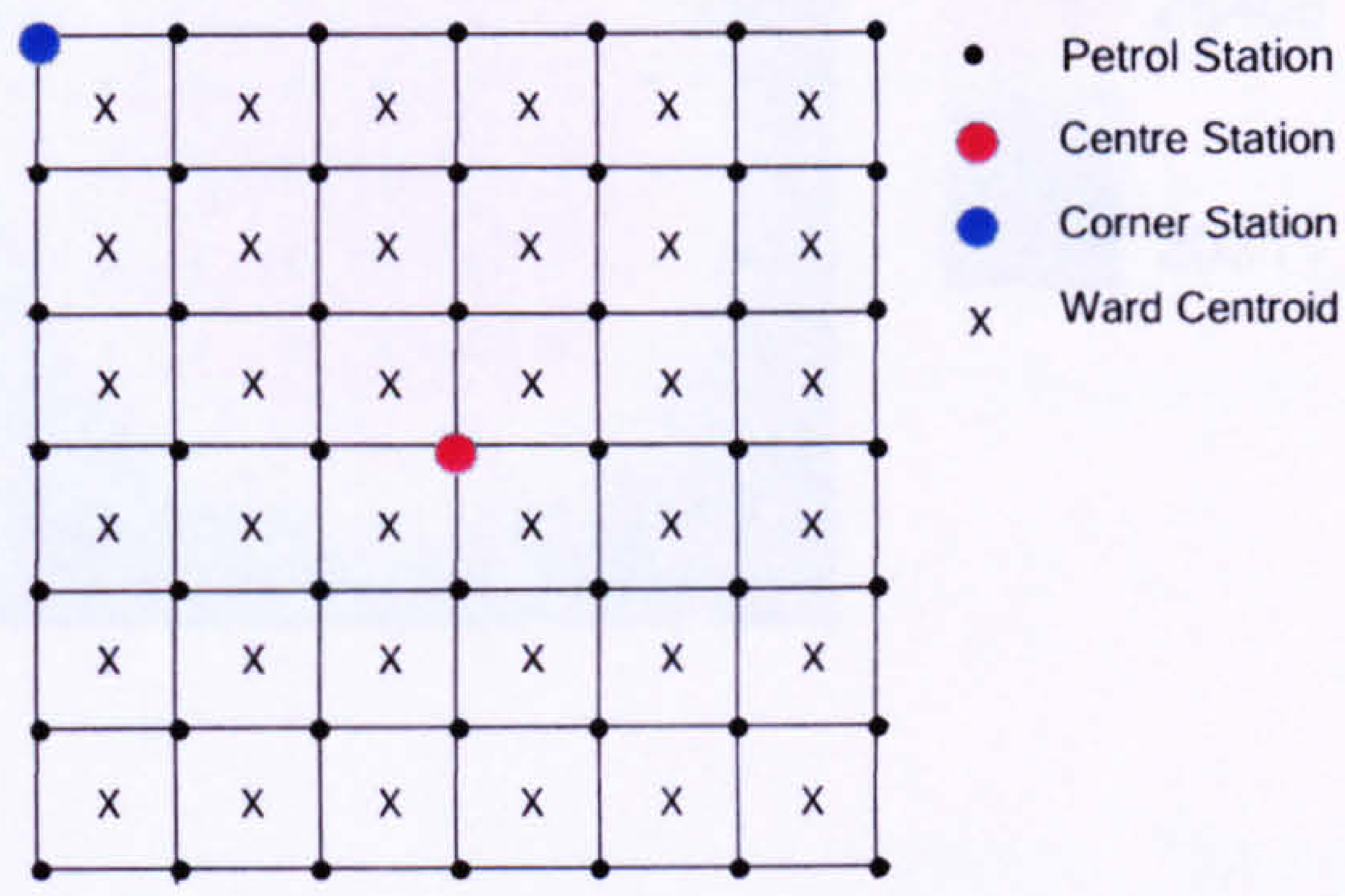


Figure 7.1: Example of the type of grid used within the idealised simulations.

One of the constraints with using such a small grid is that there will be a noticeable difference in the amount of profit that stations located at the centre and edge of the grid will make. This is shown by Figure 7.2 which quantifies the effect of the edges on sales at each station for a given β. The stations at the centre of the grid are surrounded by more potential consumers than those situated at the edge of the grid and so sell more fuel. The profit at the centre station (61) and an edge station (1) will be compared in some of the subsequent experiments to quantify this effect.

7.2.1 Experimentation with β

The parameter β was introduced in the construction of the spatial interaction model (see Chapter 6; Equation (6.8)). This parameter determines how the fuel sales vary with distance from the station. The degree of influence that β exerts can be calculated. For example, if all the stations within the idealised grid were set to the same price, the amount that a petrol station, j, sells to ward i is proportional to $e^{-\beta d_{ij}}$.

Re-arranging this equation, the spatial scale over which the stations are selling petrol can be calculated. Taking this as the distance at which the sales would halve gives

$$\begin{aligned} e^{-\beta d_{ij}} &= \frac{1}{2} \\ -\beta d_{ij} &= -\log(2) \\ d_{ij} &= \frac{1}{\beta}\log(2). \end{aligned} \tag{7.1}$$

For example, if β was set to a value of 0.0003, this would produce a distance scale of 2310m. This means that any consumer located more than 2310m from the station would be half as likely to buy petrol from it as a consumer located adjacent to the station. All the stations are initialised

[1] The size of the grid used was limited by the capacity of the spreadsheet (in particular, the number of columns that are allowed). 20km x 20km was the maximum grid size that could be computed.

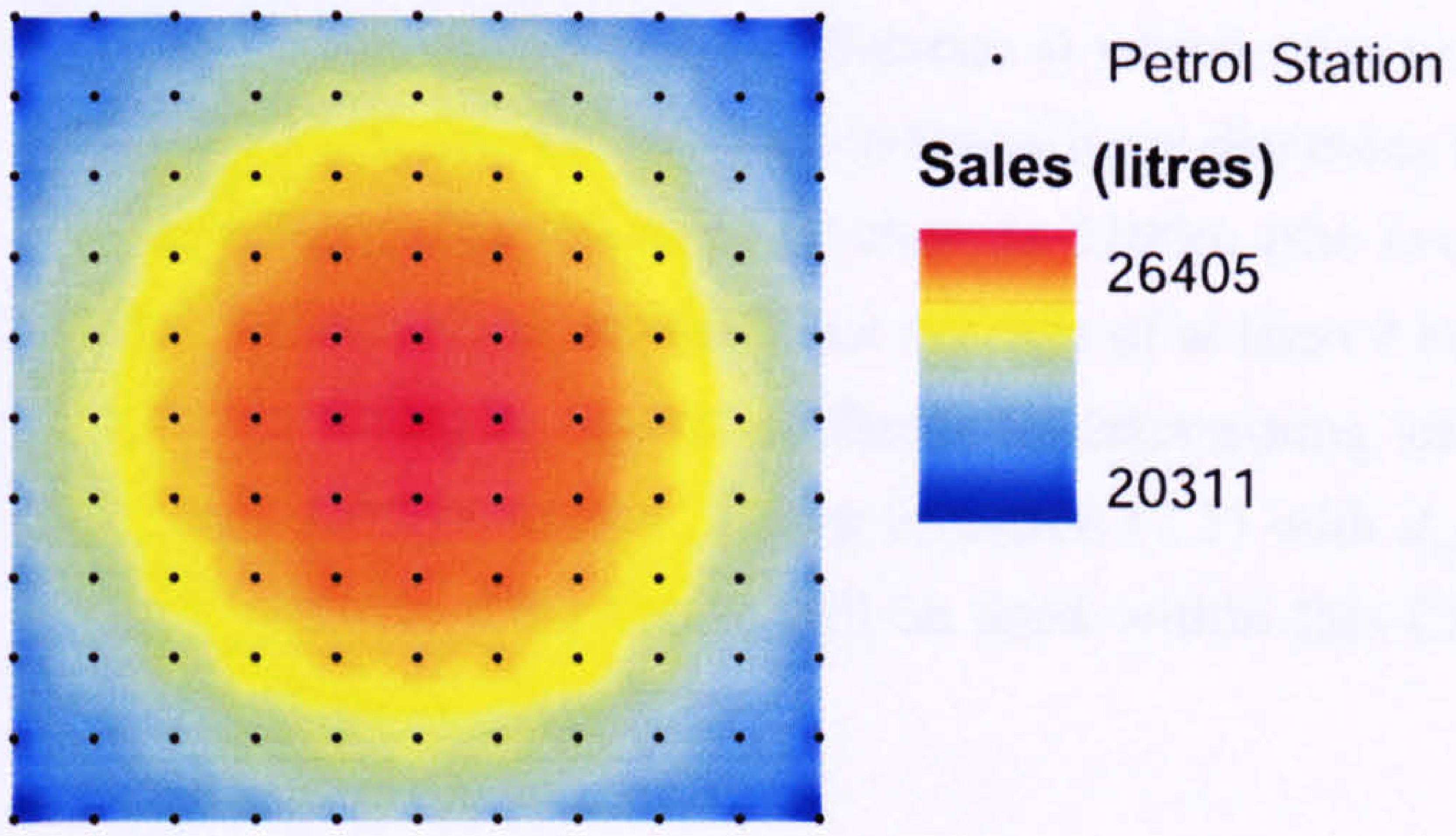


Figure 7.2: Spatial distribution of the volume of sales (litres) across an idealised grid assuming constant prices produced by the spatial interaction model. The dots mark the location of the petrol stations at the corner of the wards.

at 68p, λ is kept constant at 0.7 and the β is set to 0.03, 0.003 and 0.0003. (Note: *fixedCosts* and *costToProduce* are not needed to calculate sales.) The effects of varying β were analysed for the centre station, i.e. examining how much petrol was sold by this central station to consumers within all the wards. Figure 7.3 clearly shows that with a smaller β value, distance is less of an influence on consumers (there is not a great difference in sales within wards located near to the centre station and those located at a distance). With β set to 0.0003 and 0.003, consumers are willing to travel greater distances to purchase petrol. This is seen in the distribution of the volume of sales over the entire grid. However, increasing β to 0.03 increases the sensitivity of consumers to distance.

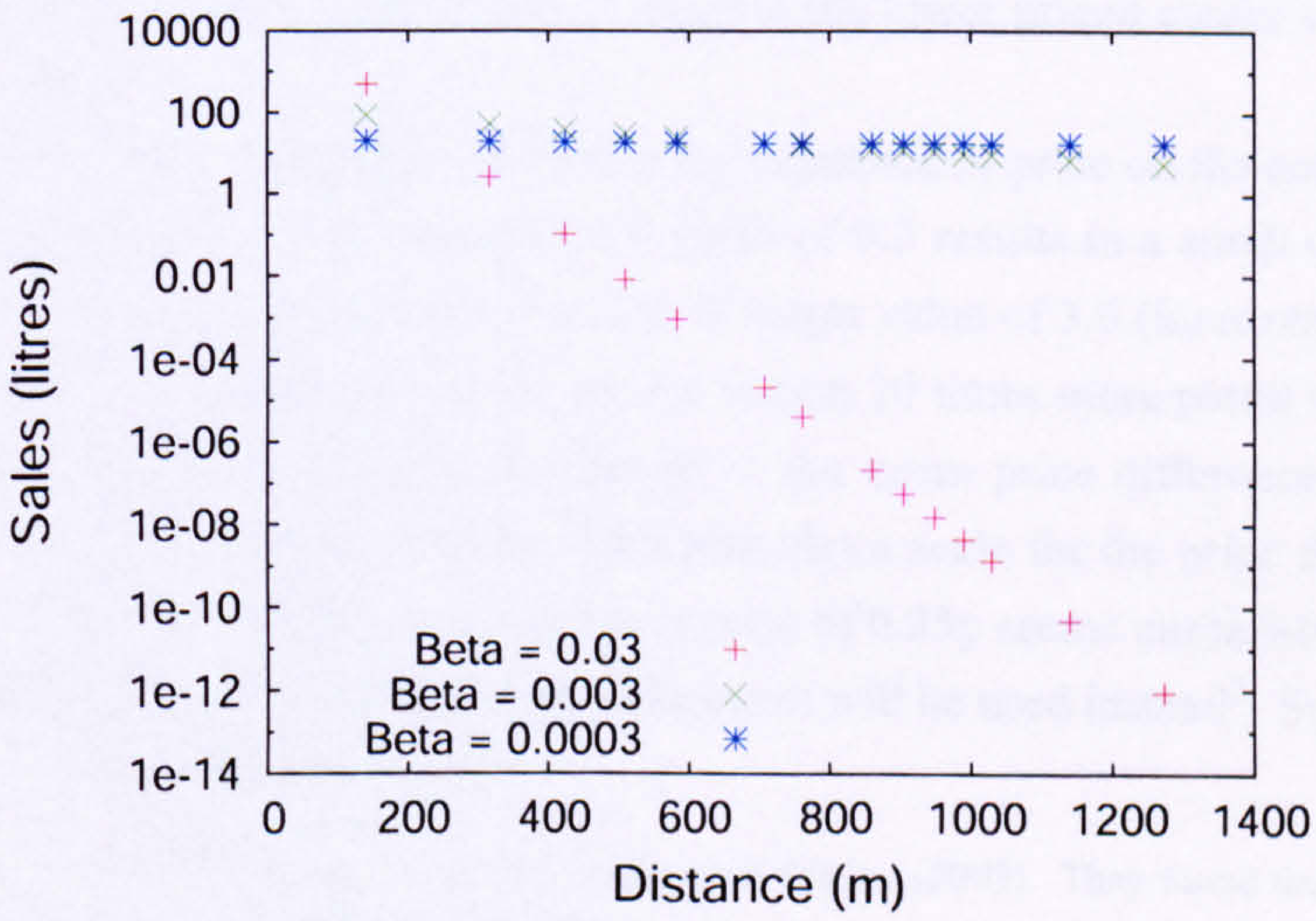


Figure 7.3: Figure showing the sales at the centre station as a function of distance from the consumer. Results are shown for different values of β.

Within West Yorkshire, a station with a 2km neighbourhood will have an average of 4 neighbouring stations. The average distance between these stations is approximately 700m (see §4.7.2). To estimate β using Equation (7.1), an estimate of the distance at which sales would drop by 50% is required. This is because consumers are less inclined to travel large distances to purchase petrol (Ning and Haining, 2003). Supposing this distance is taken as 2100m (the average distance between neighbours multiplied by three), there would be an average of at least 4 stations closer than 2100m. This would ensure that price is an important factor in determining sales between these stations thereby producing a competitive market. Using Equation (7.1) with $d_{ij} = 2100$m, β can be calculated at approximately 0.0003. This figure will be used within this Chapter, a separate figure will be derived for the real data in Chapter 8.

7.2.2 Experimentation with λ

The λ coefficient controls the importance of price to consumers. The degree of influence that it has can also be calculated using

$$\begin{aligned} e^{-\lambda p_j} &= \frac{1}{2} \\ -\lambda p_j &= -\log(2) \\ p_j &= \frac{1}{\lambda}\log(2). \end{aligned} \tag{7.2}$$

Assigning λ a value of 0.07 would mean that a 10p price increase at a station would result in the halving of its sales, i.e. custom would be lost to neighbouring competitors.

The spatial interaction model spreadsheet described in §7.2.1 was reused to examine the influence of λ upon the system. All the stations were set to 68p with the exception of the centre station (61) which was set to 67p. and β was fixed at 0.0003. The total sales at each station were analysed for different values of λ to investigate how the sales are affected by the cheaper competition at station 61. Figure 7.4 shows a comparison of sales at the lower priced centre station (61) and its neighbour (station 62).

The results show that increasing λ increases the influence of price on the consumers' decision on where to purchase petrol. For example, a λ value of 0.3 results in a small difference in sales between station 61 and its neighbour, station 62. A larger value of 3.0 (increasing the importance of price to consumers) results in station 61 selling almost 20 times more petrol than station 62.

Based on analysis of the real data in Chapter 4, the mean price difference between stations in West Yorkshire was calculated at 0.35p. This provides a scale for the price changes within the system. A 50% decrease in sales due to a price increase of 0.35p seems unrealistic. A price change of 1p (approximately 3 times the mean price difference) will be used instead[2]. Substituting 1p into Equation (7.2), λ can be calculated at 0.7.

[2]This value can be supported by the research of Ning and Haining (2003). They found that stations were never more than 1 - 2p more expensive than their neighbours.

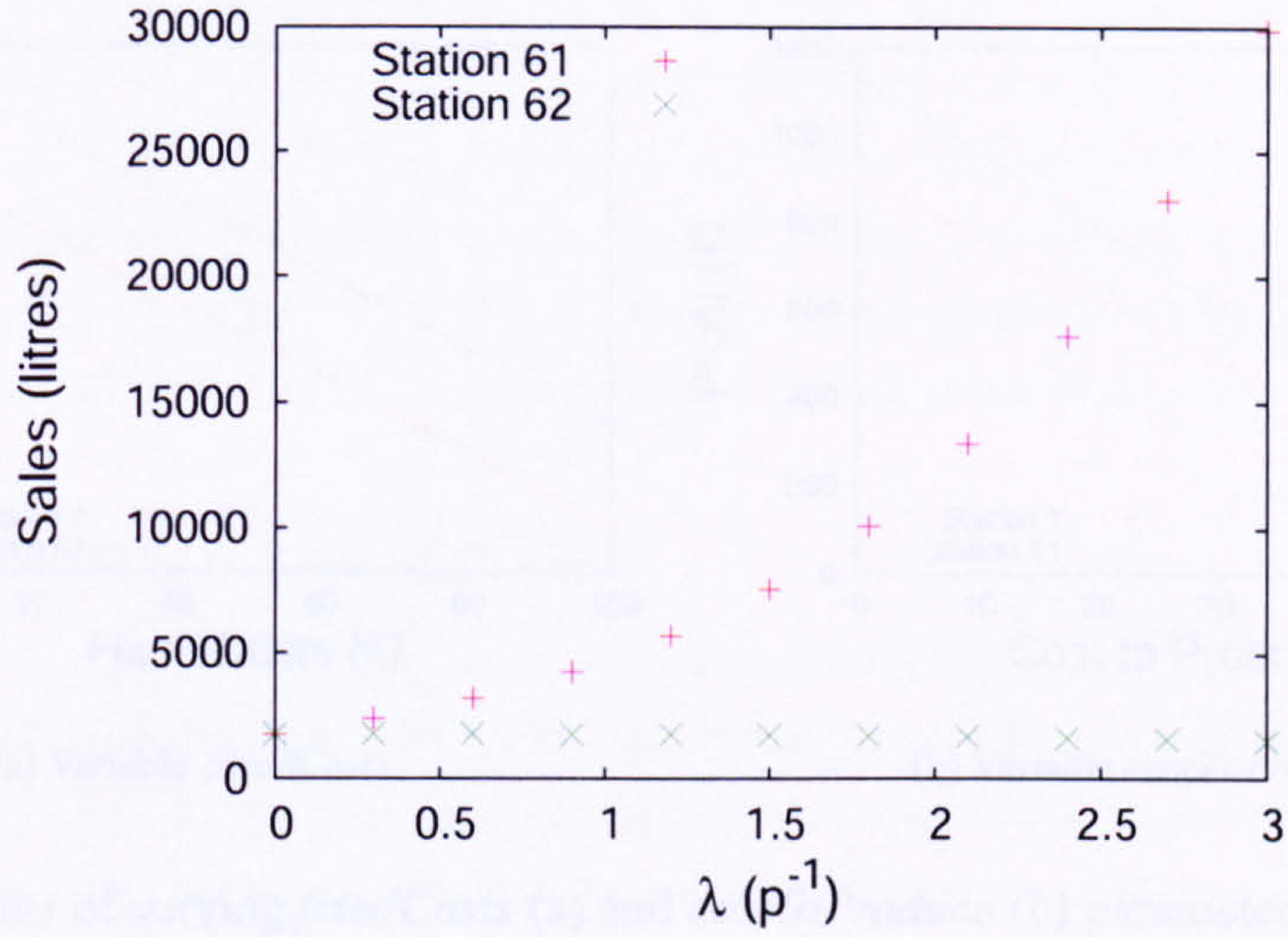


Figure 7.4: Comparison of the total daily sales (in litres) at station 61 and 62 for different values of λ.

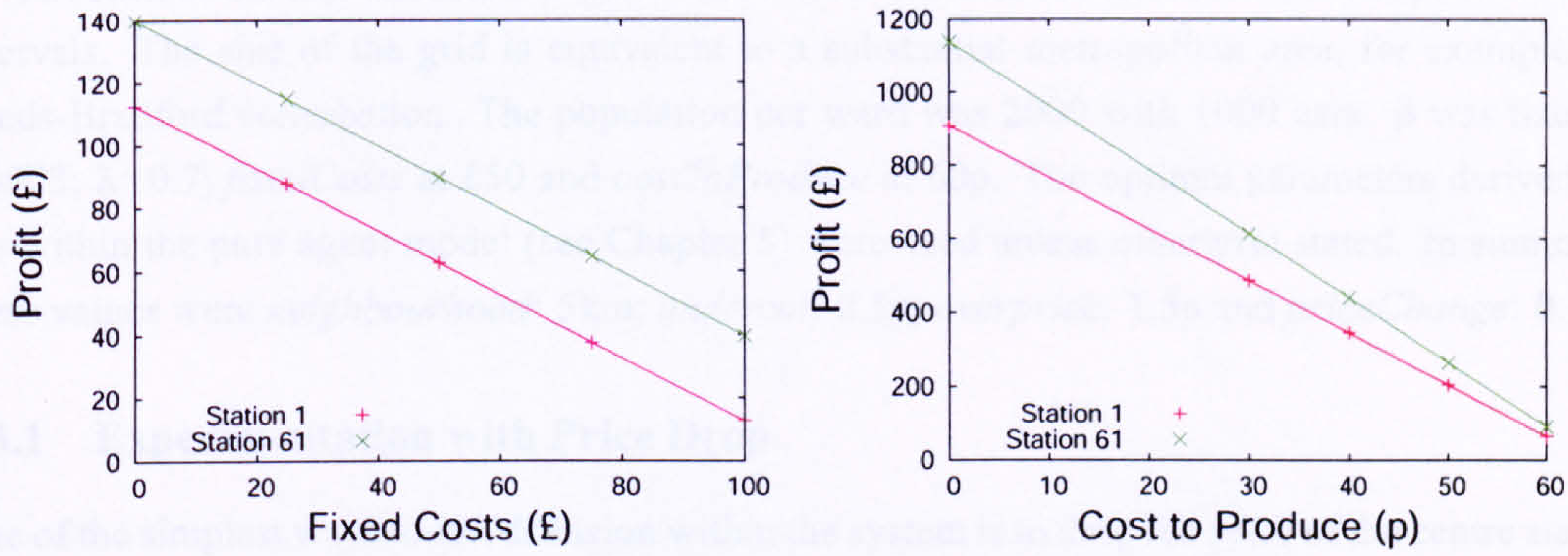
7.2.3 Experimentation with Fixed Costs and Cost to Produce

The *fixedCosts* variable represents the amount that the petrol station has to pay per day to keep running, for example, payment of staff. The *costToProduce* variable represents the amount per litre that it costs the station to produce the petrol. This includes crude oil prices and transportation costs. In practice part of the *fixedCosts* will actually be included in the *costToProduce* since, for example, a large station selling lots of petrol will be likely to employ more staff. The distinction between the two may therefore sometimes be a little hazy. It is also worth remembering that only a fraction of the *fixedCosts* are actually being considered since this thesis concentrates on just the unleaded petrol sales. Most stations will also generate income from sales of diesel and other fuels in addition to non-fuel sales from forecourt shops. The *fixedCosts* and *costToProduce* parameters will obviously exert influence on the amount of profit that the retailers make. The nature of the relationship between *costToProduce*, *fixedCosts* and profit will be examined within this section. It is hypothesised that a linear relationship will be evident, i.e. the amount of profit that each station makes will decrease with increasing costs.

The 20km x 20km grid described in 7.2 was used with all the stations initialised at 68p. β was set to 0.0003 and λ to 0.7. The *fixedCosts* variable was set to £50 when varying the *costToProduce* parameter. Similarly *costToProduce* was set to 60p when experimenting with *fixedCosts*. The value of £50 is likely to be smaller than the actual real fixed costs for the reasons discussed above.

Figure 7.5 shows that both *fixedCosts* and *costToProduce* show a strong linear relationship with profit. The amount of profit that each station makes decreases with increasing costs. In both cases, station 1 (located on the edge of the grid) makes less profit than station 61 (located at the centre of the grid).

Increasing the *fixedCosts* at a station reduces the amount of profit by the same amount regardless of the amount of petrol sold (Figure 7.5 (a)). This trend was not found with the *costToProduce* results (which are calculated per litre sold). As the *costToProduce* variable increases towards the



(a) Variable *fixedCosts*   (b) Variable *costToProduce*

Figure 7.5: Results of varying *fixedCosts* (a) and *costToProduce* (b) parameters on the amount of profit made by the centre and edge stations.

retail price of the petrol, none of the stations will make a profit. On the other hand, as the *costToProduce* decreases towards zero, the profit becomes increasingly dependent on the amount of petrol sold. As station 61 sells more petrol than station 1, a change in the *costToProduce* variable will have a larger absolute impact on its profit. This accounts for the different slopes of the cost functions in Figure 7.5 (b).

In terms of sensitivity, varying *fixedCosts* does not have a great effect on the model. This is because the profit at each station is affected in the same way. The costs that this parameter includes are not affected by the amount of petrol that is sold i.e. the station has to remain open, maintain the site and employ workers. The model may be expected to be more sensitive to the *costToProduce*. Changing this variable leads to different changes in profit for petrol stations with different sales. Since changes in profit are the principal driving force in the model this might be expected to lead to different pricing patterns for different values of the variable.

On the basis of these numerical experiments (and in the absence of any industry information), *costToProduce* will be set at £50 and *fixedCosts* to 60p in the following sections. These figures will be recalculated for application to the real data in Chapter 8.

7.3 Diffusion

Diffusion experiments provide a useful opportunity to examine the reaction of the system whilst adjusting individual parameters. This gives an understanding of how information is transmitted between stations. The experiments performed within this section use the hybrid model (see §6.6 for further details). §6.9 saw the development of a network model that could provide an alternative method of redistributing the population. However, as the data is idealised, there is no need to redistribute the population, therefore this option will not be used.

The use of the hybrid model introduces the parameters of *neighbourhood* size, *overprice*, *undercut* and *priceChange* (see §5.2.5; Table 5.1 for further details). Experiments will be carried out with each of these parameters investigating their influence and sensitivity.

For each of the experiments a 40km x 40km[3] grid was constructed with stations spaced at 2km intervals. The size of the grid is equivalent to a substantial metropolitan area, for example the Leeds-Bradford conurbation. The population per ward was 2000 with 1000 cars. β was fixed at 0.0003; λ: 0.7; *fixedCosts* at £50 and *costToProduce* at 60p. The optimal parameters derived for use within the pure agent model (see Chapter 5) were used unless otherwise stated. In summary, these values were *neighbourhood*: 5km; *undercut*: 0.5p; *overprice*: 1.5p and *priceChange*: 0.1p.

7.3.1 Experimentation with Price Drop

One of the simplest ways to test diffusion within the system is to drop the price of the centre station by varying amounts and examine if and how the cheaper price is transmitted through the network. It is hypothesised that the cheaper price at the centre station will diffuse through the system at different rates according to the amount that the centre station is dropped by.

The stations were all initialised at 68p, except the centre station where the price was dropped by 1p, 3p and 5p. In each case the initial price drop at the centre station was found to gradually spread out at an even rate to the surrounding area (Figure 7.6). The speed of the diffusion was found to be dependent on the size of the initial price drop. After 40 iterations, a 1p price drop (Figure 7.6 (a)) had not diffused through the system to the same degree as the diffusion seen in the 3p and 5p price drops. The price drop has diffused out to a greater extent in the 5p than the 3p, but in both cases, the cheaper price has affected all the system causing an overall lower price (indicated by the yellow colour).

Figure 7.7 shows the total price change as a function of distance from the centre station at various times for the 5p price drop. This illustrates that for a given time there is a relatively sharp interface between the region which has been strongly affected by the price drop and the more distant region where there is only a small drop in price. The rate at which this interface propagates through the system is relatively constant for any given price drop with a value of about 0.1km per day in the case of a 5p drop. For distances less than the interface the price change is more or less constant at about -6.5p. At greater distances the price change slowly drops during the simulation, but in this case never reaches less than about -3p before the interface propagates beyond the bounds of the grid. As a result of the non-local formulation of the spatial interaction model (i.e. all stations are affected to some degree by changes at the centre station), a drop in price at the centre station will affect all stations throughout the domain immediately. Obviously, the more distant edge stations are affected less than those close to the centre. This leads to the slow decrease in price at all stations. Those stations within the neighbourhood of the centre station are much more strongly affected because of the undercutting and overpricing rules. The sphere of influence of these stations slowly moves out and it is this that causes a sharp interface between the central region of low prices and the outer regions that are less affected by the price drop.

To objectively measure the position of the interface the data for each day is fitted to a hyperbolic tangent curve of the form $y(x) = a\tanh(bx - c) + d$. This shape was chosen since it closely fits the pattern of the data. The best fit curves are shown by the solid lines on Figure 7.7 and provide a good approximation to the data. The front position is taken as the midpoint of the tanh

[3]The hybrid model was not run within Excel and therefore did not have the same grid constraints as the spatial interaction model.

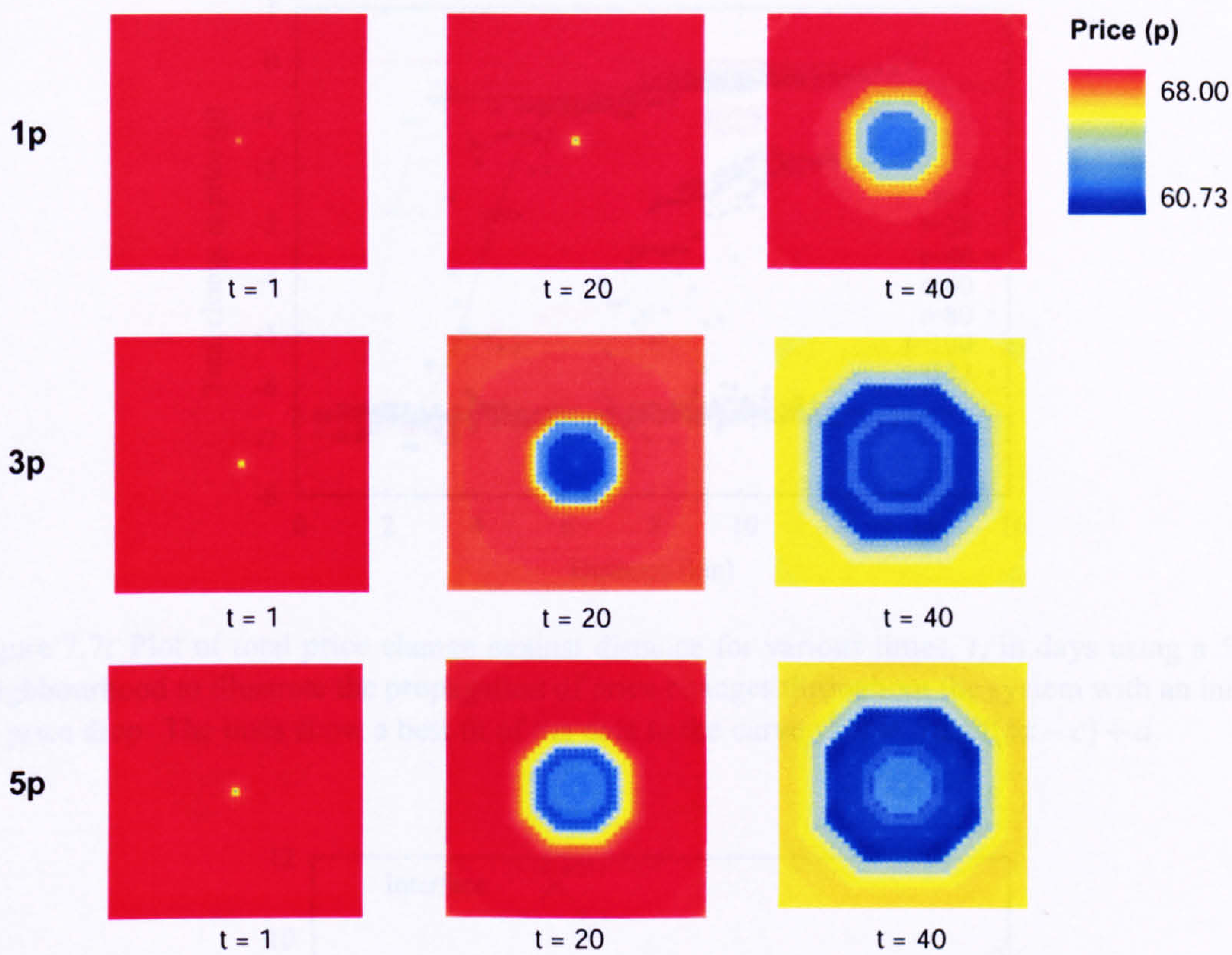


Figure 7.6: Spatial diffusion of price changes across the grid after different numbers of days, t, using a 1p, 3p and 5p price drop.

curve, c/b. Figure 7.8 show the position of the diffusion interface from the best fit curves plotted against time. The graph shows that the data is well modelled by a straight line. The linear regression line on the figure has a slope of 0.0989 corresponding to a speed for the propagation of the interface of approximately 0.1 km/day.

7.3.2 Experimentation with Influence of Neighbourhood

Increasing the price drop at the centre station has been proven to affect the rate of dispersion within the system. The size of the neighbourhood is also hypothesised to have an effect on the rate of information exchange between stations. The neighbourhood is the area within which the petrol station assess its competitors prices. It is taken as a circular area around the station with a given radius. The 40km x 40km grid was used with β and λ set at 0.0003 and 0.7; *fixedCosts*: £50 and *costToProduce* 60p . All the stations were initialised at 68p with the exception of the centre station which was set to 63p. The model was run with the *neighbourhood* parameter set to 3km, 5km and 10km[4].

Figure 7.9 shows that the larger the neighbourhood size, the faster the diffusion of the price drop. The diffusion patterns within each of the simulations are similar. The rate at which the price

[4] 1km was not used because the spacing of the petrol stations was 2km. A 1km neighbourhood would therefore have no impact.

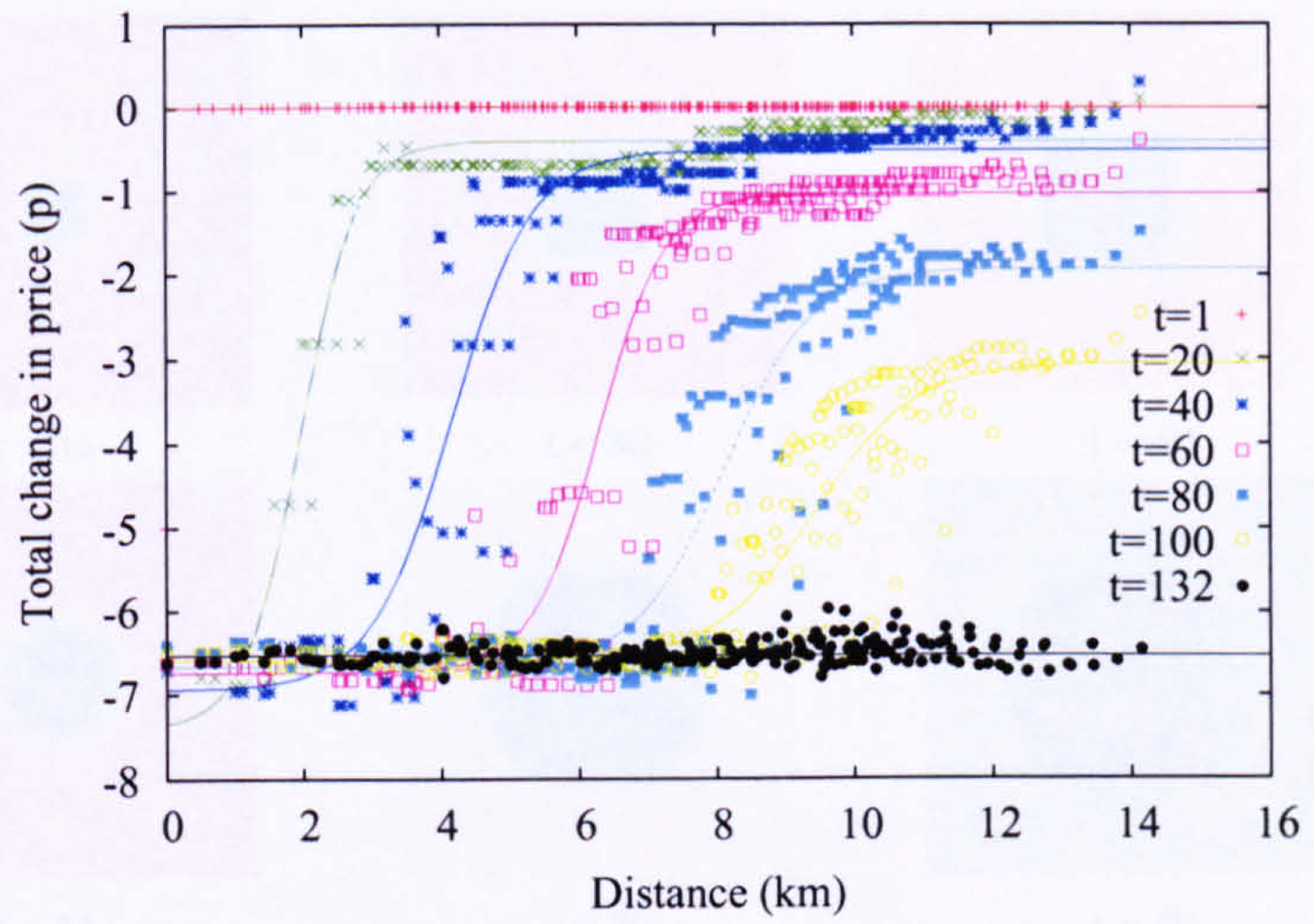


Figure 7.7: Plot of total price change against distance for various times, t, in days using a 5km neighbourhood to illustrate the propagation of price changes throughout the system with an initial 5p price drop. The lines show a best fit of the data to the curve $y(x) = a\tanh(bx - c) + d$.

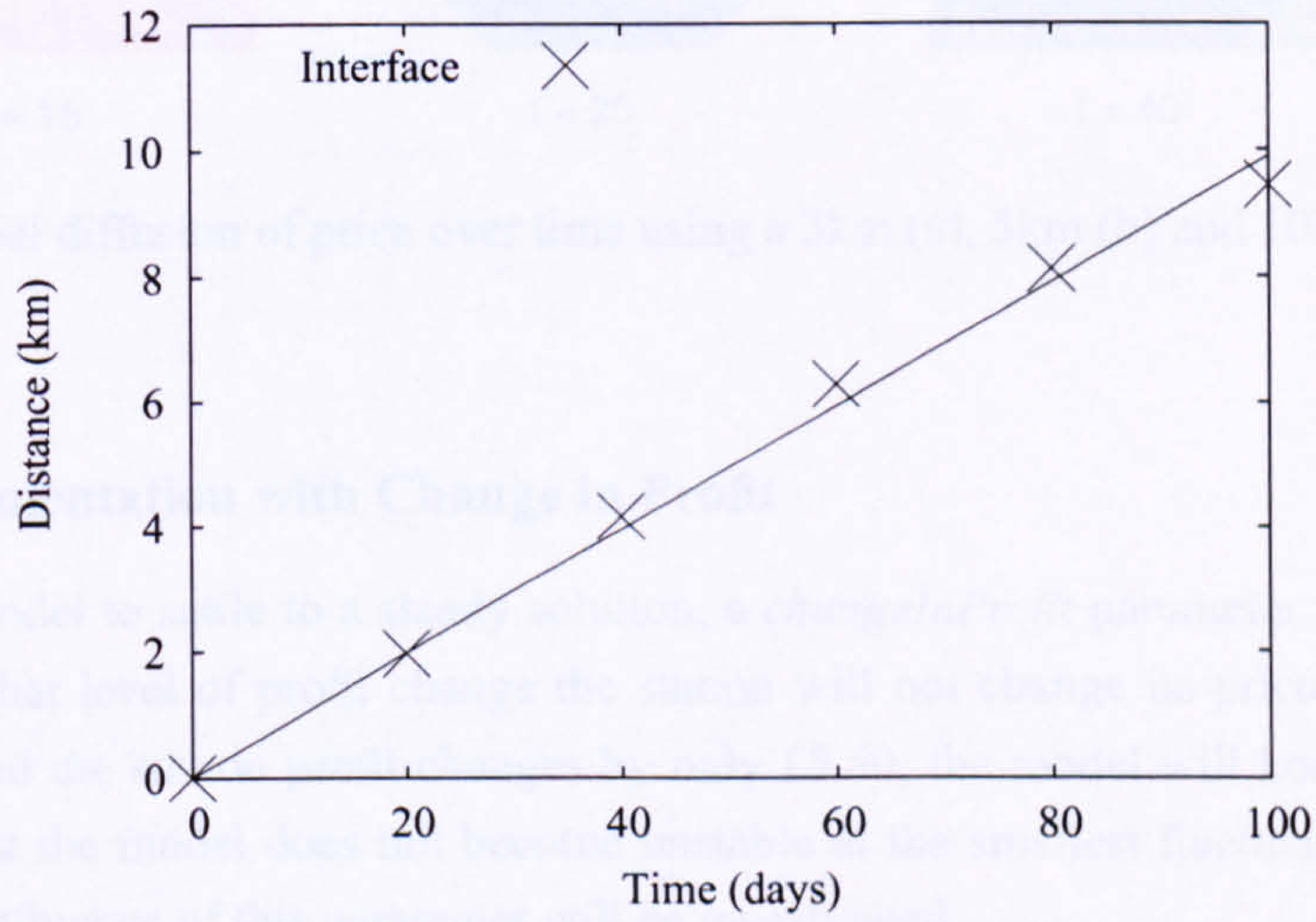


Figure 7.8: Plot of interface position (in km) against time (in days) for the simulation show in Figure 7.7. The solid line is a linear best fit to the data with the equation $y(x) = 0.0989x$.

drop diffuses out is proportional to the size of the neighbourhood. This can be seen more clearly in Figure 7.10 where the 10km neighbourhood precipitates a faster and spatially greater diffusion than either 3km or 5km (the stair effect is a result of the discretisation of the grid). This increase in diffusion rate is due to the greater number of stations within the larger neighbourhoods that can react to changes in petrol prices.

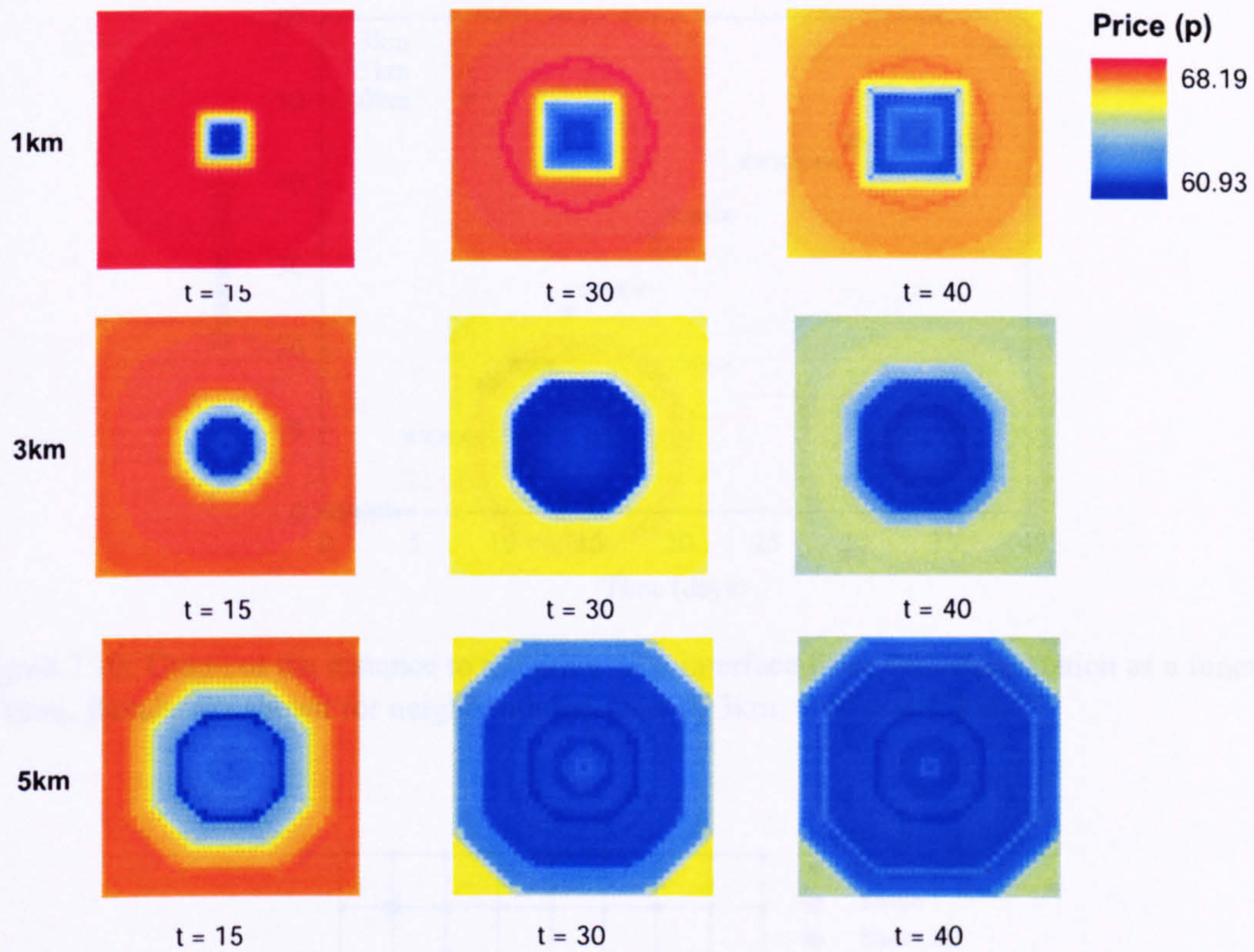


Figure 7.9: Spatial diffusion of price over time using a 3km (a), 5km (b) and 10km (c) neighbourhood.

7.3.3 Experimentation with Change in Profit

To enable the model to settle to a steady solution, a *changeInProfit* parameter was created. This determines at what level of profit change the station will not change its price. For example, if it is set to £6 and the station profit changes by only £5.50, the model will keep the price fixed. This ensures that the model does not become unstable at the smallest fluctuation in profit. The sensitivity and influence of this parameter will be investigated.

The experiments were performed using a 40 x 40km grid with β and λ set to 0.0003 and 0.7, *fixedCosts* £50 and *costToProduce* 60p. All the stations were initialised at 68p with the exception of the centre station which was set to 63p. The price was dropped to create a situation where the system had to adjust. This enabled the *changeInProfit* parameter to be assessed. To understand behaviour across the system, the price at several stations will be examined as shown in Figure 7.11.

With *changeInProfit* values of £1 to £15 (Figure 7.12(a) - (d)), the effect upon the system is not entirely predictable. With a value of £1 (Figure 7.12(a)), the corner station (1) is cycling between prices (60p - 63p). Increasing the *changeInProfit* to £5, £10 and £15 precipitates an increase in price at this station, while the other stations decrease in price before reaching equilibrium. The price at which the other stations hit equilibrium varies and tends to increase with increased *changeInProfit* values, particularly at stations 211 and 421 which are the nearest stations to the edges. These results show that the system is very sensitive to the value of the *changeInProfit*.

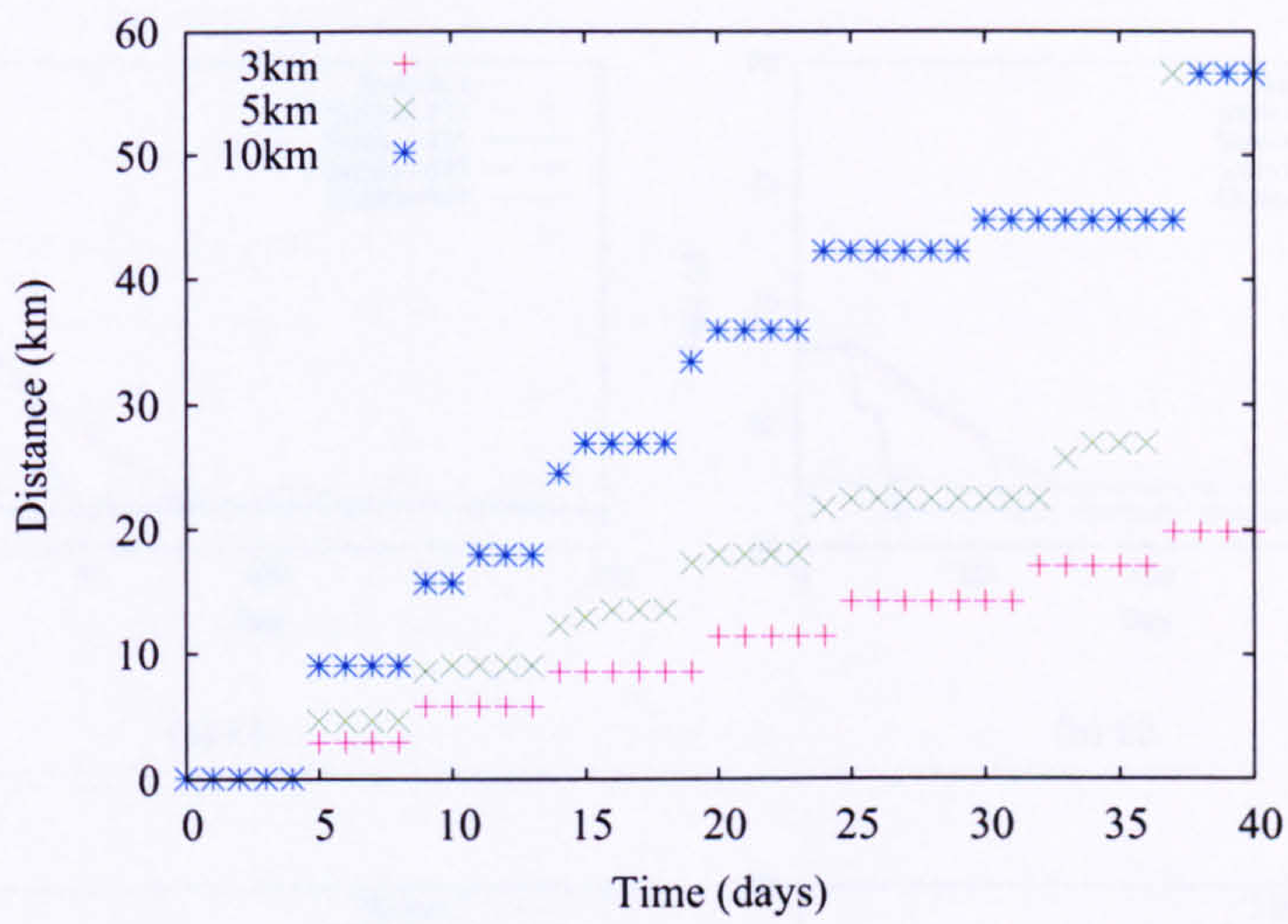


Figure 7.10: Graph of the distance to the price drop interface from the centre station as a function of time. Results are shown for neighbourhood sizes of 3km, 5km and 10km.

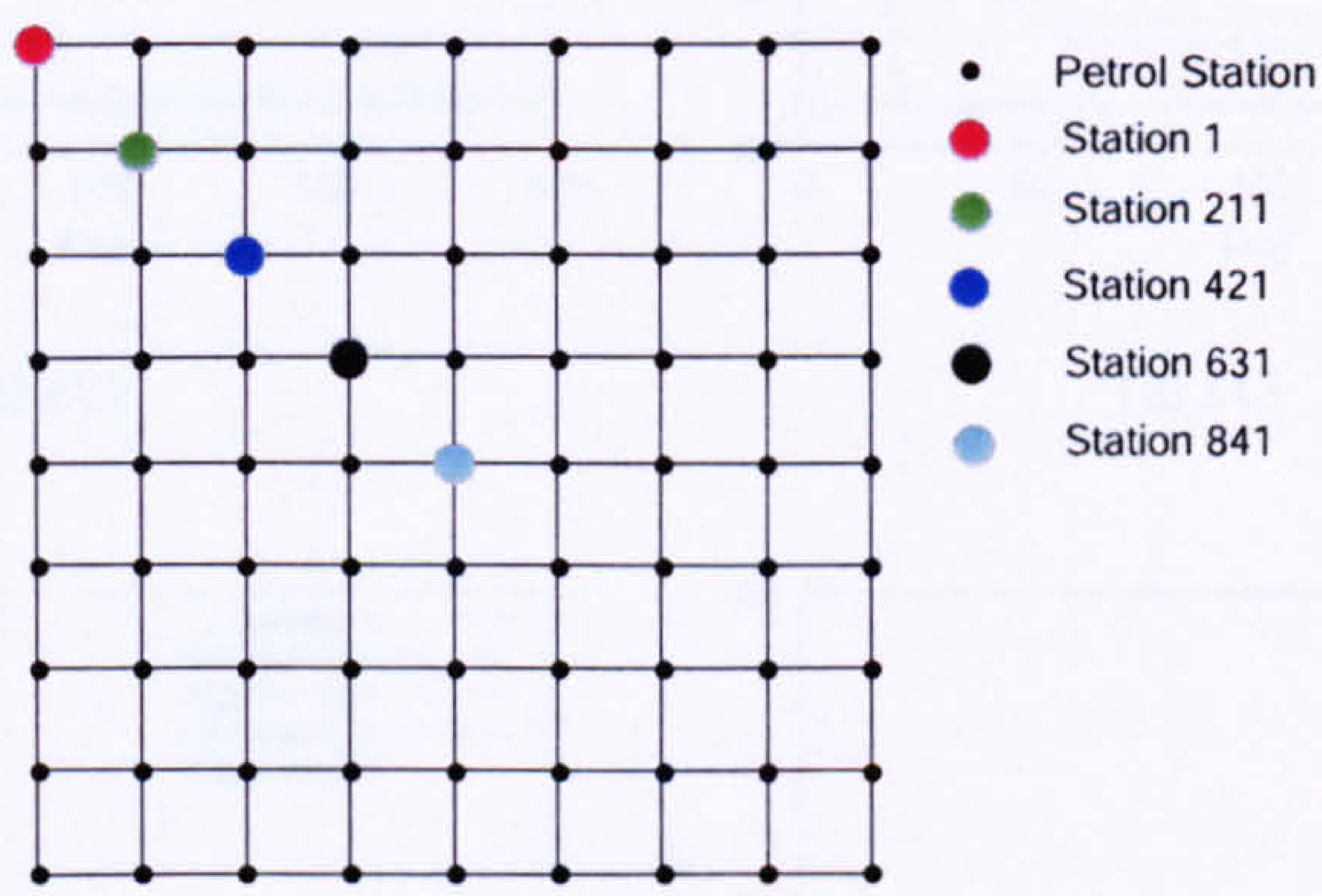
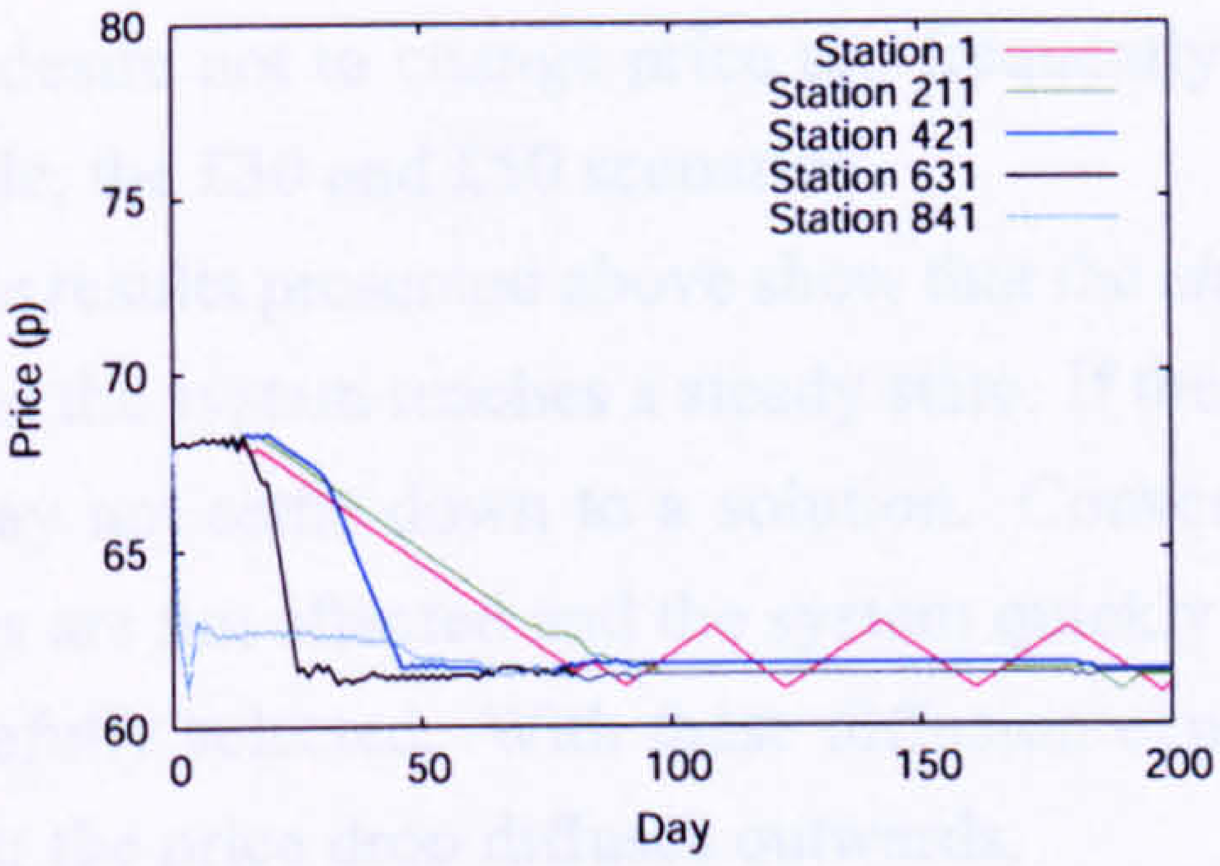


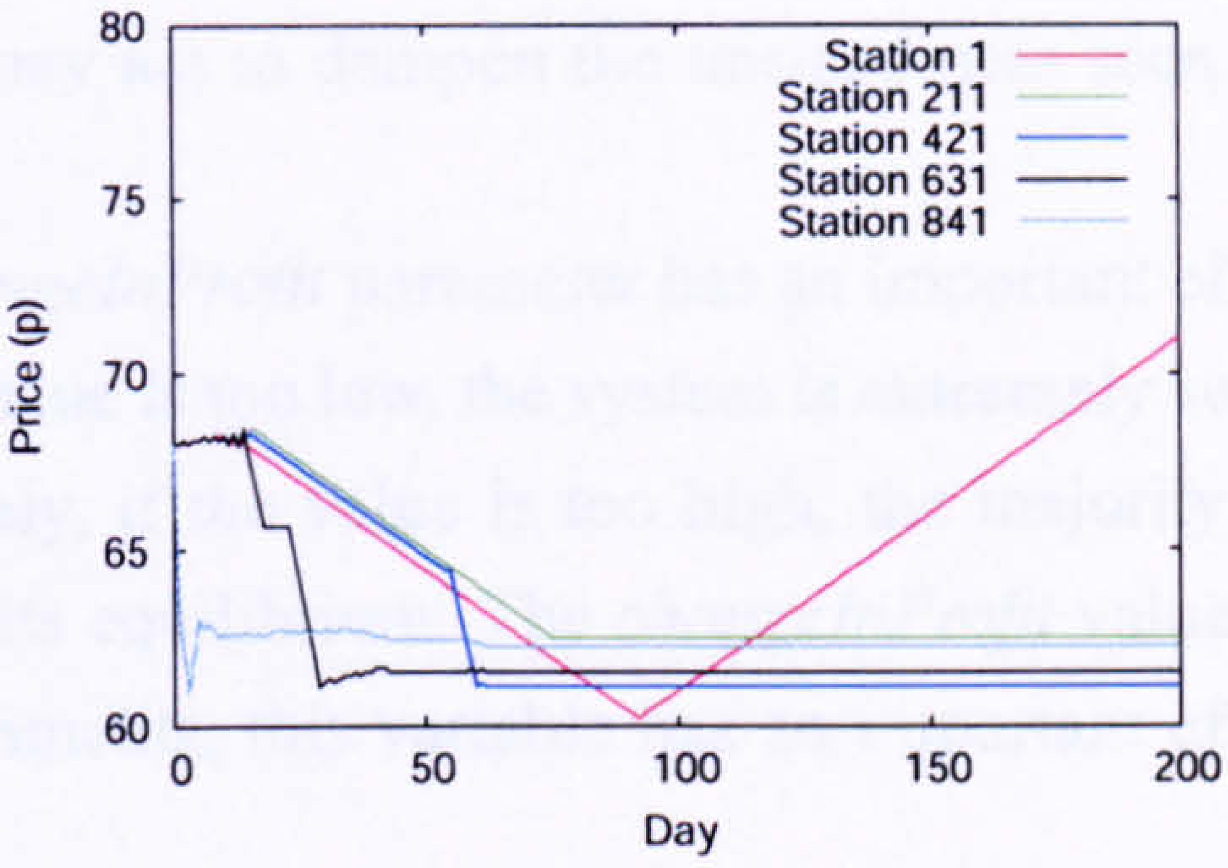
Figure 7.11: Illustration of the idealised grid showing the relative positions of the stations used in the analysis.

Increasing the *changeInProfit* variable to £20 - £25 results in most of the station prices decreasing and reaching an equilibrium after approximately 25 days (Figure 7.12(e), (f)). The exception to this is station 841 (the centre station) which fluctuates before hitting a steady solution. Increasing the *changeInProfit* further to £30 - £50 (Figure 7.12(g), (h)) results in all the stations remaining constant with the exception of station 841 (the station that has dropped its price). This station oscillates in price to in an attempt to increase its profit. The other stations are changing their price (by small amounts) in reaction to the fluctuations in price of 841. A *changeInProfit* variable of £30 - £50 only affects station 841, it is simply too large to affect any of the other stations. This type of reaction is similar to the behaviour observed within other chaotic systems, for example predator-prey systems (see Lokta, 1910; May, 1974, 1976, for further details).

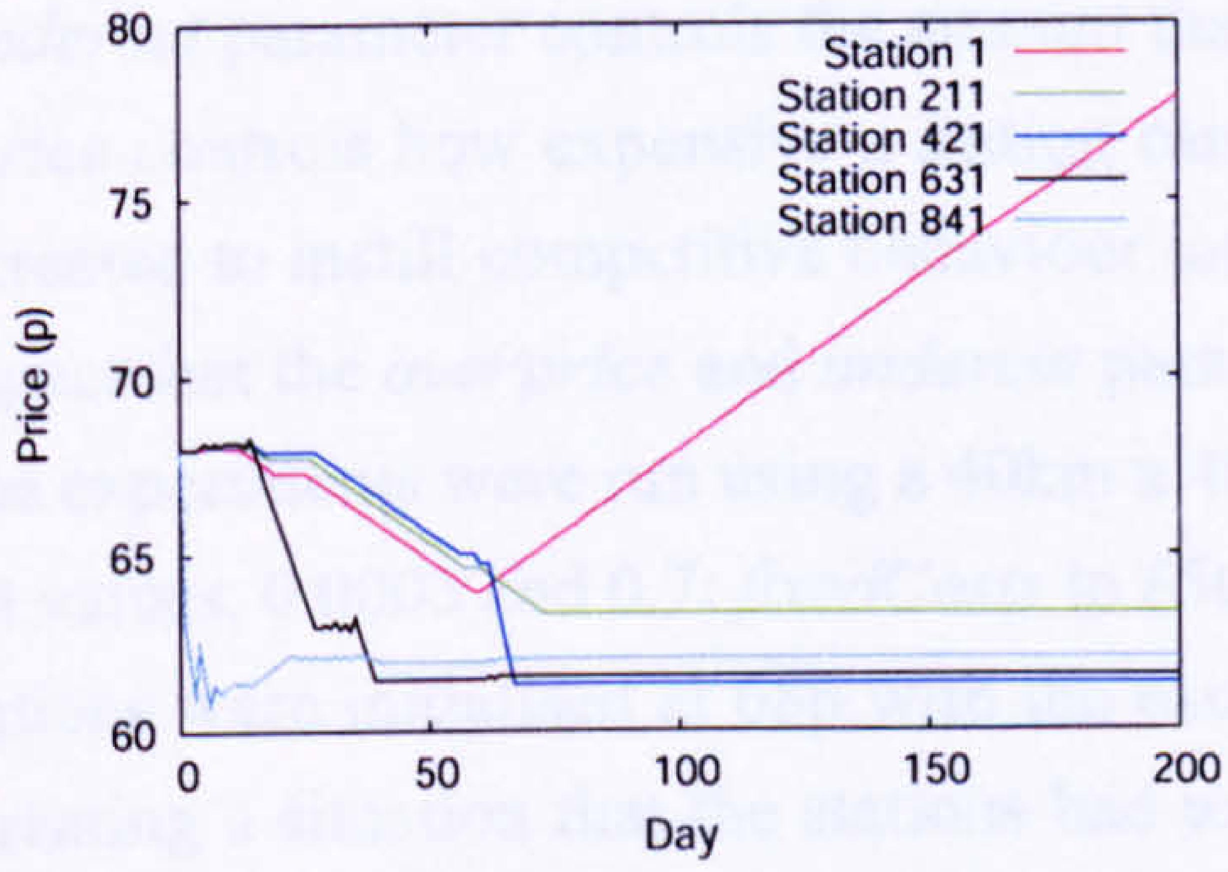
Within the model, the *changeInProfit* variable controls the pricing strategy. In reality, it may be a combination of factors that determines pricing. This combination could include subjective



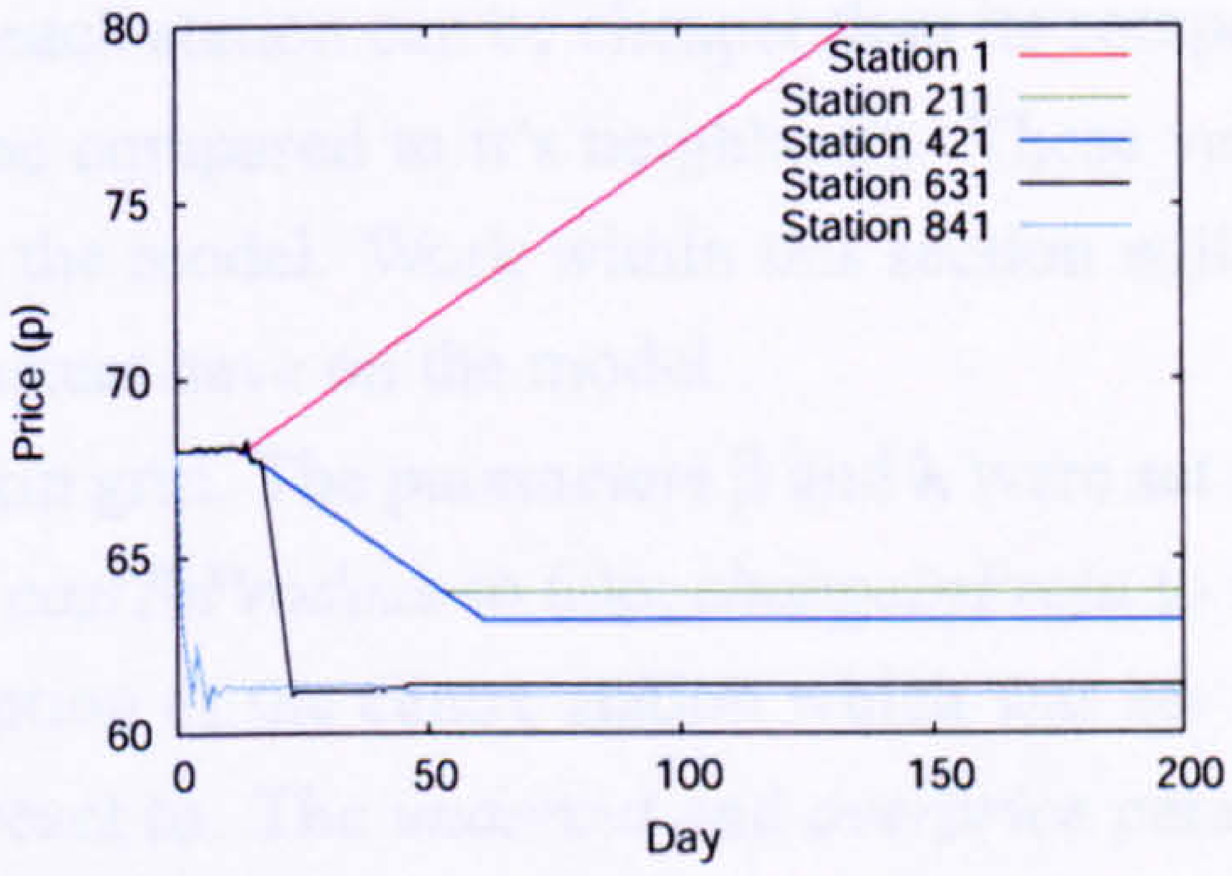
(a) £1



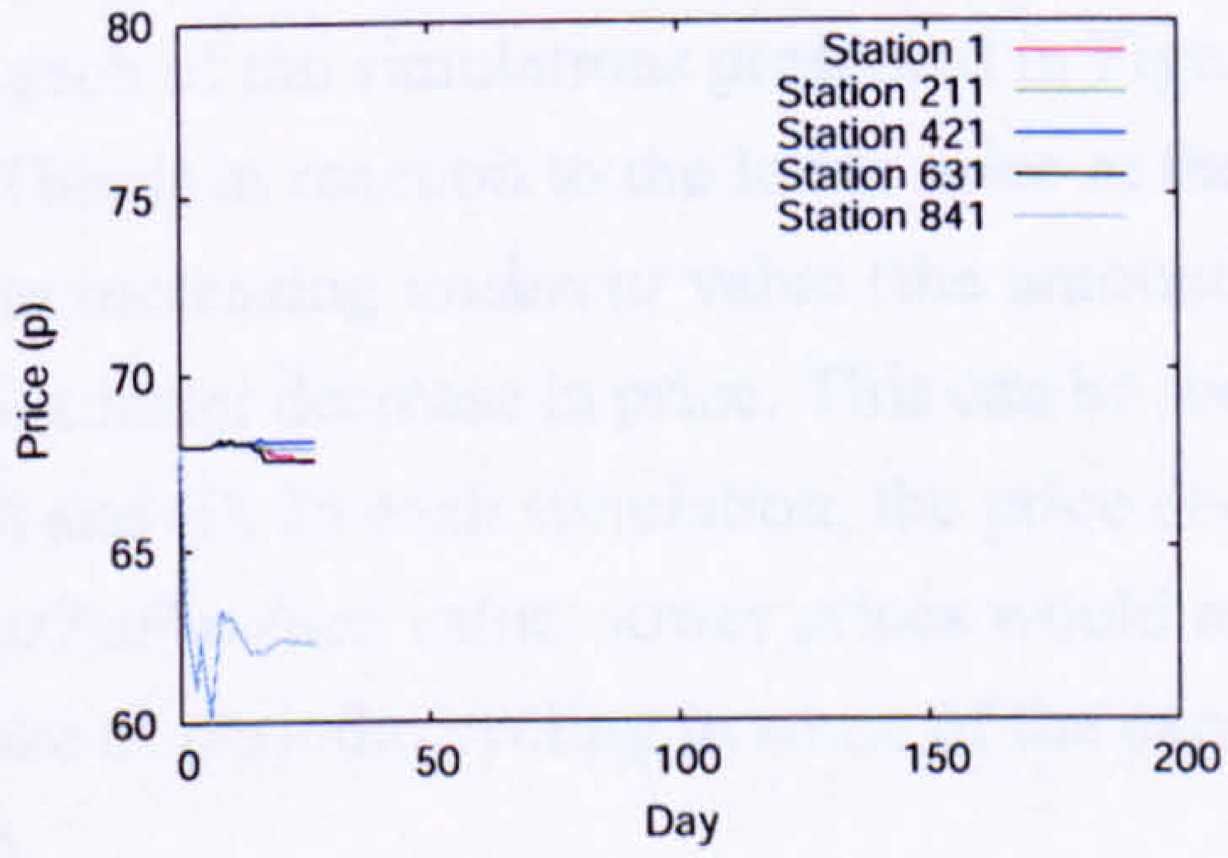
(b) £5



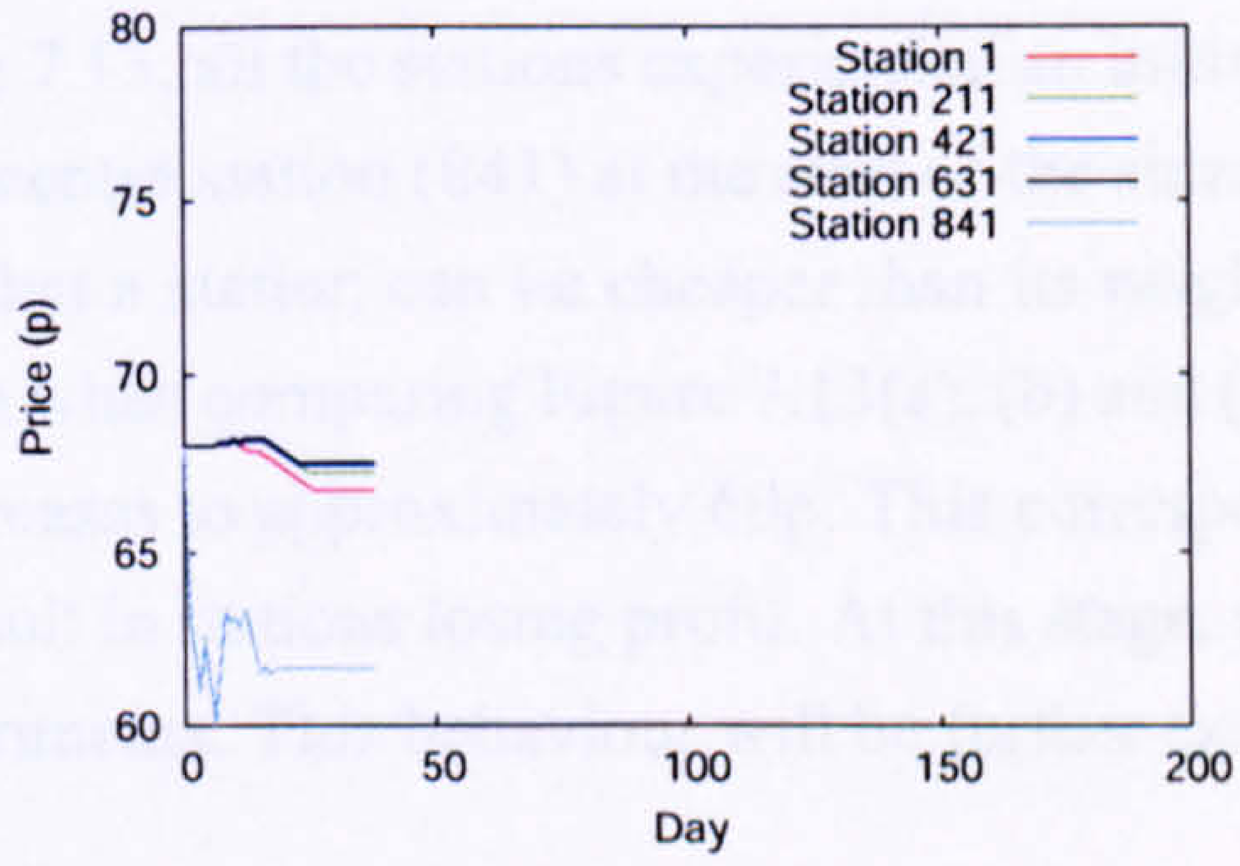
(c) £10



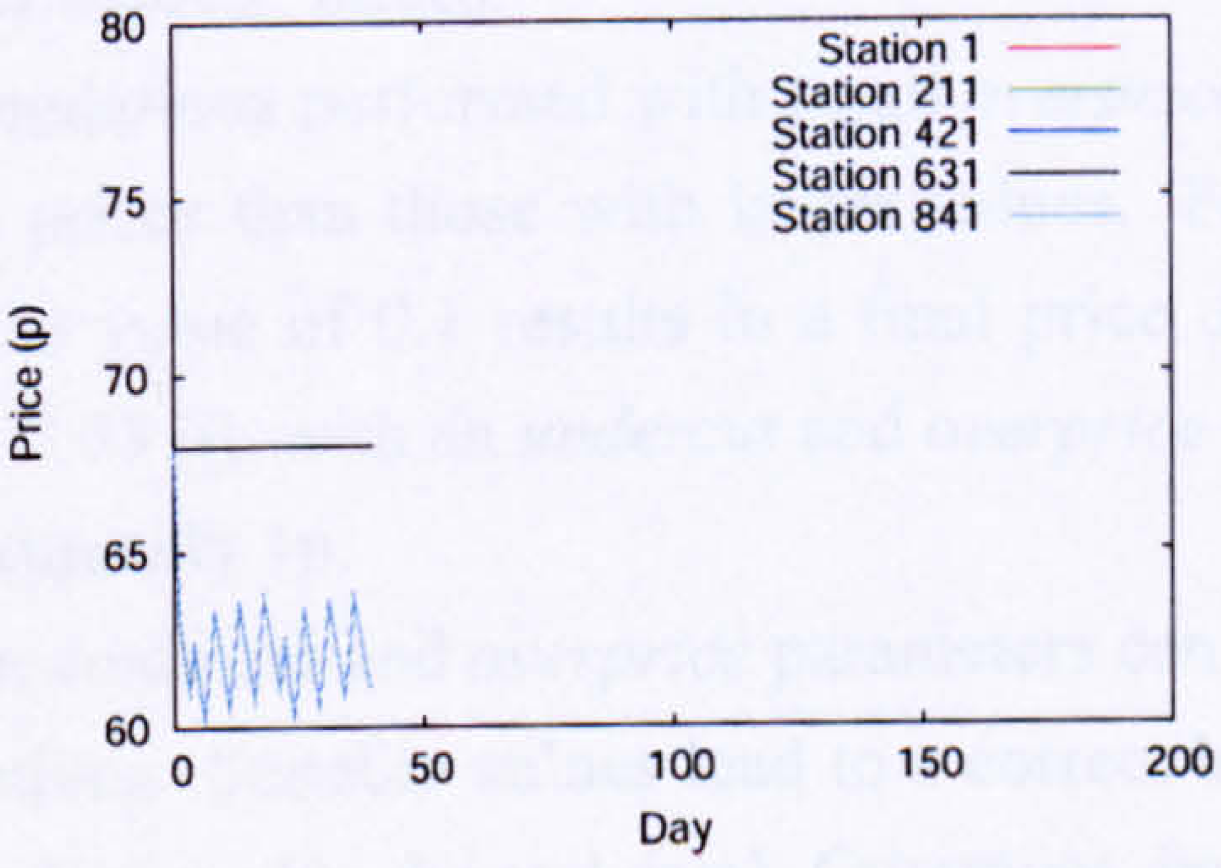
(d) £15



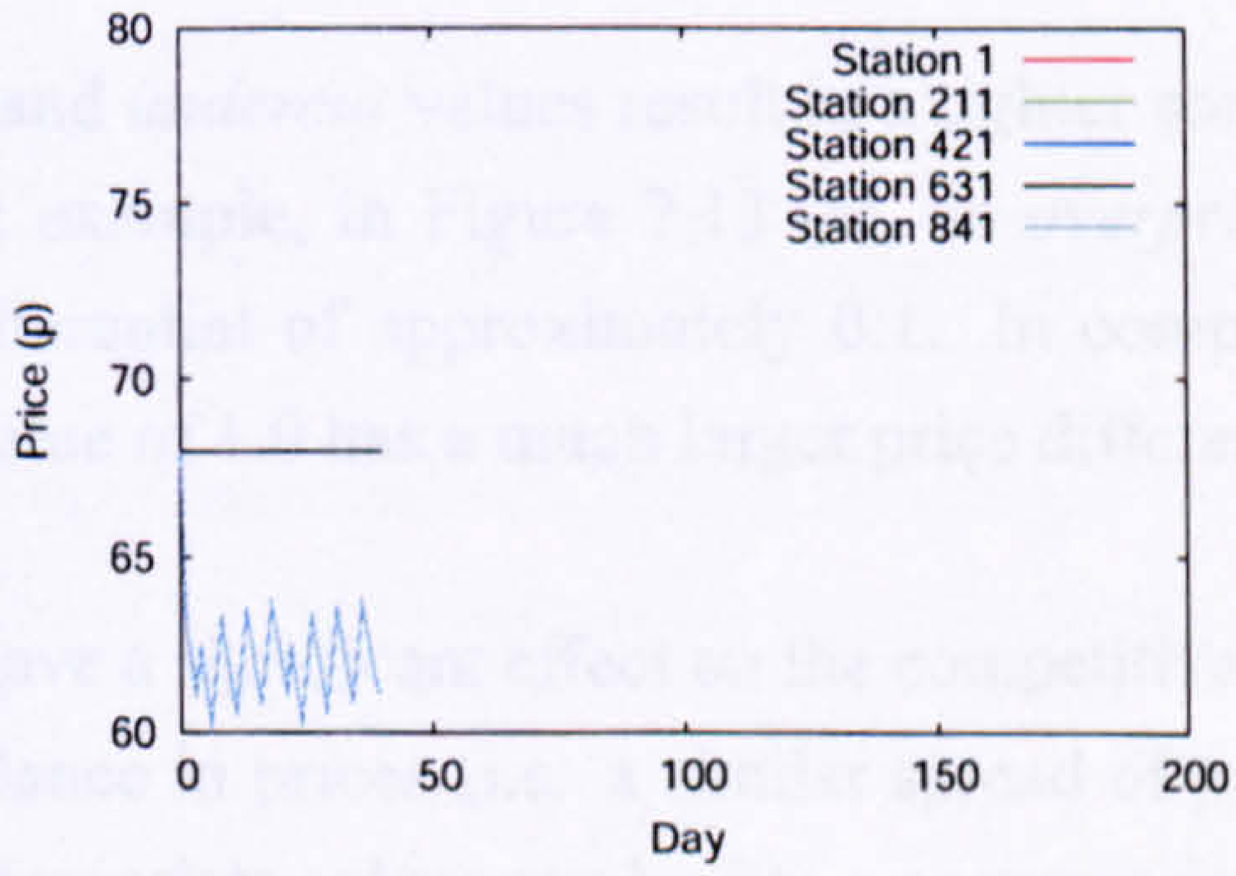
(e) £20



(f) £25



(g) £30



(h) £50

Figure 7.12: Plot of price against time for various stations on the idealised grid. Each graph corresponds to a different value of the *changeInProfit* parameter.

elements such as the perceived effect of price changes on the consumers. Human factors, such as the desire not to change price too frequently may act to dampen the unsteadiness seen in, for example, the £30 and £50 scenarios.

The results presented above show that the *changeInProfit* parameter has an important effect on whether the system reaches a steady state. If the value is too low, the system is extremely sensitive and may not settle down to a solution. Conversely, if the value is too high, the majority of the stations are not affected and the system quickly hits equilibrium. The *changeInProfit* value has to be carefully selected. With these diffusion experiments, this variable has an important effect on how far the price drop diffuses outwards.

7.3.4 Experimentation with Undercutting and Overprice

The *undercut* parameter controls the amount that each station can be cheaper than its competitors. *Overprice* controls how expensive a station can be compared to it's neighbours. These variables were created to instill competitive behaviour into the model. Work within this section will assess the impact that the *overprice* and *undercut* parameters have on the model.

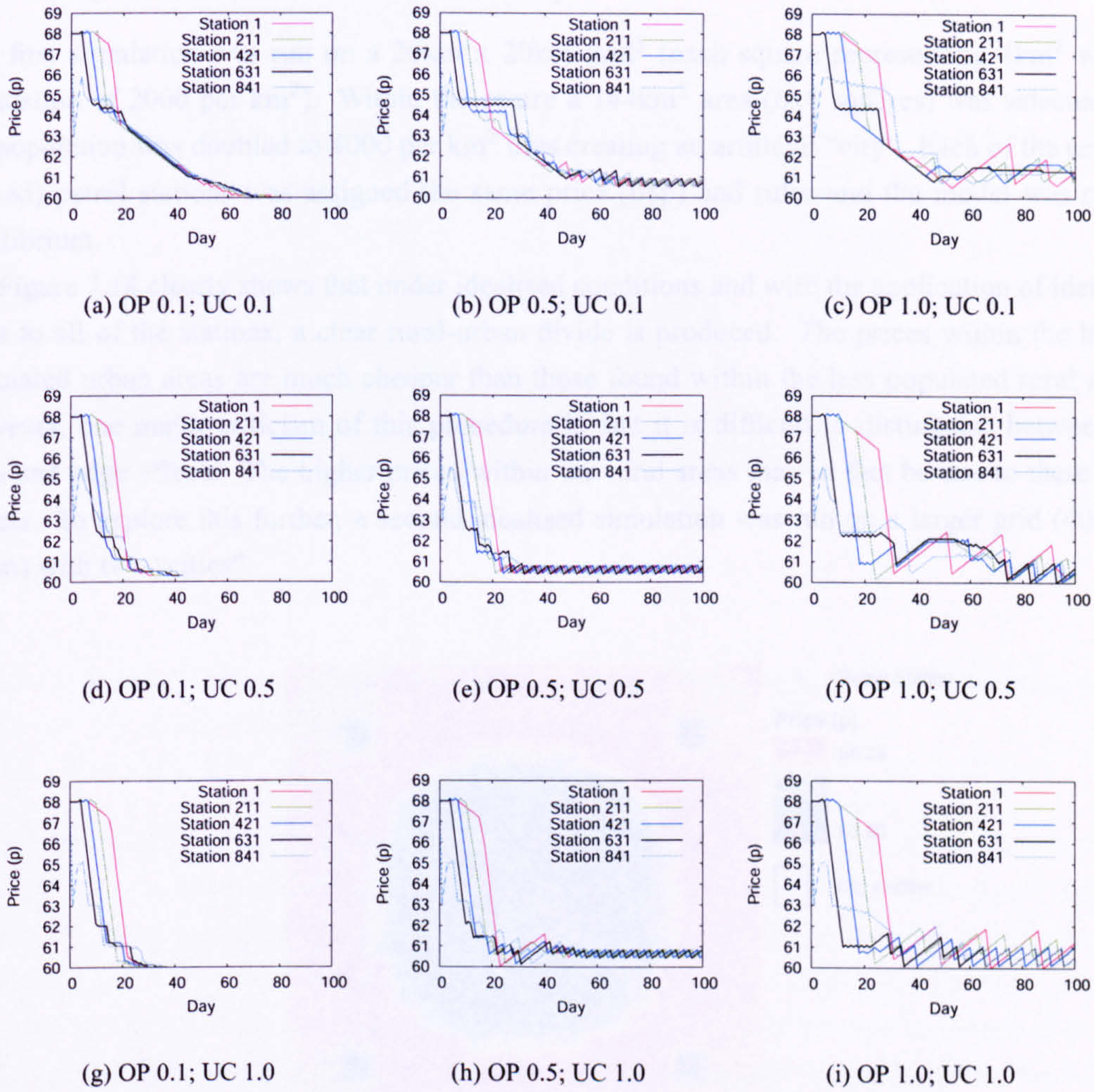
The experiments were run using a 40km x 40km grid. The parameters β and λ were set at their default values, 0.0003 and 0.7; *fixedCosts* to £50; *costToProduce* to 60p; *changeInProfit* to £2. All the stations were initialised at 68p with the exception of the centre station which was set to 63p, thus creating a situation that the stations had to react to. The *undercut* and *overprice* parameters were independently varied with values of 0.1, 0.5 and 1.0p (values selected after analysis of the real data). The impact of these varying values were assessed at several stations across the grid as shown in Figure 7.13.

In each of the simulations presented in Figure 7.13, all the stations experienced an initial price drop. This is in reaction to the lower price at the centre station (841) at the start of the simulation. With an increasing *undercut* value (the amount that a station can be cheaper than its neighbour), there is a faster decrease in price. This can be seen when comparing Figure 7.13(a), (b) and (c) with (g), (h) and (i). In each simulation, the price decreases to approximately 60p. This corresponds to the *costToProduce* value: lower prices would result in stations losing profit. At this stage, there is evidence of periodic cycling in some of the experiments. This behaviour will be further examined in §7.5.

As the *overprice* variable increases, for example in figures (a), (b) and (c), the mean price at which the model settles out increases, although the model does not reach equilibrium for these larger *overprice* values.

Simulations performed with small *overprice* and *undercut* values result in a tighter constraint on the prices than those with larger values. For example, in Figure 7.13 (a), an *overprice* and *undercut* value of 0.1 results in a final price differential of approximately 0.1. In comparison, Figure 7.13 (i), with an *undercut* and *overprice* value of 1.0 has a much larger price differential of approximately 1p.

The *undercut* and *overprice* parameters can have a significant effect on the competitiveness of the stations. Suitable values lead to a correct balance in prices (i.e. a similar spread of prices to those found within the real data). Otherwise, inappropriate values can lead to a narrow price range with prices becoming either too high or low.



(a) OP 0.1; UC 0.1 (b) OP 0.5; UC 0.1 (c) OP 1.0; UC 0.1

(d) OP 0.1; UC 0.5 (e) OP 0.5; UC 0.5 (f) OP 1.0; UC 0.5

(g) OP 0.1; UC 1.0 (h) OP 0.5; UC 1.0 (i) OP 1.0; UC 1.0

Figure 7.13: Time series of the fuel price at the analysis stations for various combinations of the *undercut* (UC) and *overprice* (OP) parameters.

7.4 Simulations Introducing Geographical and Competitive Variations

Examination of the real data within Chapter 4 provided evidence of a rural - urban price difference. Within rural areas, the prices were generally more expensive than those of petrol stations located in urban areas. This is hypothesised to be related to the density (and therefore level of competition) of petrol stations within an area as well as the differing population densities across an area. Work within this section will focus on testing out these hypothesises.

Unless otherwise stated, the following rules were used: β: 0.0003; λ: 0.7; *fixedCosts*: £50; *costToProduce*: 60p; *neighbourhood* 5km; *undercut* 0.5p; *overprice* 1.5p and *priceChange* 0.1p.

7.4.1 Experimentation with Different Population Densities

The first simulation was run on a 20km x 20km grid[5] (each square representing 4km^2 with a population of 2000 per km^2). Within the centre a 144km^2 area (6x6 squares) was selected and the population was doubled to 4000 per km^2 thus creating an artificial "city". Each of the (evenly spaced) petrol stations was assigned the same price (68p) and rules and the model was run to equilibrium.

Figure 7.14 clearly shows that under idealised conditions and with the application of identical rules to all of the stations, a clear rural-urban divide is produced. The prices within the highly populated urban areas are much cheaper than those found within the less populated rural areas. However, one major criticism of this procedure is that it is difficult to distinguish between the rural and edge effects. The higher prices within the rural areas may in fact be due to these edge effects. To explore this further, a second idealised simulation was run on a larger grid (40km x 20km) with two "cities".

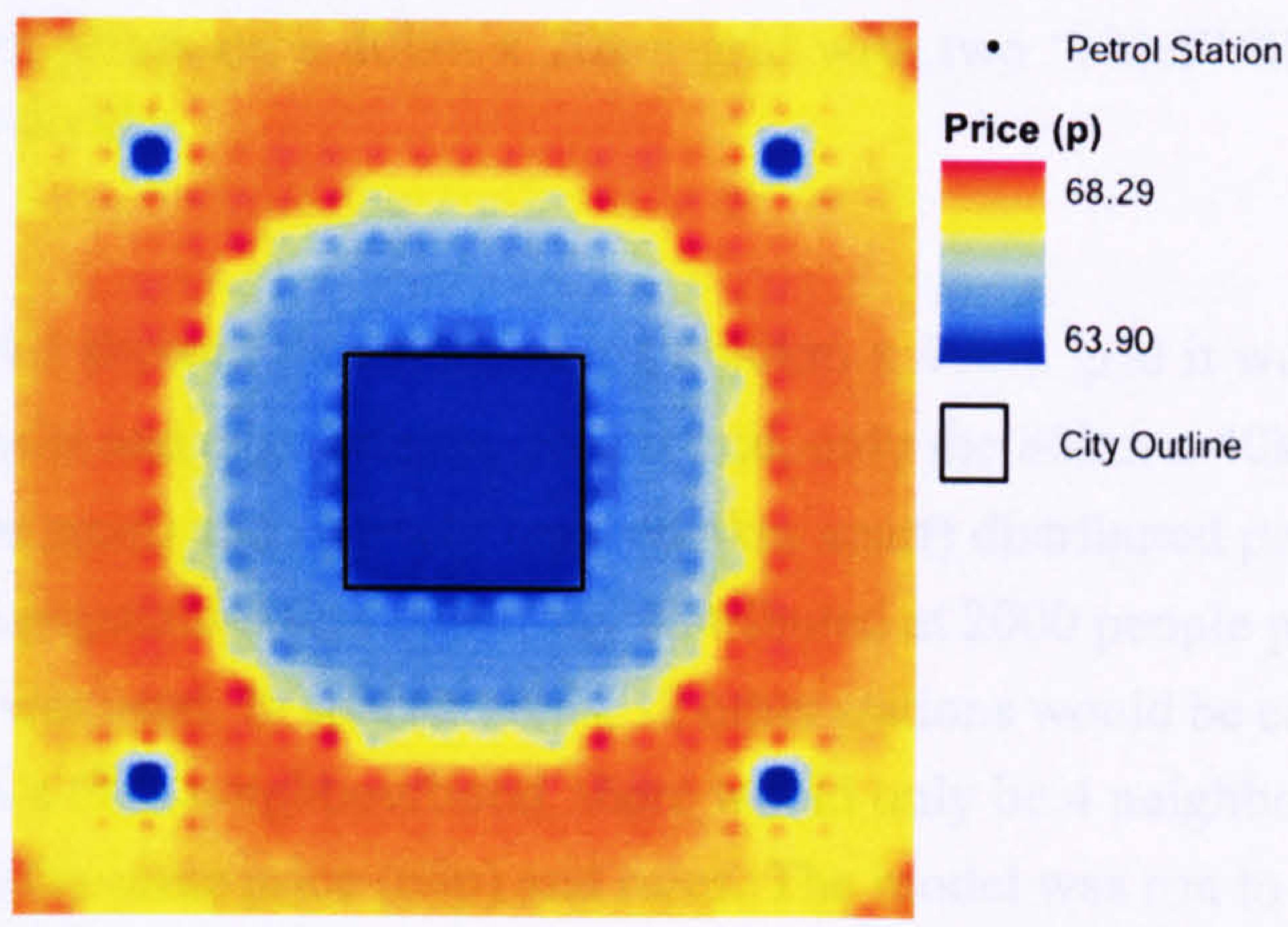


Figure 7.14: Map of petrol prices on a 20km x 20km grid with one "city" illustrating the re-creation of the rural-urban divide.

The patterns in Figure 7.15 are similar to those seen in Figure 7.14. However, the rural region situated between the two cities has a higher price than the centre of either city. As this region is also situated away from the edge of the grid, it can be confidently stated that the urban-rural divide is a result of the different population densities rather than the edge effects.

7.4.2 Experimentation with Different Petrol Station Distributions

What is the effect of varying the density and distribution of stations on the system? It is hypothesised that areas of high density stations (akin to urban areas) will sustain lower prices than the sparsely distributed petrol stations (rural areas).

[5]The size of the grid used was limited by the amount of memory available on the PC.

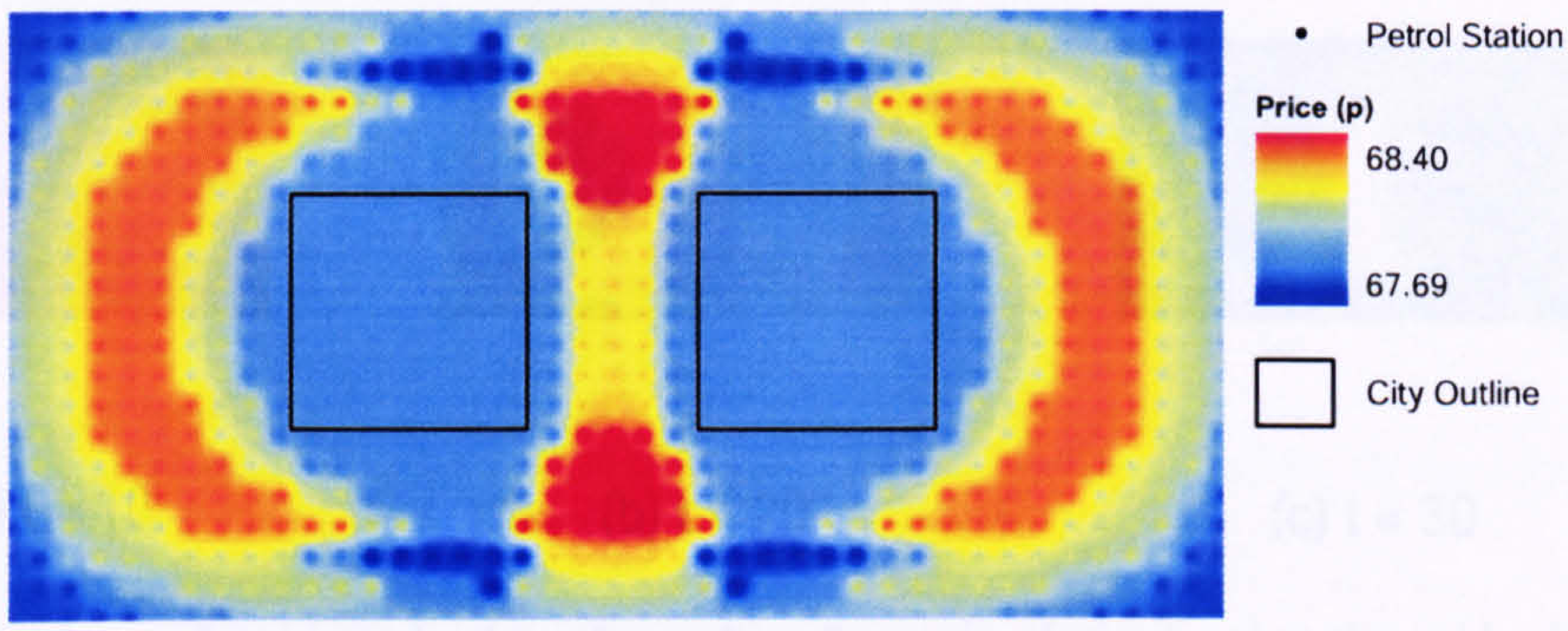


Figure 7.15: Map of petrol prices on a 40km x 20km grid with two "cities" illustrating the re-creation of the rural-urban divide.

The results within §7.4.1 showed that when using a 40km x 40km grid it was hard to distinguish between genuine results and edge effects. To this end, only the 80km x 40km grid with two areas of dense (spaced 2km apart) and sparsely (spaced 4km apart) distributed petrol stations will be used (Figure 7.16). The population was uniformly distributed at 2000 people per km^2. An 8km *neighbourhood* parameter was used. In sparser areas, 12 other stations would be contained in a stations neighbourhood. With a 5km neighbourhood, there would only be 4 neighbours. Each petrol station was initialised with the same price (68p) and rules. The model was run to equilibrium.

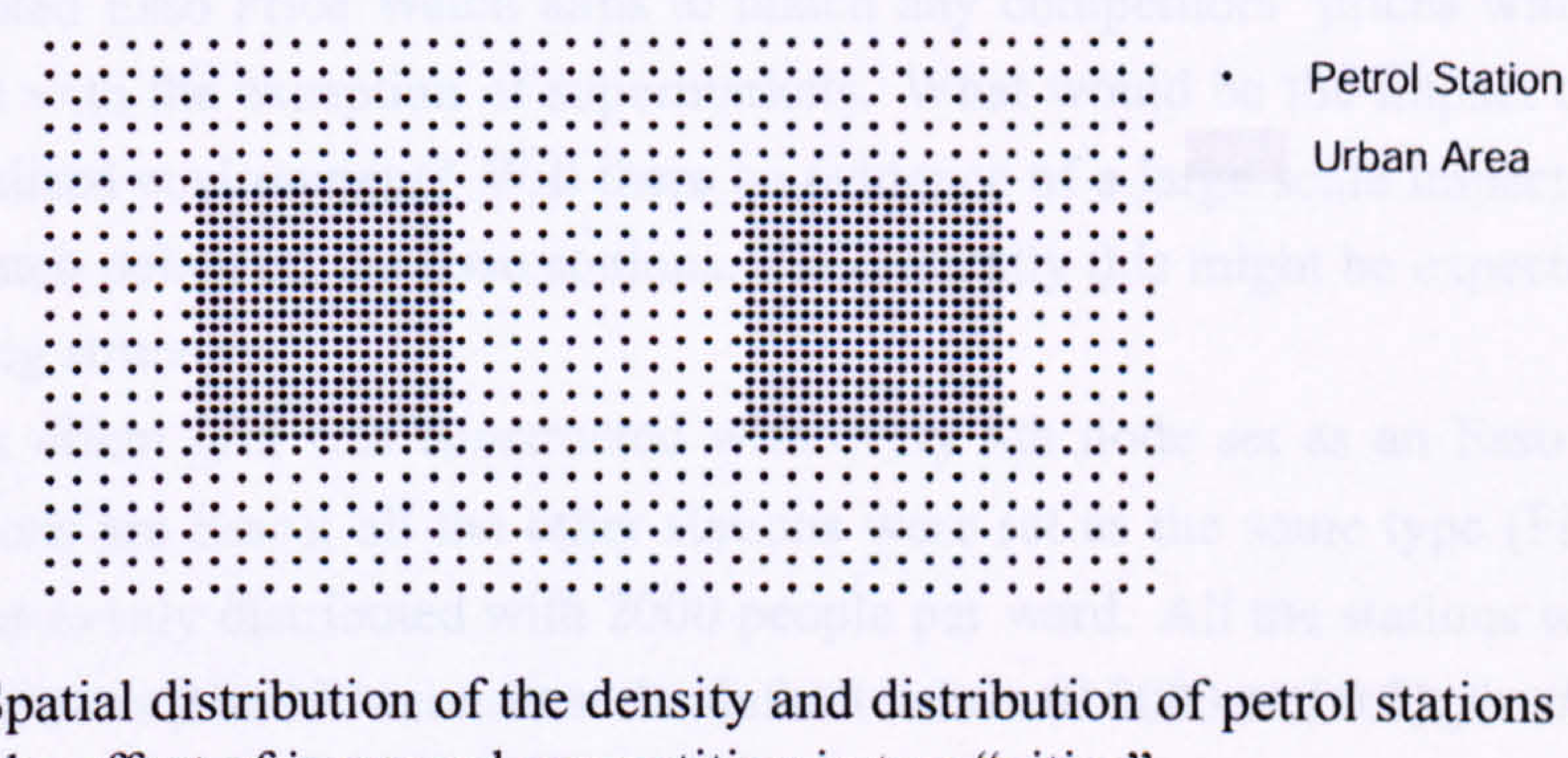


Figure 7.16: Spatial distribution of the density and distribution of petrol stations on the grid used to investigate the effect of increased competition in two "cities".

The results of the simulation (Figure 7.17) clearly shows a price divide between the sparsely and densely populated areas of stations. The prices within the sparse (rural) areas are much more expensive than those found within the denser (urban) areas. The region of sparse stations between the two dense areas has a higher price than the centre of either of the denser areas. As this situation is away from the edge of the grid, this suggests that the price divide is a result of the different petrol station densities rather than the edge effects. The increased competition in those areas with a higher density of stations leads to a decrease in price in order to increase sales.

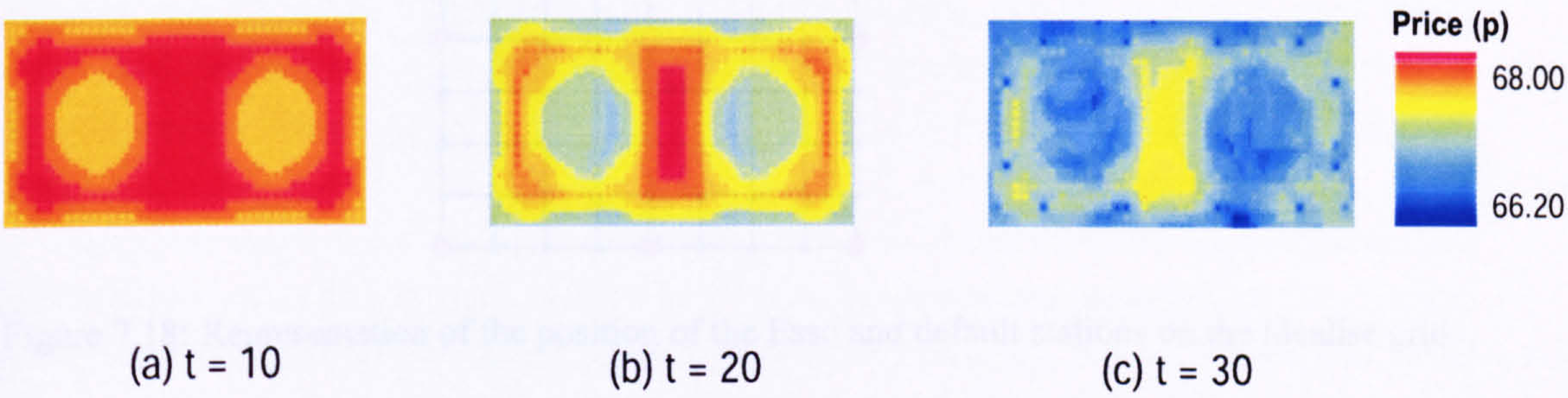


Figure 7.17: Maps of the petrol price after t days from simulations using the grid with varying densities of petrol stations.

7.4.3 Experimentation with Different Petrol Station Types

Varying population and station distributions influence the behaviour of the system. However, this work has been undertaken with the assumption that all the stations operate the same pricing strategy. This does not occur in the real system. Esso and supermarkets have been identified as operating distinct rules and have been previously experimented with (Chapter 5). The pricing behaviour of both these stations are different and the effects can be easily understood using idealised conditions.

Esso

The often quoted Esso Price Watch aims to match any competitors' prices within a distance of 3 miles (5km) with the exception of supermarkets. What would be the impact of this behaviour within an idealised environment? Will there be evidence of a large scale impact of imposing the Esso Price Watch policy on the Esso stations. Theoretically this might be expected to give rise to a tighter pricing structure.

A 40km x 40km grid was constructed with every 4th node set as an Esso station (i.e 1 in every 16 stations are Esso); all the other stations were set as the same type (Figure 7.18). The population was evenly distributed with 2000 people per ward. All the stations were initialised at 68p. The parameters β and λ were set at the default values (0.0003 and 0.7); *fixedCosts* set to £50; *costToProduce* 60p; *changeInProfit* £2.

The default stations were given a 10km *neighbourhood* and 1p value for *undercut*, *overprice* and *priceChange* rules. The Esso stations were given a 5km neighbourhood (anything smaller than this would only allow them to interact with its immediate neighbours) and a 1p *undercut* and *priceChange*. The significant difference was that the Esso stations were given a 0p *overprice* value, i.e. they were forced to match the prices of all other stations within their neighbourhood.

Figure 7.19 (a) shows that initially all the Esso stations located within the centre of the grid are matching the prices of those around them. Only the stations towards the edge of the grid

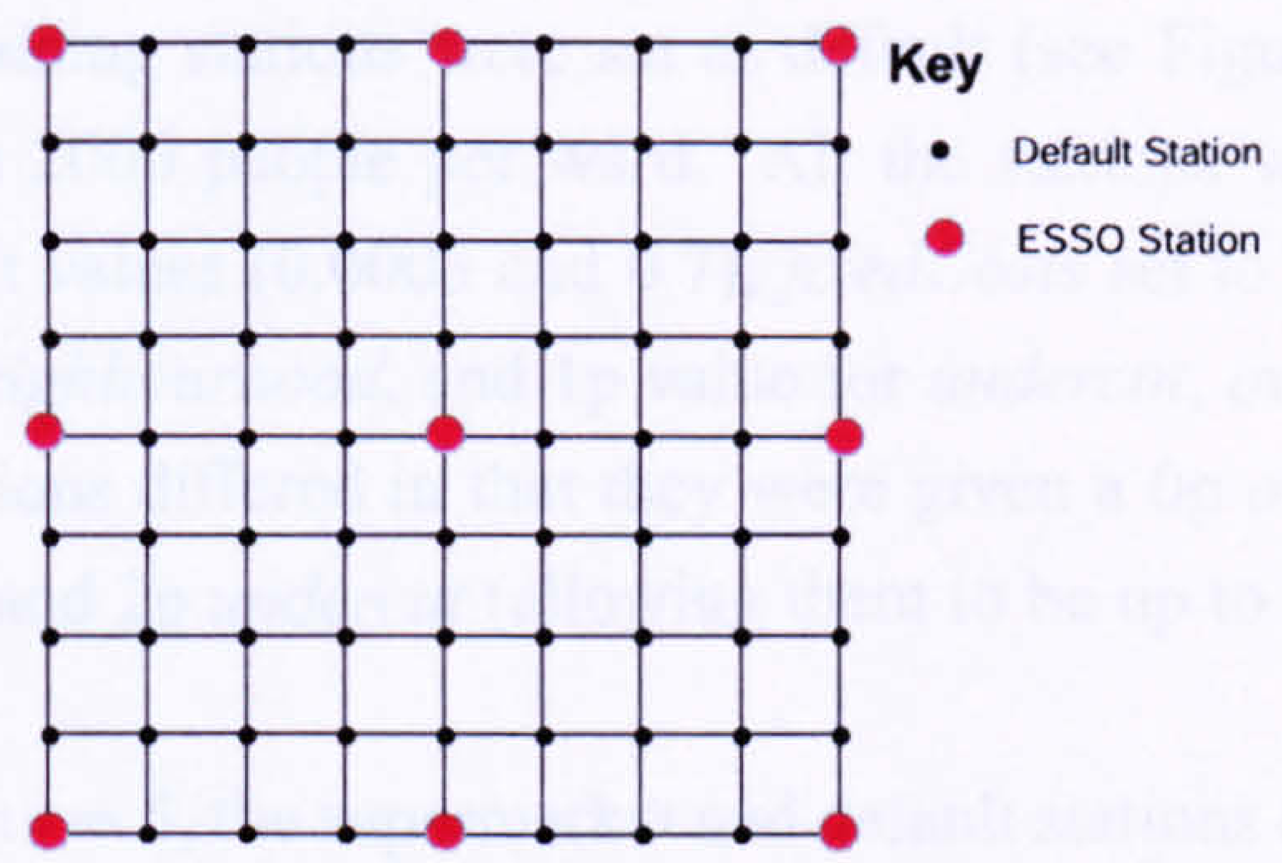


Figure 7.18: Representation of the position of the Esso and default stations on the idealise grid.

are dropping their price. This trend continues over time (Figures 7.19 (b) and (c)) with the Esso stations all dropping by 1p at the edges by $t = 15$. All the stations are aiming to maximise their profit and, as outlined in §6.8.1, will implement price increases or decreases to achieve this. The stations at the edge of the grid are not surrounded by as much population as those in the centre. This means that there are not as many people buying petrol, thus causing a decrease in sales and profit. The Esso stations are therefore trying to "buy" market share at lower prices in order to increase profits. By $t = 15$ (c), the Esso stations within the centre area can also be distinguished by slightly lower prices. This is due to the effect of the Esso Price Watch.

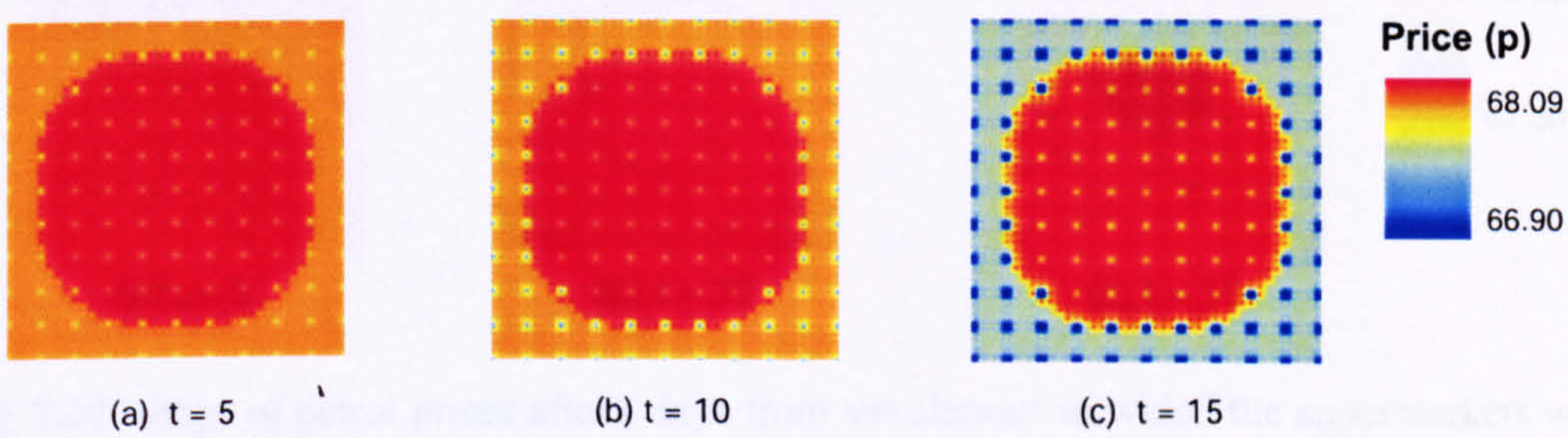


Figure 7.19: Comparison of the petrol prices at Esso and default station prices after (a) $t = 5$, (b) $t = 10$ and (c) $t = 15$ days.

In terms of price distribution, there is only a 1.7p difference between the highest and lowest petrol station. Assigning the Esso stations their own rule set has resulted in a tighter distribution of prices within the system.

Supermarkets

Evidence from the literature and analysis undertaken in Chapter 4 have shown supermarkets to be the most competitively priced of all stations. In this section, some of the stations will be assigned rules similar to those that supermarkets are thought to operate and the impact on the system assessed. For example, we can assume that supermarket stations will not let themselves be overpriced and will readily undercut nearby stations.

A 40km x 40km with every 4th node set as an supermarket station (i.e 1 in every 16 stations are supermarkets); the remaining stations were set as default (see Figure 7.18). The population was evenly distributed with 2000 people per ward. All the stations were initialised at 68p. β and λ were set at the default values (0.0003 and 0.7); *fixedCosts* set to £50; *costToProduce* 60p; *changeInProfit* £2; 10km *neighbourhood*, and 1p value for *undercut*, *overprice* and *priceChange* rules. The supermarket stations differed in that they were given a 0p *overprice* (forcing them to match competitors pricing) and 2p *undercut* (allowing them to be up to 2p cheaper than competition).

Figure 7.20 shows that at $t = 5$, the supermarket and default stations on the edge of the grid are the first to lower their prices. These stations have fewer consumers surrounding them and therefore have to lower their prices to try and attract more custom to retain profit. At $t = 10$, the impact of the edge stations lowering their prices has affected all the stations in the grid. The centre stations have lowered their prices. By $t = 15$, the system has settled down with the centre stations (with more potential consumers surrounding them) having the lowest prices. The supermarket stations are the most competitively priced within their neighbourhoods. These stations, due to their cheaper prices are also selling more petrol. For example, at $t = 15$, the centre station (a supermarket) is selling 6400 litres per day whilst its neighbours (non-supermarkets) are only selling 5186 litres. A similar trend is seen at other supermarket stations.

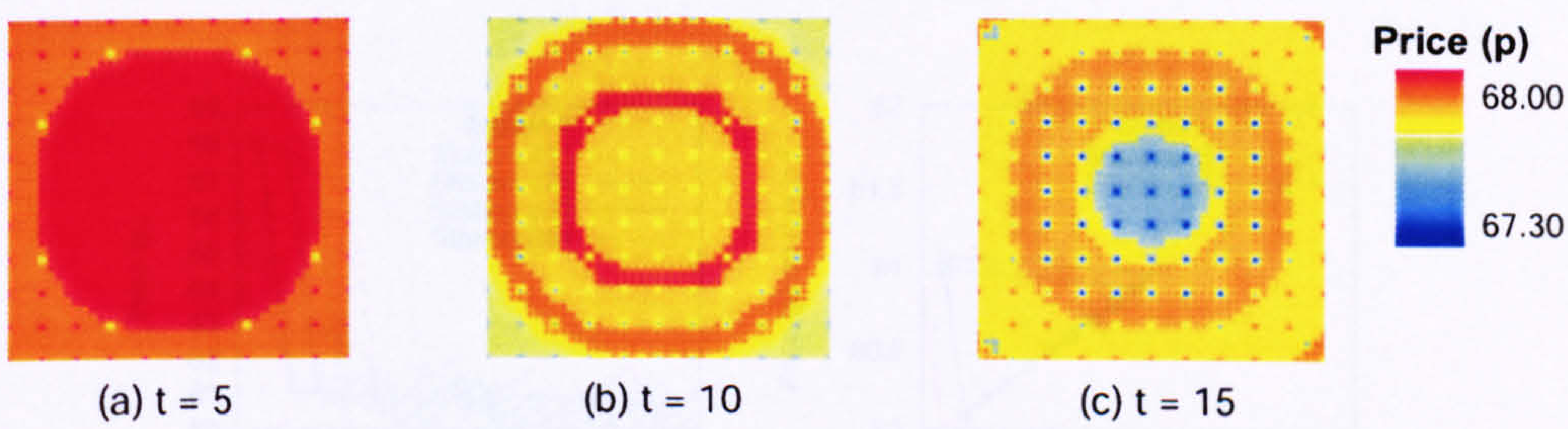


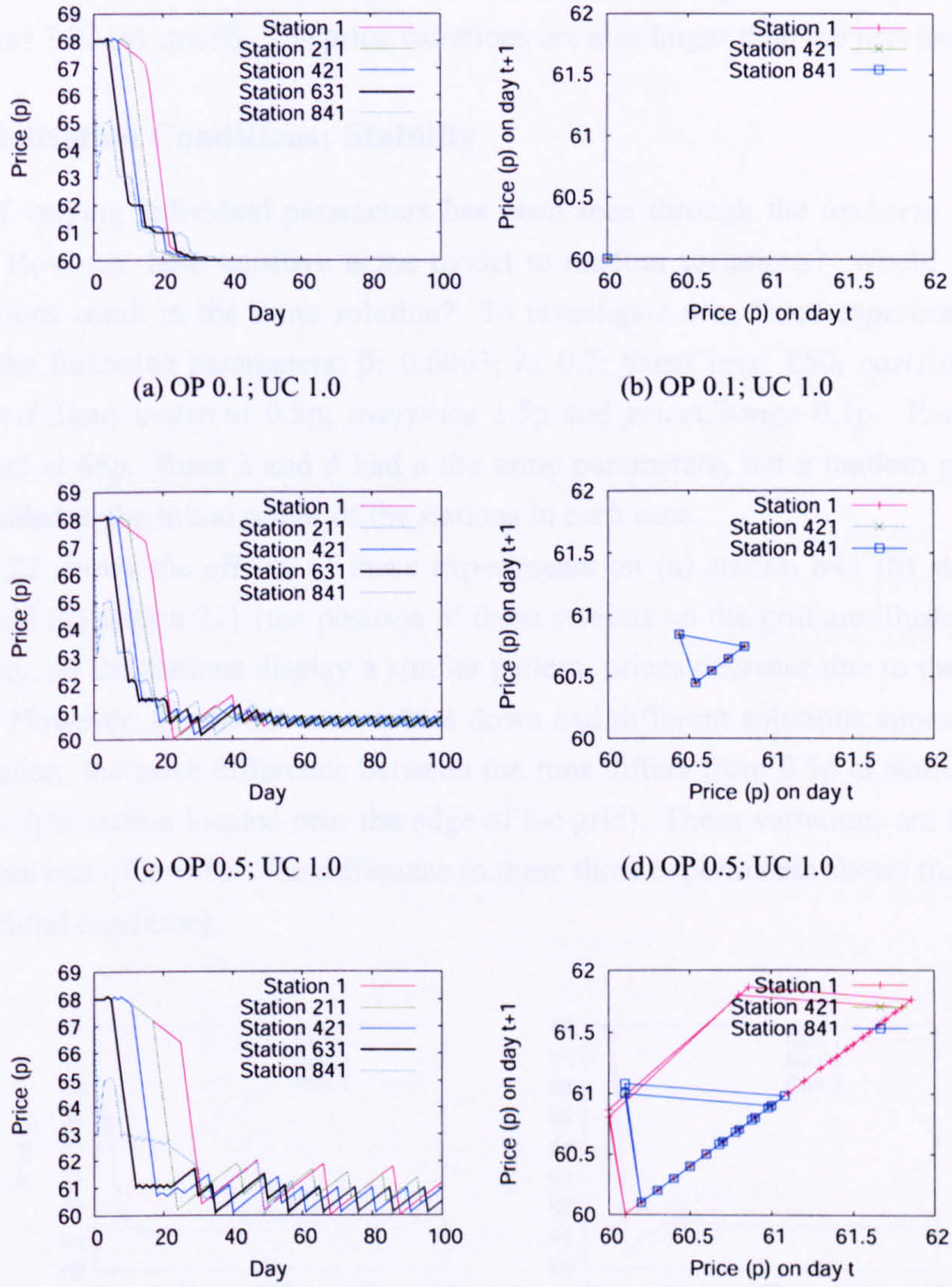
Figure 7.20: Maps of petrol prices after t days from simulations in which the supermarkets were assigned a competitive rule set whilst the other stations operate the default rule set.

7.5 Sensitivity of the System

Some of the experiments undertaken have shown evidence of periodic and cyclical behaviour. In this section, the sensitivity of the system to small changes in parameters and initial conditions will be investigated further. The ramifications of this for the model will also be discussed.

7.5.1 Periodic Cycling in Individual Parameters

During experimentation with individual parameters, evidence of period cycling and chaotic behaviour was seen. This was particularly evident in experimentation with the *costToProduce* and *undercut* and *overprice* parameters. The *undercut* and *overprice* simulations produced a variety of behaviour and these will be investigated here. The price over time (Figure 7.21 (a)) was plotted along with a corresponding phase diagram (Figure 7.21 (b)). The phase diagrams plot the price



(a) OP 0.1; UC 1.0

(b) OP 0.1; UC 1.0

(c) OP 0.5; UC 1.0

(d) OP 0.5; UC 1.0

(e) OP 1.0; UC 1.0

(f) OP 1.0; UC 1.0

Figure 7.21: Time series and phase diagrams of petrol price at the analysis stations for simulations with different *undercut* (UC) and *overprice* (OP) parameters.

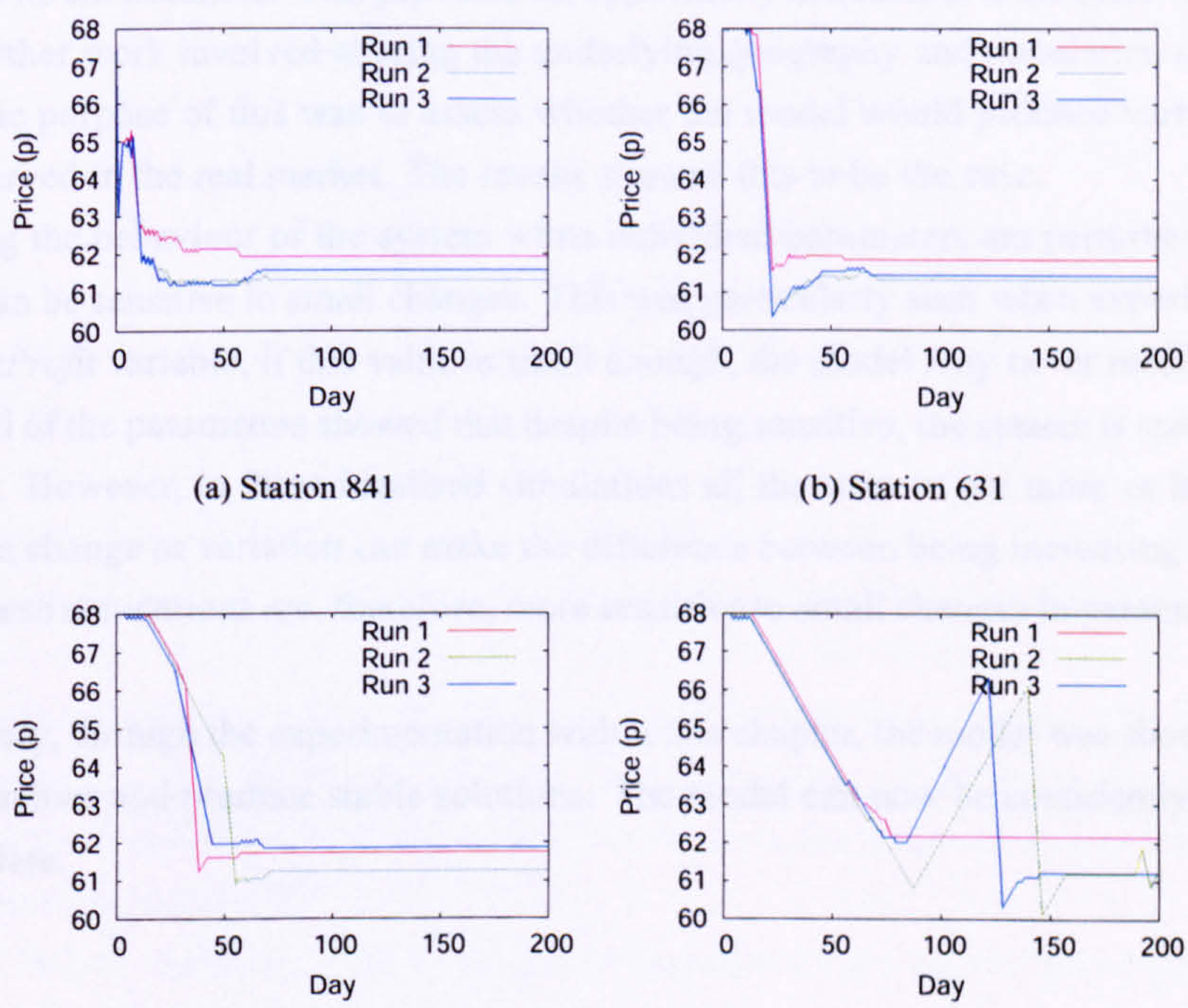
on day t against the price on day $t+1$ for stations 841, 421 and 1 (the centre, middle and corner station; stations 211 and 631 were omitted to keep the diagram clear). The data was plotted after day 60 when the model had settled into a solution. Plotting the data in this manner allows periodic behaviour to be easily identified.

Figure 7.21 (a) shows the model settling down quickly to an equilibrium. The corresponding phase diagram (Figure 7.21 (b)) reflects this. The point is fixed, the price is constant at all of the plotted stations. Figures 7.21 (c) and (d) show a different type of behaviour. After the model settles down, a 5 day periodic cycle appears. The stations follow the same cycle but, as Figure 7.21 (c) shows, not all the stations have the same price at the same time. The behaviour displayed in Figures 7.21 (e) and (f) is even more complicated than the previous examples. There is evidence of cyclical behaviour, but it is not quite periodic. The cycle in the prices is longer than the 5 days seen in Figures 7.21 (e) and (f). The price variations are also larger than the previous simulation.

7.5.2 Initialisation Conditions: Stability

The effect of varying individual parameters has been seen through the *undercut* and *overprice* parameters. However, how sensitive is the model to random variations? Would perturbing the initial conditions result in the same solution? To investigate this, three experiments were run. Run 1 used the following parameters; β: 0.0003; λ: 0.7; *fixedCosts*: £50; *costToProduce*: 60p; *neighbourhood* 5km; *undercut* 0.5p; *overprice* 1.5p and *priceChange* 0.1p. Each station was initially priced at 68p. Runs 2 and 3 had a the same parameters, but a random perturbation of ±0.1p was added to the initial prices of the stations in each case.

Figure 7.22 shows the effects of these experiments on (a) station 841 (b) station 631, (c) station 421 and (d) station 211 (the position of these stations on the grid are illustrated in Figure 7.11). Initially, all the stations display a similar pattern, prices decrease due to the price drop at station 841. However, the model soon settles down and different solutions appear. By the end of the simulation, the price difference between the runs differs from 0.5p at station 841 to 1.0p at station 211 (the station located near the edge of the grid). These variations are larger than the initial perturbations of ± 0.1p. The difference in these three experiments shows that the model is sensitive to initial conditions.



(a) Station 841 (b) Station 631

(c) Station 421 (d) Station 211

Figure 7.22: Time series of petrol price for three runs with similar initial conditions. Results are shown at 4 of the analysis stations.

The sensitivity of the model means that small random variations can significantly change which stations are the most successful. For example, if the threshold of the *changeInProfit* variable (the level of profit change at which the station will not change its price) is exceeded, the model will increase or decrease its price by *xp*. However, if the amount of profit made falls below the threshold, no change would be made. This can result in discontinuous behaviour, i.e. a small change in initial conditions may make the difference between a station changing price by *xp* and a station not changing price. Such behaviour is realistic, in the real market stations have a tendency to alter their prices in "jumps" (0.5p - 1p) rather than in small increments. This sensitivity of the model to small changes in the initial conditions and parameters can have a significant effect on the whole system. These idealised simulations are particularly sensitive to small changes in conditions because of the homogeneous distribution of stations and initial prices. The stations are only differentiated by small differences in sales and profit, therefore they are sensitive to small changes in the controlling parameters. In contrast, simulations using real initial data may be expected to be more robust because of the naturally occurring larger variation in sales, price and profit between stations.

7.6 Conclusion

Using idealised simulations has allowed the sensitivity and robustness of the system to be tested in a controlled environment. By use of simple diffusion experiments, the model was forced to react to changes in its environment. This provided an opportunity to examine if the rules were operating sensibly. Further work involved altering the underlying geography and population distribution of the grids. The purpose of this was to assess whether the model would produce variations similar to those observed in the real market. The results showed this to be the case.

Assessing the behaviour of the system when individual parameters are perturbed showed that the system can be sensitive to small changes. This was particularly seen when experimenting with the *changeInProfit* variable, if this value is small enough, the model may never reach equilibrium. Perturbing all of the parameters showed that despite being sensitive, the system is stable (it reaches equilibrium). However, in these idealised simulations all the stations are more or less equal so a small random change or variation can make the difference between being increasing or decreasing profit, and these simulations are, therefore, more sensitive to small changes in parameters or initial conditions.

In summary, through the experimentation within this chapter, the model was shown to produce sensible behaviour and produce stable solutions. The model can now be confidently tested on the real market data.

Chapter 8

Experimentation with Real Data

8.1 Introduction

Chapter 7 was concerned with testing the behaviour of the model and model parameters under idealised conditions. This "standardised" approach removed the effect of geography, distributing both population and petrol stations at equal intervals. By using diffusion experiments and recreating spatial variations, the behaviour, sensitivity and robustness of the system were tested. The results of these experiments concluded that the model, under idealised conditions, produced sensible results and was reasonably robust.

The next step in testing and validating the model involves application to a more realistic system. Analysis of the real data in Chapter 4 highlighted the irregular spatial distribution of petrol stations (and population) within West Yorkshire and the Yorkshire region. The examination of the petrol market in Chapter 4 brought to light the complexity of the interaction of factors that contribute towards the setting of a petrol price. The availability of the real data provides an opportunity to test the ability of the model to replicate patterns and trends that occur in the real world. This will be achieved by employing a mixture of diffusion and sensitivity tests similar to those used in Chapter 7.

The construction of a spatial interaction model, linked to the agent model, to account for customer behaviour was detailed within Chapter 6 (the "hybrid model"). A simple network model (the "network hybrid model") was also constructed to distribute these consumers around the study area (based on non-cascading travel to work data). The network model was redundant within Chapter 7 as the population within the idealised experiments was uniform. The use of the real data provides the first opportunity to compare the performances of the pure agent model, hybrid model and network hybrid model. Does the inclusion of a consumer and network model improve the overall model performance?

However, before the models can be compared, suitable values for the model parameters have to be derived. The work within this chapter will first focus on deriving suitable parameters, then comparing the performance of each model over time and its ability to recreate spatial variations. A series of diffusion and sensitivity experiments will then be performed on the model that gives the best performance. These will aim to examine the robustness of the model and its ability to produce sensible results under a variety of initial conditions.

8.2 Choice of Parameters

The parameters used to run the hybrid model in Chapter 7 were derived through experimentation with the idealised data. This will render them largely unsuitable for application to data that is irregular both in terms of petrol station and population distribution. This section of work will draw on information gleamed from the literature in Chapter 2, the analysis of the real data presented in Chapter 4 and behavioural information about the model from Chapter 7. The parameters will be derived to fit the West Yorkshire data from July 27th. For all parameters, these are at best estimates. A more objective method of selecting parameters will be presented in Chapter 9 with the use of a genetic algorithm.

8.2.1 β and λ

In West Yorkshire, the average distance from a petrol station to its nearest neighbour is 1423m (see §4.7.2, Table 4.4). Within a neighbourhood of this size, each petrol station would only have, on average, between 2 and 4 neighbours. For β to exert an influence on the system, there needs to be significant price competition between stations. To enable this, 1000m was added onto the average distance giving a value of 2400m. This gives an average of 7 stations within the neighbourhood (see §4.7.3, Table 4.5). Taking this value, a consumer located 2400m from the garage would be half as likely to buy petrol at the station than from an equivalently priced station situated adjacent to the consumer. Using Equation (7.1), a distance of 2400m produces a β value of 0.0003.

λ was calculated by using a combination of the average and maximum price change per day (§4.7.1), derived from the real data. The mean absolute price change (calculated over those stations where the price altered) was 0.27p per day, the maximum price change was 4.6p. The maximum value of 4.6p seem rather large, but on the other hand the maximum price change should be larger than the mean value. To this end, a higher figure of 1.0p will be used. Using Equation (7.2), 1.0p translates to a λ value of 0.7.

8.2.2 Fixed Costs and Cost to Produce

In §7.2.3 it was concluded that profit decreases with increasing *fixedCosts* and *costToProduce*. In principle, these values could be known. However, data of this nature is commercially sensitive and not available within the public domain. It is known that the profit margin is very tight due to the constraining factors of fuel tax and duty (see §2.2.2). We can therefore conclude that the *costToProduce* will be a high proportion of the actual price. *fixedCosts* is the amount that it costs the station per day to keep running irrespective of the amount of petrol sold. This includes part of the maintenance of the site and payment of staff. The *costToProduce* parameter is the amount per litre that it costs to produce the petrol, for example transportation and refinery costs. In practice the *costToProduce* will also include a portion of the staff costs and maintance since a large garage selling a large volume of fuel will need more staff than a smaller garage. The values for the *fixedCosts* will also be lower than might be anticipated because this chapter is only considering unleaded fuel, whereas in reality a petrol station would also sell other types of fuel as well as often selling other goods. The profit from all these would also contribute towards the fixed costs.

§2.2.2 presented the breakdown of a petrol price of 74.9p; 57p accounts for the duty and tax, 13.9p covers the cost of the product and the remaining 4p goes to the retailer. It can be assumed that the greater proportion of the 13.9p covers the *costToProduce* and the remainder covers the *fixedCosts*. This could be approximately 13p for *costToProduce* and 0.9p for *fixedcosts*. The total *costToProduce* in this example was 57p (tax) + 13p (*costToProduce*) = 70p (these are the variable costs). The remaining 0.9p covers the *fixedCosts*. The total *fixedcosts* is therefore, 0.9p × the amount in litres sold. This is on average about 8,000 litres (based on data from Catalist (accessed 2003)). This gives £80. However petrol in the real data was, on average, 4p cheaper (see §4.4.1), therefore the *costToProduce* will be taken as 66p rather than 70p. The *fixedCosts* will remain at £80.

8.2.3 Change In Profit

Experimentation undertaken in §7.3.3 found the system to be very sensitive to variability in the *changeInProfit* value. This value controls how readily a station reacts to changes in its profit level. The model rapidly settled into a steady solution when parameterised with values over £30, and remained in flux with values under £15. This suggests that an appropriate value to use would fall between these two values. The results presented in Figure 7.12 showed that the results for *changeInProfit* values of £20 and £25 were almost identical with the system reaching equilibrium after approximately 25 days.

However, these values are based on the idealised simulations and not necessarily appropriate for the real data. A useful guide for determining values for the real data is to compare the average profit made by each station in the real and idealised data. This was achieved by using Equation (6.8) (see §6.6.1) to calculate the average number of litres of unleaded petrol sold at each station per day. Using the 40x40 idealised grid (1681 stations), each station sold an average of 3675 litres of unleaded petrol per day. There are 517 petrol stations in West Yorkshire, but only 262 of these are used in the experimentations on July 27th (the remaining 255 are either out of industry or do not have a price recorded for that day). Using Equation (6.8), these 262 stations were calculated as selling an average of 5959 litres of unleaded petrol per day. This is almost double the amount calculated for the idealised data. This suggests that the *changeInProfit* value used in the idealised simulations should be scaled accordingly for use in the real data simulations. Figures between £20 - £25 were suggested as the most appropriate for use within the idealised simulations (see §7.3.3). Values between £40 - £50 are therefore appropriate for the real data. As the real data contains more variation than the idealised data, this figure is a general guideline. A default value of £40 will be used in these experiments.

8.2.4 Undercutting and Overpricing

The *undercut* parameter controls how much a station can undercut the price of its competitor (in pence). Selecting a small value for this parameter may squash any variations that exist within the real data. Conversely, choosing a parameter that is too large may result in the rule having no effect even if an unrealistic price difference occurred. The standard deviation (SD) is an indicator of the amount of variation within the prices. From 27th July to 19th August for West Yorkshire, the SD ranges between 0.32 - 0.90 (see Table 4.3). There are no significant differences from this range in any of the geographical classifications. An alternative approach to assessing suitable *undercut* values is to extend the neighbourhood analysis in §4.7.4 to find the cheapest and second cheapest stations in each neighbourhood. The difference between these two gives the maximum price by which the cheapest station undercuts the others in the neighbourhood. For a 5km neighbourhood size, 157 of the neighbourhoods had an undercut value of less than 1p, 71 neighbourhoods had an undercut value of 1-1.9p, while only 30 had a values of 2-2.9p. None of the cheapest stations were more than 2p cheaper than the others in their neighbourhood. These two analyses suggests that an appropriate figure to use for the *undercut* would be 1.0p. This value is large enough to allow natural variations to exist, but not too large to be ineffectual.

The *overprice* parameter determines the amount by which a station can be more expensive than its nearest competitors (in pence). As with the *undercut* parameter, choosing a value that is too large may result in the rule becoming ineffectual. If the value is too small, variations within the data will be suppressed. In the analysis of the real data, the maximum price change between 27th July to 19th August was found to range between 2.0 - 4.6p (Table 4.3). This suggests that the *overprice* variable should be set to at least 5p to maintain price variation. The choice of this parameter is substantiated by the calculation of the variation in price ranges with a 5km neighbourhood size. The price range in almost all of the neighbourhoods was less than 5p (Table 8.1). Setting the *overprice* value to this amount will therefore prevent unrealistically large price variations in a neighbourhood without suppressing naturally occurring variations.

Price Range (p)	Number of Neighbourhoods
0.0 - 0.9	7
1.0 - 1.9	11
2.0 - 2.9	74
3.0 - 3.9	64
4.0 - 4.9	89
5.0 - 5.9	12
6.0 - 7.0	4

Table 8.1: Price range within a 5km neighbourhood for July 27th.

8.2.5 Neighbourhood Size

The *neighbourhood* is taken to be a fixed circular area around the petrol station with a given radius. In §4.7.2, the average distance between stations was calculated at 1423m. This equates to an average of 4 neighbours (see §4.7.3). Considering the varying distribution of stations throughout the study area (as shown in Figure 4.2), the neighbourhood needs to be significantly larger to ensure that even the rural stations have some neighbours to generate competition.

Analysis of the real data in Chapter 4 (see 4.7.2) showed that a neighbourhood of 5km gives an average of 18 neighbours. Experimentation with different neighbourhood values (§7.3.2) using idealised conditions illustrated that using this neighbourhood size (containing about 20 stations) produced sensible results. 5km also corresponds to the neighbourhood size that is employed by

Esso in their Price Watch policy (see §2.3.4). This suggests that Esso consider prices within this distance to be important to competition.

8.2.6 Summary

A combination of numerical experiments, analysis of the real data and idealised simulations have been used to determine appropriate values for each of the parameters within the model for the first half of the data set (July 27th) for West Yorkshire. Table 8.2 presents a summary of these values:

Parameter	Value
β	0.0003
λ	0.7
fixedCosts	£80
costToProduce	66p
changeInProfit	£40
undercut	1p
overprice	5p
neighbourhood	5km

Table 8.2: Optimal values derived from the real data and literature for the model parameters.

Unless otherwise stated, these values will be used in the following experimental sections and referred to as the default parameters.

8.3 Comparison of Model Performances

In this section, the agent, hybrid and network hybrid models will be initialised with the real price data from July 27th for West Yorkshire. The parameters summarised in Table 8.2 will be used. The output on day 10 from each model will be compared with day 10 of the real data (August 6th). This method of comparison has been previously detailed in §5.3.7. The comparisons will show which of the three models best predicts the values and spatial distributions of the real data prices.

The second half of the data set (August 19th - September 1st) and the Yorkshire region will be used for further testing of the model's performance in §8.8.

8.3.1 Visual and Statistical Comparisons

Figure 8.1 shows the spatial distribution of price difference (in pence) between the real and model data on day 10. There are no great differences between the spatial distribution of the price differences for the model output and real data (Figure 8.1 (a)). In general, positive price differences are found in urban areas with negative variations in rural areas. There is however, some variation in the magnitude of these price differences. The hybrid and network hybrid model (Figure 8.1 (c), (d)) both display almost identical price differences to the real data. The price differences in the agent model (Figure 8.1 (b)) are greater and more widely distributed across the study area.

Assessing these spatial patterns is largely a qualitative exercise. Quantifying the degree of these differences can be achieved by using various statistical techniques such as mean and standard

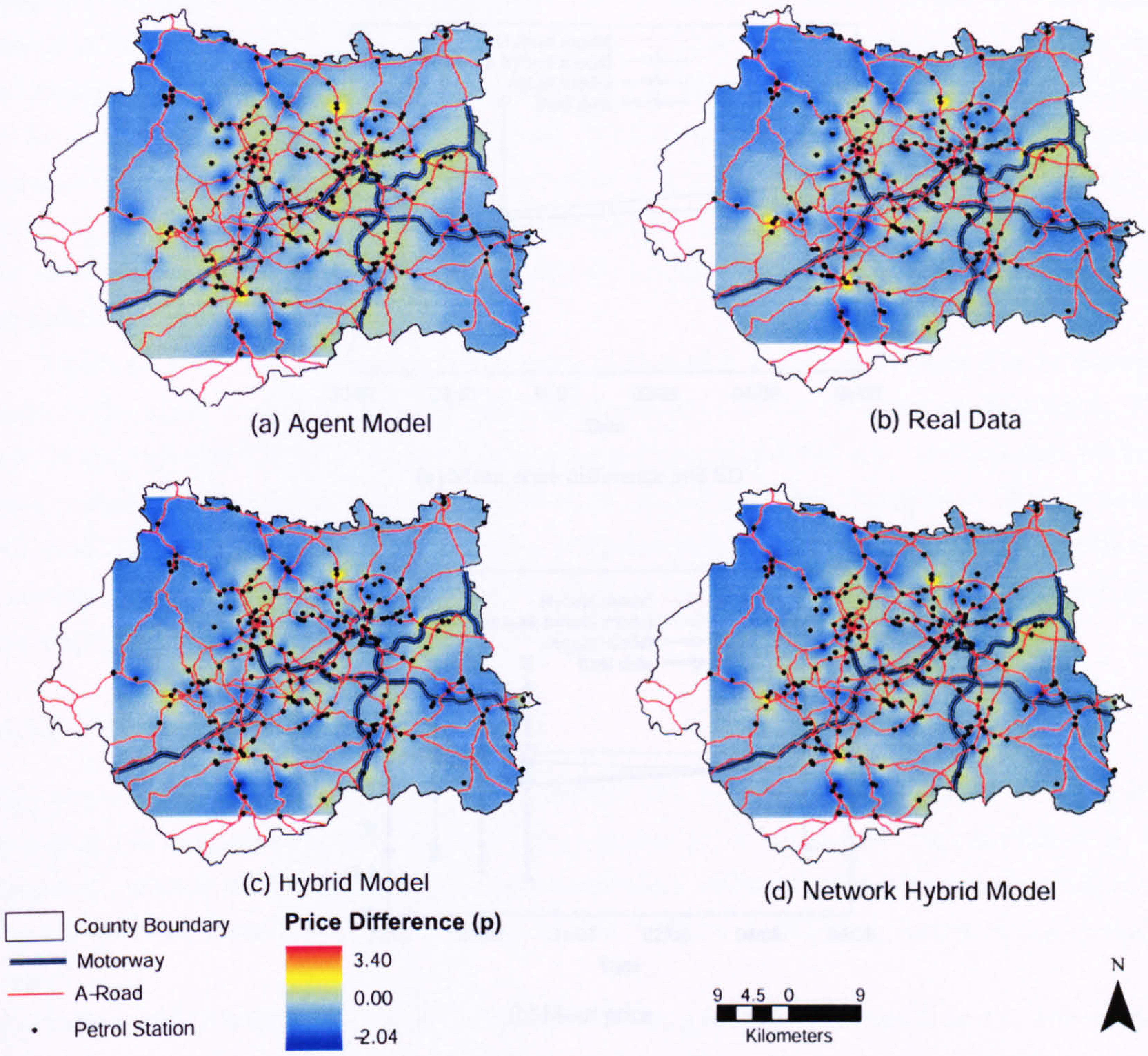
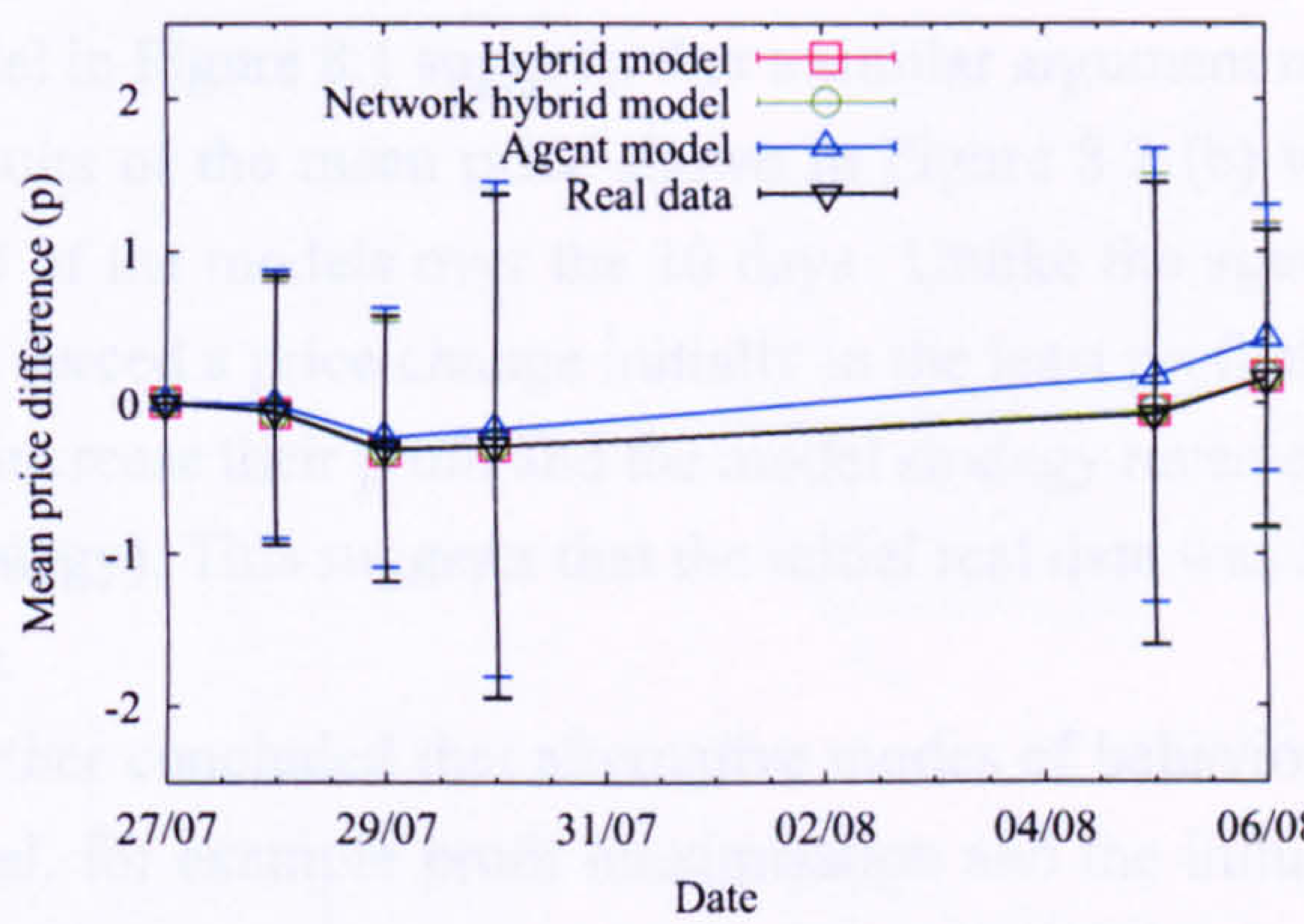


Figure 8.1: Price difference between the real data (day 10) and model output for the (b) agent model (c) hybrid model and (d) network hybrid model for West Yorkshire. The difference between the real data (July 27th and August 6th) (a) is included for comparison.

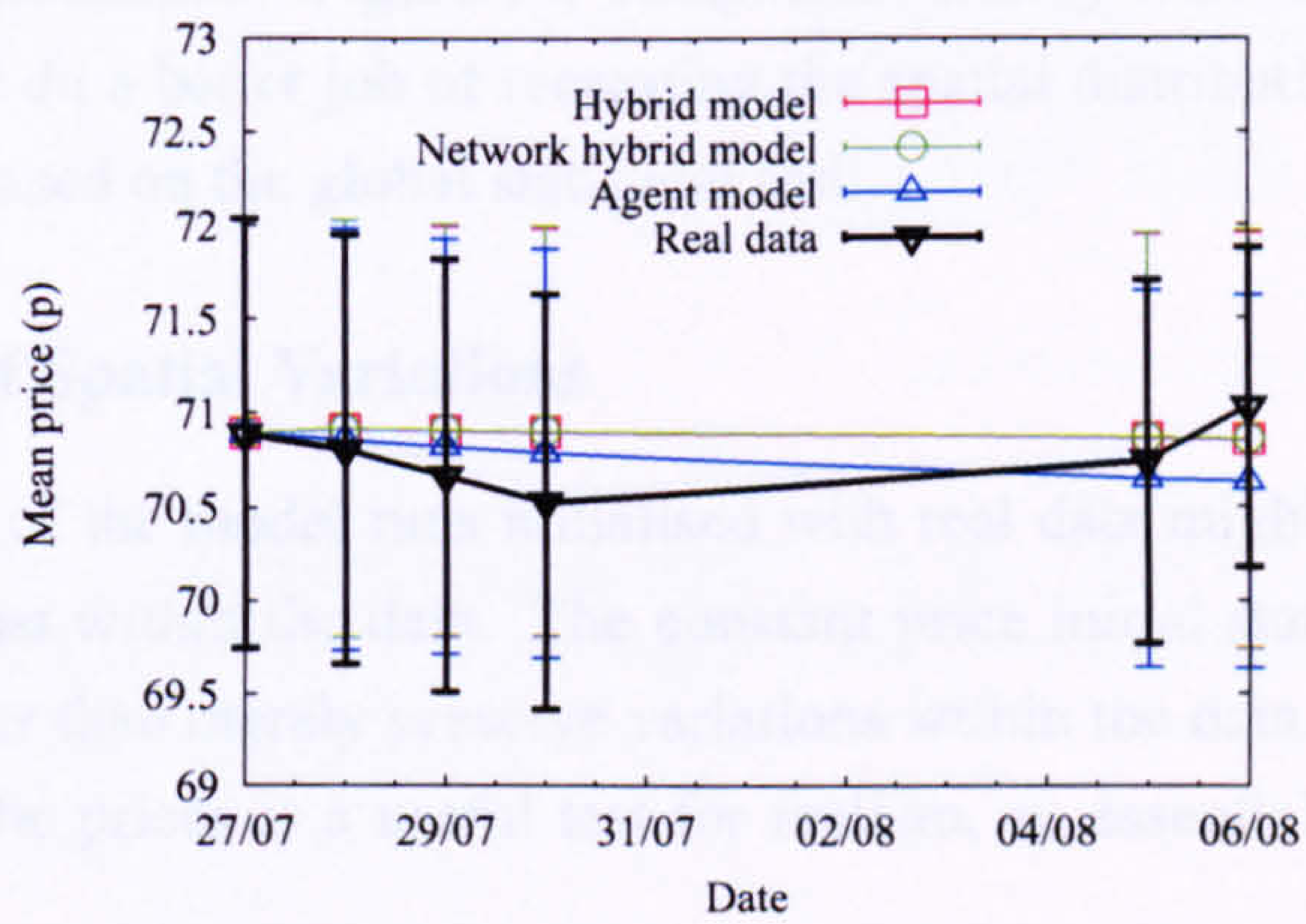
deviation (SD) of the price difference between the real and model data and standardised root mean square error (SRMSE). A detailed explanation of these techniques was presented in §4.9 and §5.3.7.

Figure 8.2 (a) shows that the mean price differences are almost identical for the hybrid and network hybrid model. The SD of the hybrid model is marginally larger than that of the network hybrid model. However, both model results mirror the pattern of the real data almost perfectly. The agent model also follows the patterns of the real data, but exhibits a slightly larger mean price difference and SD that increases over time.

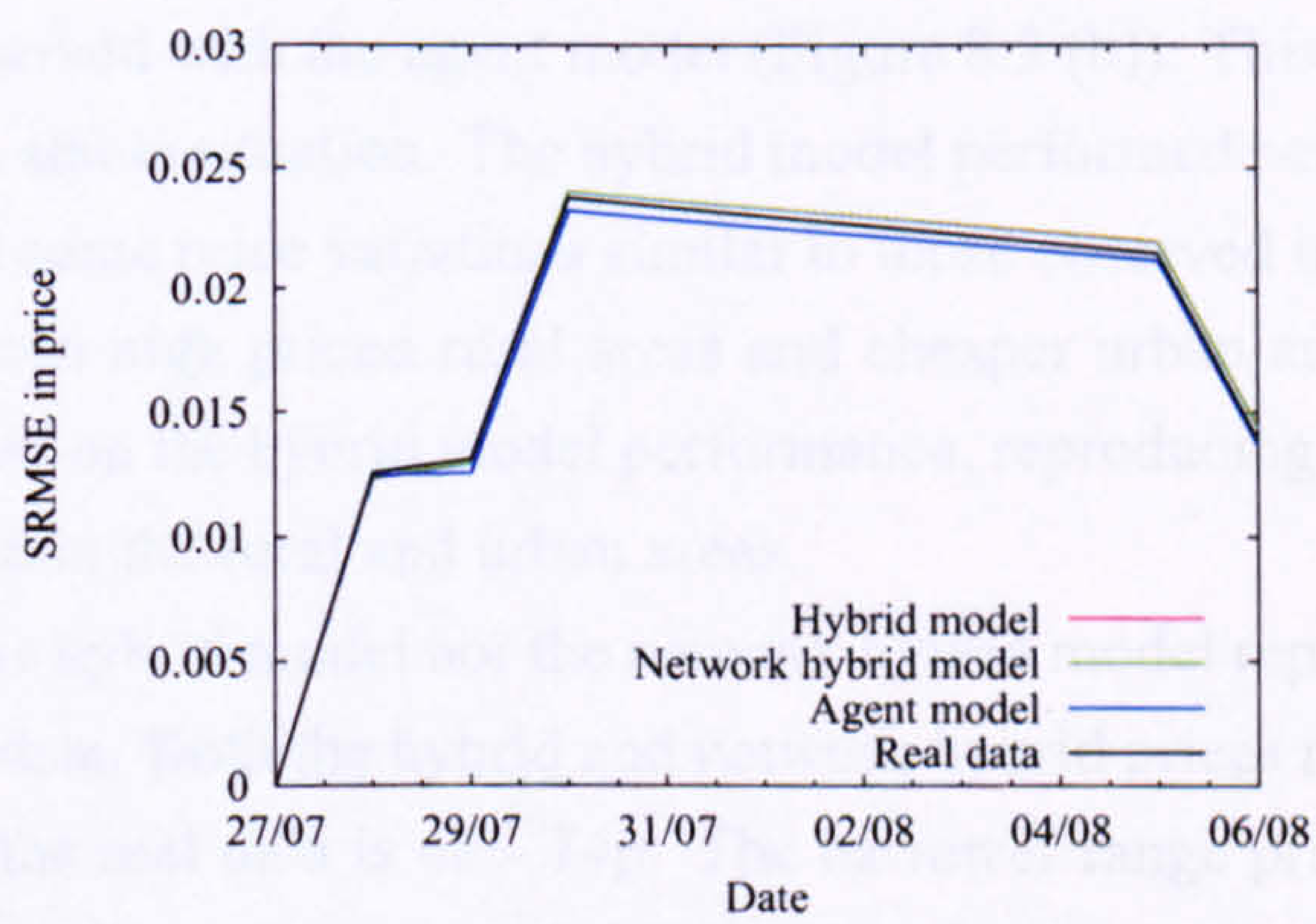
Over time, Figure 8.2 (c) shows that all the models have similar values for the SRMSE and follow the trends displayed by the real data. From this Figure, it appears that the agent model is producing the closest fit to the real data. However, this is misleading. Investigation of the agent model in Chapter 5 concluded that any variations in the difference between real and model data were due to the changes in the real data (the agent model only took a few days to hit equilibrium,



(a) Mean price difference and SD



(b) Mean price



(c) SRMSE

Figure 8.2: Comparison of (a) the mean price difference over time (SD is indicated by the vertical bars) and (b) the SRMSE between the real (day 10) and model data for the agent, hybrid and network hybrid model for West Yorkshire (Model initiated on July 27th). The real data differences are plotted for comparison.

clearly not enough time to operate a rule set). The similarities of the hybrid and network hybrid model to the agent model in Figure 8.1 suggests that a similar argument may also be applied. This is supported by the results of the mean price shown in Figure 8.2 (b) which show little change in the mean price of all of the models over the 10 days. Unlike the agent model, the hybrid and network-hybrid models forced a price change initially in the least profitable stations. However, in most cases this did not increase their profit and the model strategy reversed this change (see §6.8.1 for details of model strategy). This suggests that the initial real data was in fact close to the model equilibrium in this case.

However, §5.10 further concluded that alternative modes of behaviour needed to be incorporated to the agent model, for example profit maximisation and the influence of consumers. The hybrid and network hybrid models both account for this type of behaviour. Both models are built on a sounder theoretical basis than the agent model having the ability to replicate real processes, for example profit maximisation. Figure 8.1 compounds this by illustrating that the hybrid and network hybrid models do a better job of recreating the spatial distribution of prices even if they are slightly less good based on the global statistical test.

8.3.2 Recreation of Spatial Variations

One possible criticism of the model runs initialised with real data might be that they are merely preserving the variations within the data. The constant price initial state tests the ability of the model to generate rather than merely preserve variations within the data. Examining whether the models generate realistic prices is a useful test for realism, an essential feature in any dynamic model.

Each model was run with all the petrol stations assigned the same initial price. The average price of the real data on July 27th (71p) was chosen. The model was run to equilibrium. No price changes were observed with the agent model (Figure 8.3 (b)). This is because the rules used assumed that this was a stable situation. The hybrid model performed better (Figure 8.3 (c)) and is beginning to reproduce some price variations similar to those observed in the real data (Figure 8.3 (a)), for example between high priced rural areas and cheaper urban areas. The network hybrid model (8.3 (d)) improved on the hybrid model performance, reproducing with greater accuracy the price differentiation within the rural and urban areas.

However, neither the hybrid model nor the network hybrid model reproduce the range in prices present within the real data. Both the hybrid and network hybrid prices range from approximately 70 - 72p, the range of the real data is 68 - 74p. The narrower range produced by the hybrid and network hybrid model suggests that the model is limiting the price range too much. Larger values for the *undercut* and *overprice* parameter might help with this. A more objective selection of the parameters in Chapter 9 will show whether this is the case.

Based on the evidence presented here and in §8.3.1, the agent model has been comprehensively outperformed by both the hybrid and network hybrid model. A comparison of these two models identifies the network hybrid as the superior, reproducing the spatial trends of the real data with the greatest accuracy (Figure 8.3). This is reflected statistically, with the network hybrid model accurately modelling the prices over a 10-day period. Experimentation undertaken in the following sections will use only the network hybrid model.

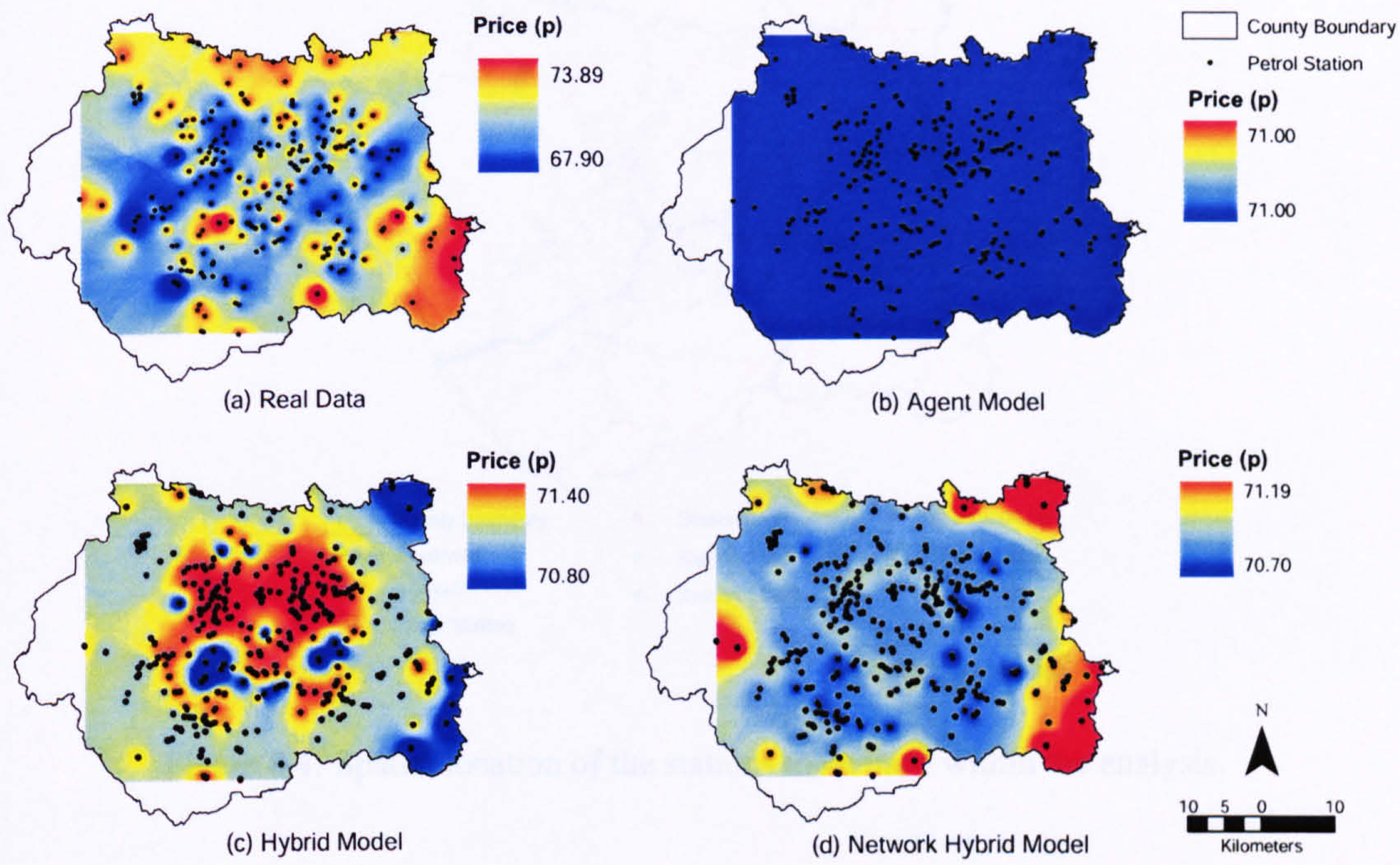


Figure 8.3: Price distributions for the various models 30 days after runs started with all stations initialised at 71p: (b) agent model, (c) hybrid model and (d) network hybrid model. Data from day 10 of the real data set (a) is included for comparison.

8.4 Space-Time Diffusion

The experiments performed here will follow the methodology used in the diffusion experiments in §7.3.1, but with the real distribution of petrol stations and data for West Yorkshire. One station will be dropped in price on the starting day and the impact on the system over subsequent days will be studied. Different experiments will use constant prices or real prices on July 27th to initialise the model. The parameters input are outlined in §8.2.6. The reaction of the real system in each of these situations will contribute important information about system behaviour, sensitivity and robustness as well as indicating how well the model performs.

The station that will be dropped in price is 8537 (Figure 8.4). This station is located within the centre of Leeds. The central position of station 8537 in a neighbourhood of other urban stations will ensure some symmetry in the diffusion of the new price. The reaction of several stations to the price drop will be analysed (Figure 8.4). These stations were selected at roughly equal intervals on a north-west transect from station 8537. They are drawn from different geographical parts of the study area, e.g. cities, small towns and rural areas. With increasing distance from 8537, the character of the area surrounding the stations changes from urban, through semi-urban to rural. Analysing the reaction of these stations to the price drop will increase understanding of the behaviour of the real system.

Station 8537 will be dropped by 5p to initiate the diffusion. Experiments with the idealised data in §7.3.1 concluded that using 3p or 5p precipitated better diffusion patterns than 1p. How-

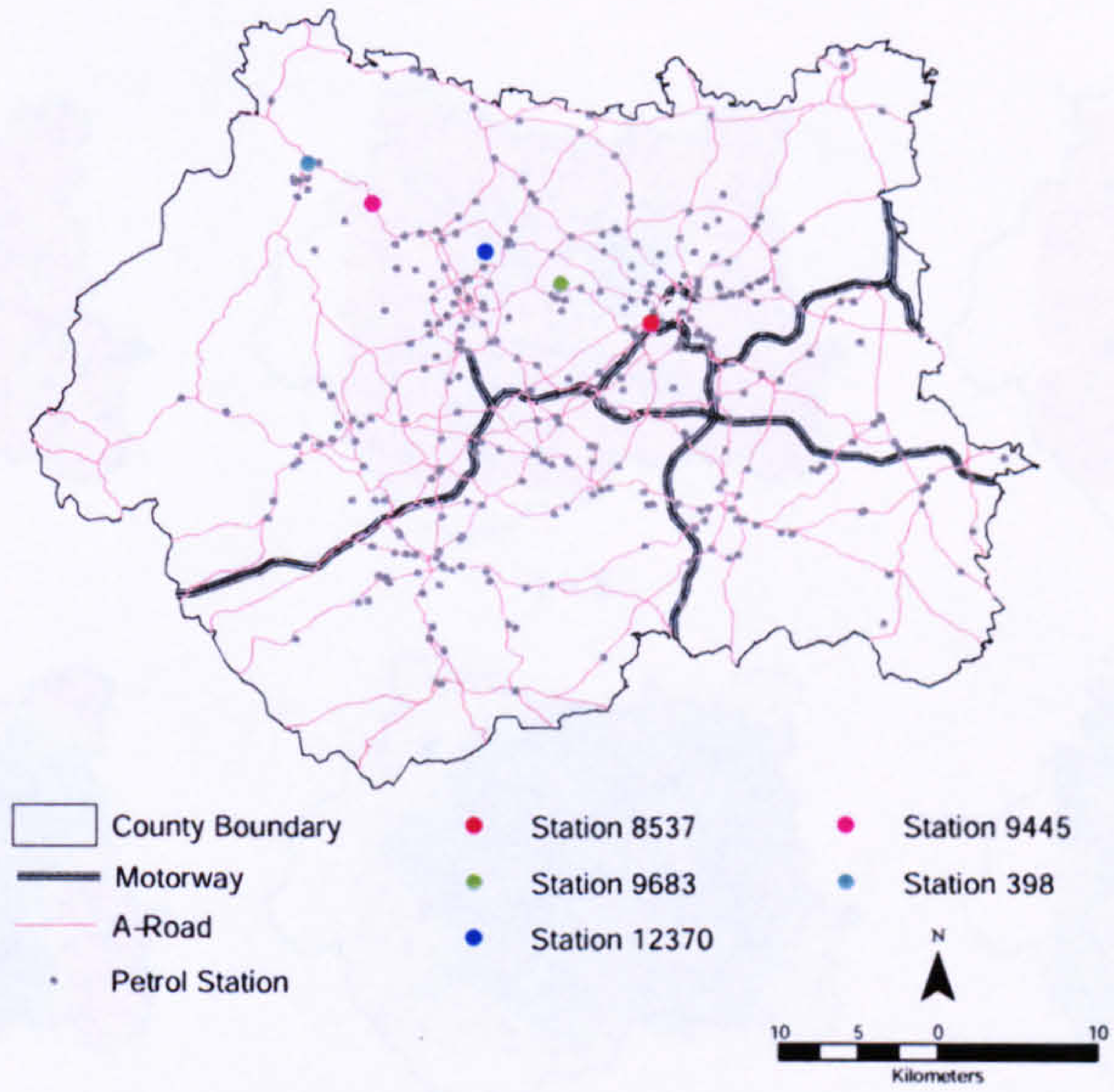


Figure 8.4: Spatial location of the stations to be used within the analysis.

ever, as we are interested in the reaction of the real system, a price drop of 5p will force a response. Dropping the price by 1p or even 3p would produce a change that could be absorbed by surrounding neighbours. A 5p difference is greater than any naturally occurring differences in the data but is still small enough to ensure that the stations continue to make some profit.

8.4.1 Experimentation with Constant Initial Prices

The model was initialised with all the stations set to 71p except for station 8537 which was set 5p cheaper at 66p. The parameters summarised in Table 8.2 were used except for the *changeInProfit* which was set to £2. This was to prevent the system reaching equilibrium too rapidly before the price drop had time to make an impact on the model (see §7.3.3 for further details). Figure 8.5 shows that the lower price gradually diffuses throughout the study area. By day 40, most of the stations have lowered their price by an average of 4p. Each of the selected stations in this simulation experience this 4p price drop, dipping under the price of 8537 before a gradual increase to a steady state (Figure 8.6 (a)). These differences are also reflected in the levels of profit made (Figure 8.6 (b)). The full impact of the price drop does not reach station 398 until day 40. By day 46, this station has also settled into a steady state. From this point onwards, and in the absence of any external price changes, there are only small fluctuations within the system.

The behaviour exhibited by the system is a direct result of the rules implemented. These in turn are directly linked to the profit maximisation strategy of the stations. Between days 10 and 20, the stations closest in proximity to 8537 (stations 9683 (green) and 12370 (dark blue) in Figure 8.6) begin to lose profit due to this new aggressive pricing by station 8537 (Figure 8.6(b)). Their response is to drop their prices by up to 1p per day (Figure 8.6 (a)) in an attempt to attract consumers back and thereby maximise profits. Over the course of the simulation, the effect of the price drop diffuses across the area triggering stations located further away from station 8537 to

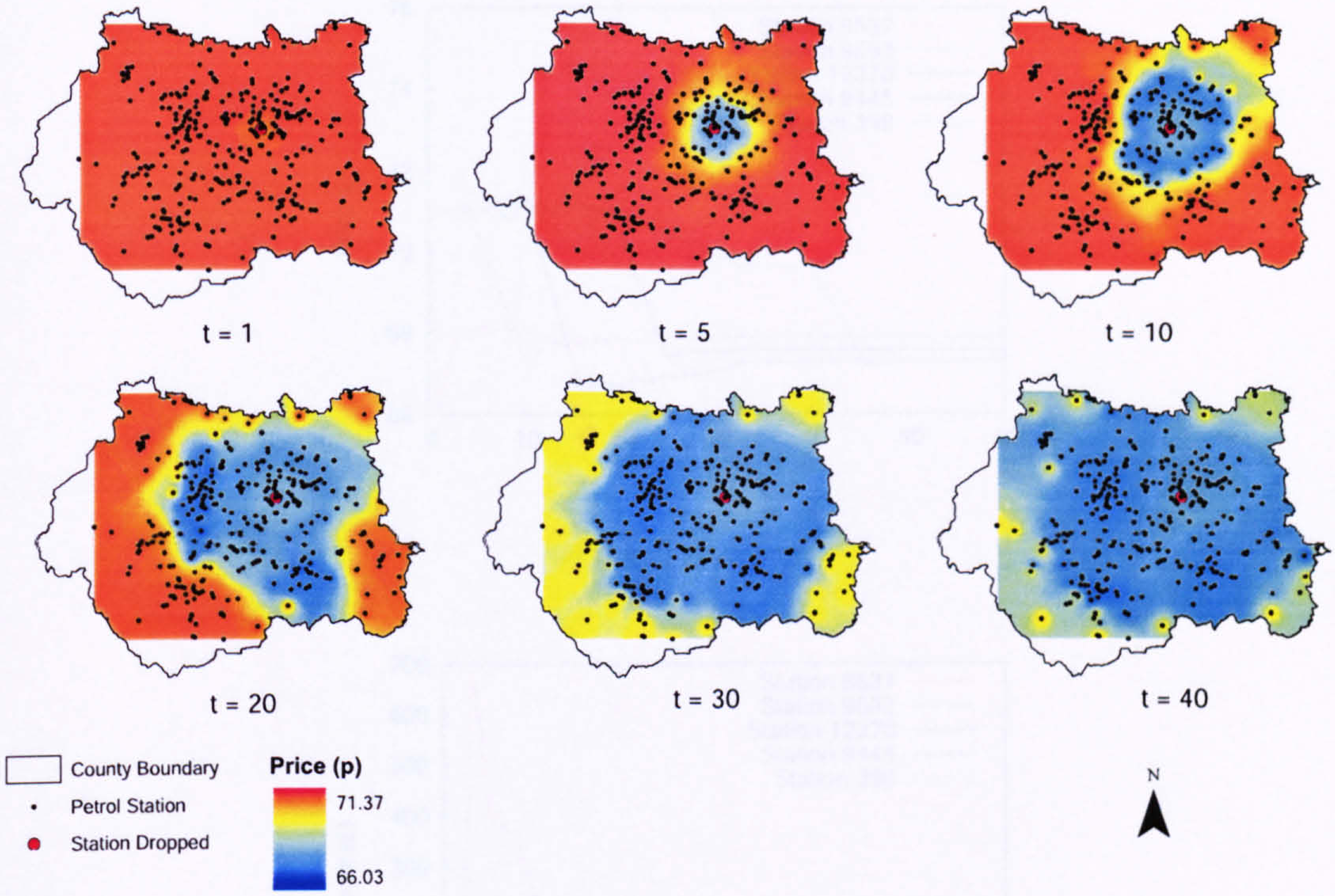


Figure 8.5: Results of the spatial diffusion of prices over time with all stations initialised at a constant price of 71p except for station 8537 which experiences a price drop of 5p.

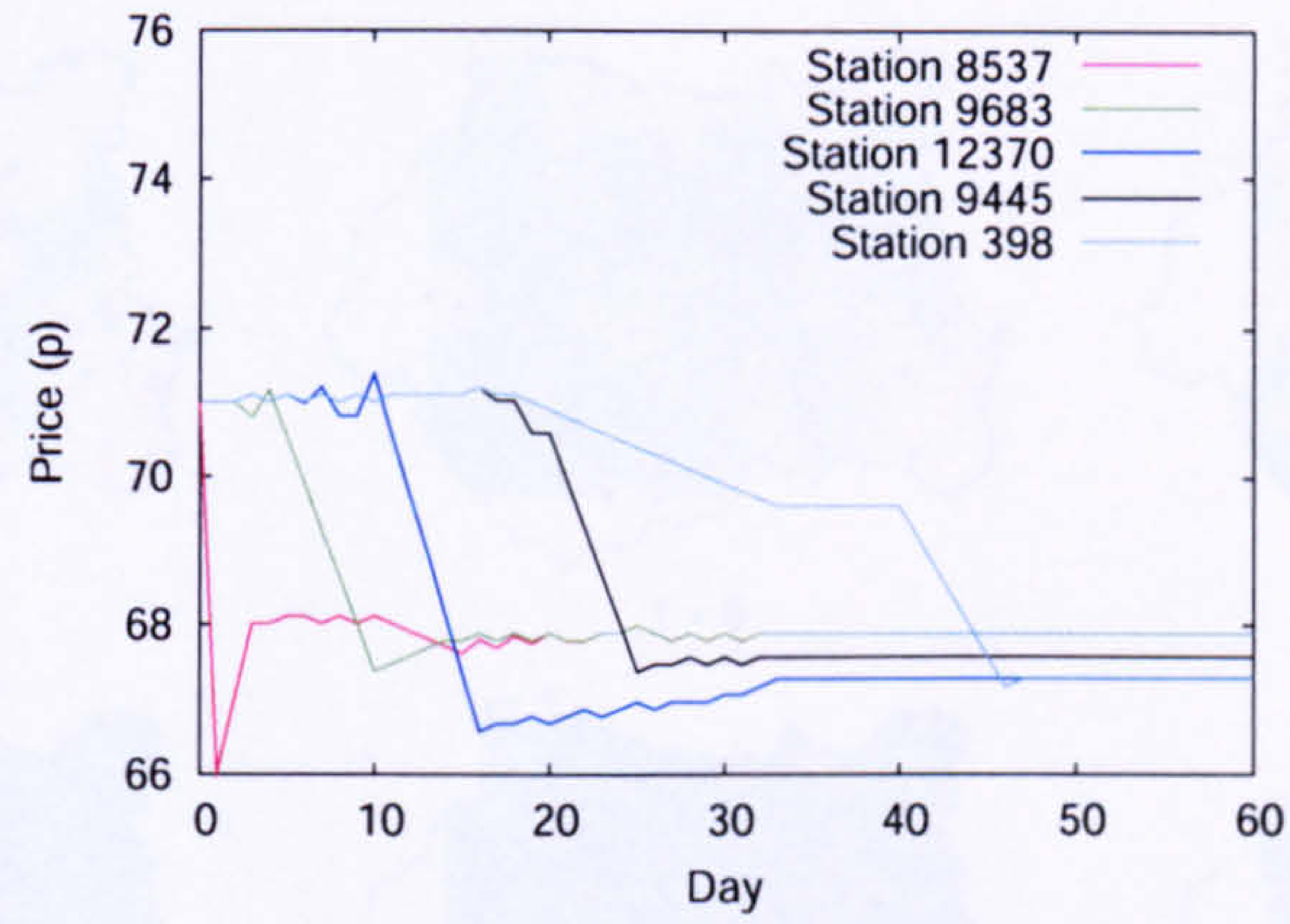
implement the same strategy; for example, stations 9445 (pink) and 398 (light blue) between days 20 to 45.

At the same time, station 8537 has large sales but a low profit margin due to its cheap prices and so increases its price in an attempt to maximise profit levels. This price increase slows down around day 5 and then exhibits minor fluctuations until day 30. At this stage, the prices appear to converge at a point where all the stations are making profit. This happens around 66-67p (the *costToProduce* variable is set to 66p, any prices lower than this will precipitate negative profits; Figure 8.6(a)). With the absence of any external influences (e.g. rise in tax or crude oil prices) to force an upwards kick in the prices, the simulation exhibits little fluctuation in prices or profit from day 45 through to day 60.

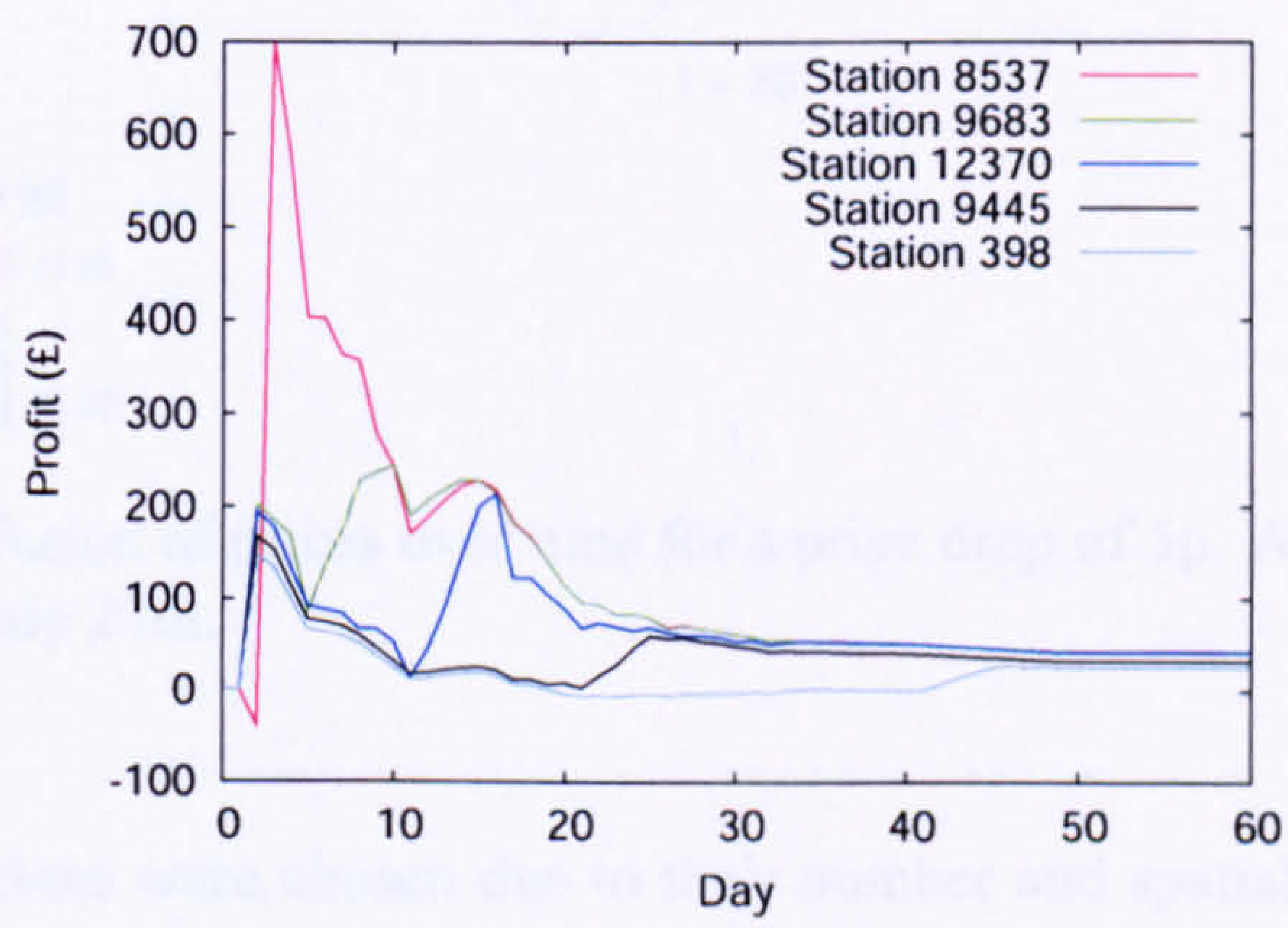
8.4.2 Experimentation with Real Initial Prices

The model was initialised with the real data (July 27th) and the default values detailed in Table 8.2 were used. Station 8537 was dropped by 5p (from 73p to 68p).

Figure 8.7 shows a slightly faster diffusion of the new price throughout the study area than observed in §8.4.1. By day 20, most of the stations have lowered their prices by an average of 3 - 4p (66 - 67p). On day 40, the price at the majority of stations has settled to approximately 66p and do not fluctuate again. Interestingly, in the south west corner of the study area, a genuinely occurring low price also precipitates a similar price drop diffusion across the study area (this is



(a) Price (p) over time



(b) Profit (£) over time

Figure 8.6: Change in (a) price and (b) profit over time for the selected stations in the constant price experiment.

particularly evident at $t = 10$).

Figure 8.8 shows the reaction of the price and profit are almost identical to that seen in §8.4.1. The drop in price of station 8537 precipitates a decrease in the profit of the surrounding stations. This forces a price drop until a point is reached where all stations are making profit once again (around day 35). This similarity in behaviour to the diffusion experiment using the constant price data shows the network hybrid model to be relatively robust to variations within the system.

8.4.3 Experimentation with Multiple Price Drops

So far, only price drops at individual stations have been considered. It might be hypothesised that all the stations belonging to a particular brand may drop their prices simultaneously. A diffusion experiment was carried out to examine the reaction of the system to such a price drop.

The model was initialised with the real data (July 27th) and the default values detailed in Table 8.2 used. The Esso stations were all dropped in price by 5p at the beginning of the simulation.

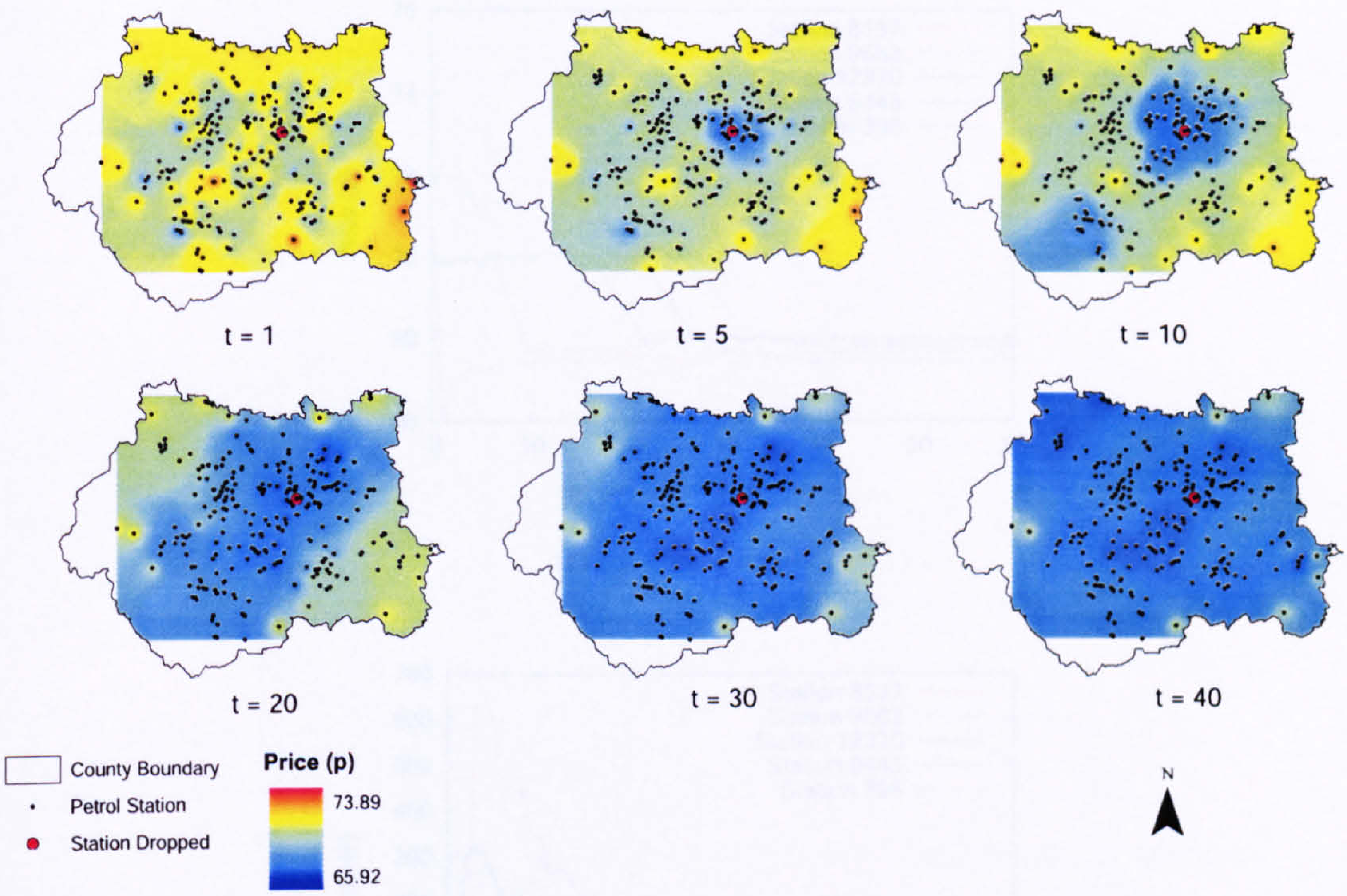
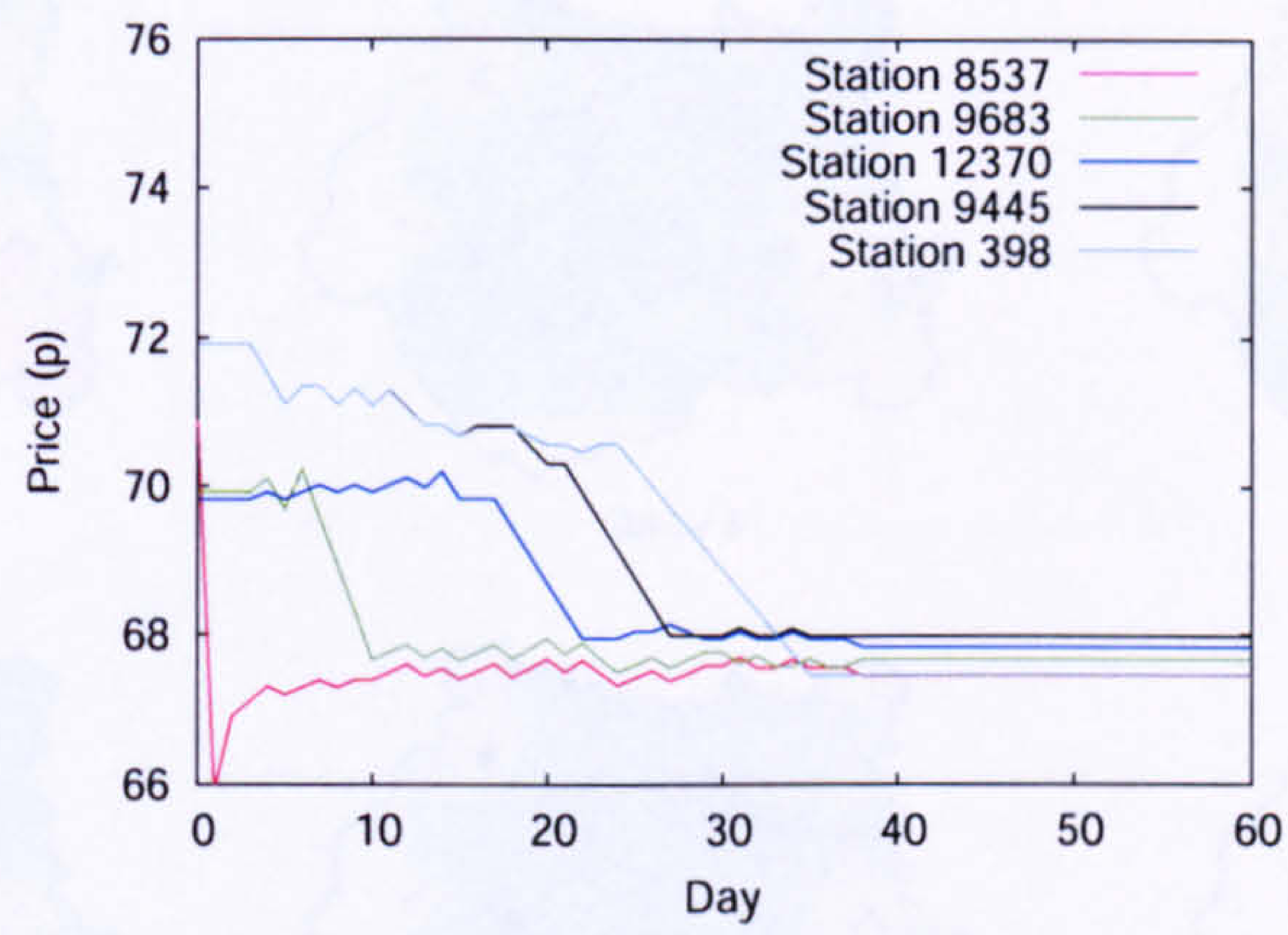


Figure 8.7: Spatial diffusion of prices over time for a price drop of 5p. All stations are initialised with real prices from July 27th.

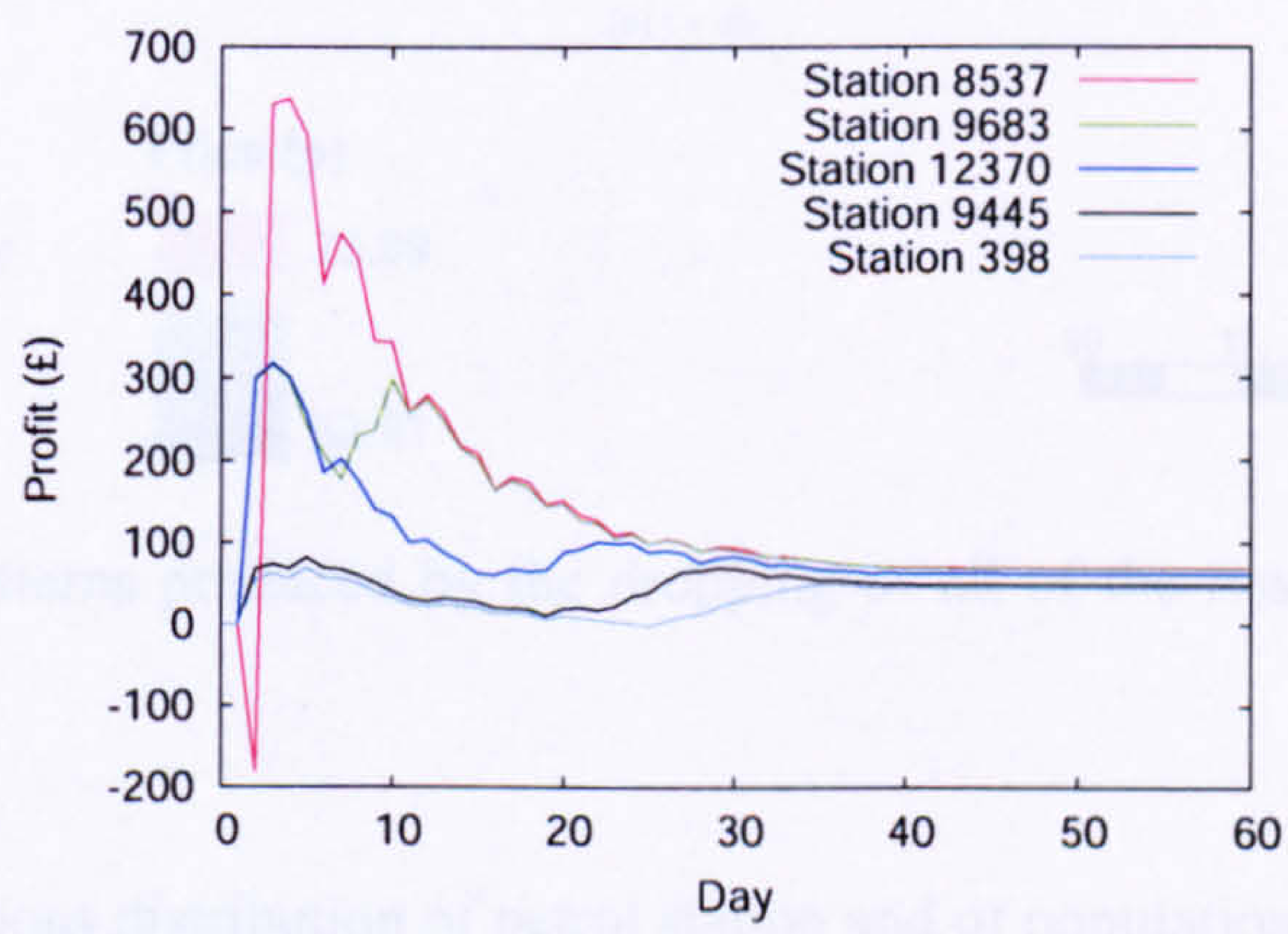
The Esso group of stations were chosen due to their number and spatial distribution around the study area (see Figure 8.9). This would allow examination of the reaction of the system to multiple simultaneous changes.

Figure 8.9 shows that the system responds rapidly to the drop in prices. By day 5, most of the non-Esso stations have dropped their prices by 3p-4p. The Esso stations remain fairly constant in price after the drop. This means that the range of prices decreases with time. For example, the price range is 11p on day 1, 8p on day 5 and 6p on day 40. However, this appears to be the only discernible impact of the multiple drop on the entire study area.

The reaction in price and profit of several different stations (non-Esso) was again analysed (Figure 8.10). These stations are the same as those used in §8.4 and form a transect moving out from the centre of Leeds. The graph showing the station prices as a function of time illustrates the rapid diffusion of the price drop out to these stations over a few days. After this the prices remain constant at values between 69 and 71p. The profit graph shows a sharp fall on day 1 as a result of the Esso stations attracting a large increase in sales with their cheaper petrol. As the price of the non-Esso stations falls they begin to attract back trade. This results in an increase in profits, even though their profit margins fall. The increase in profit continues up to about day 16 even though the price drop stops around day 5. This is presumably an indication that prices in other parts of the system are still adjusting, and affecting the sales around this area. From the maps in Figure 8.9 the changes seem primarily to be in the very cheap regions around Leeds and Halifax at day 5 (indicated by the light blue areas) which gradually increase in price through to around day 20.



(a) Price (p) over time



(b) Profit (£) over time

Figure 8.8: Change in (a) price and (b) profit over time for the selected stations in the real data diffusion experiment.

The work in this section illustrates several important features in the model. Stations react sensibly to price drops by reducing their prices; this shows that dropping the price can have a positive impact on profit. It also demonstrates the stability of the system with all the stations quickly reaching a steady equilibrium.

8.4.4 Summary

The purpose of the diffusion experiments was to examine the reaction of the system under several different conditions. This was achieved by using a constant price and real data to initialise the model. A price drop of 5p was used to force the system to react. The results from each experiment shared several common themes. The new price spread out through the surrounding areas triggering pricing adjustments by neighbouring stations. The end result was a much narrow price range than initially observed in the real data. Finally, all the stations within each simulation experienced a decrease in profit and implemented sensible strategies to regain and maximise profit levels. Al-

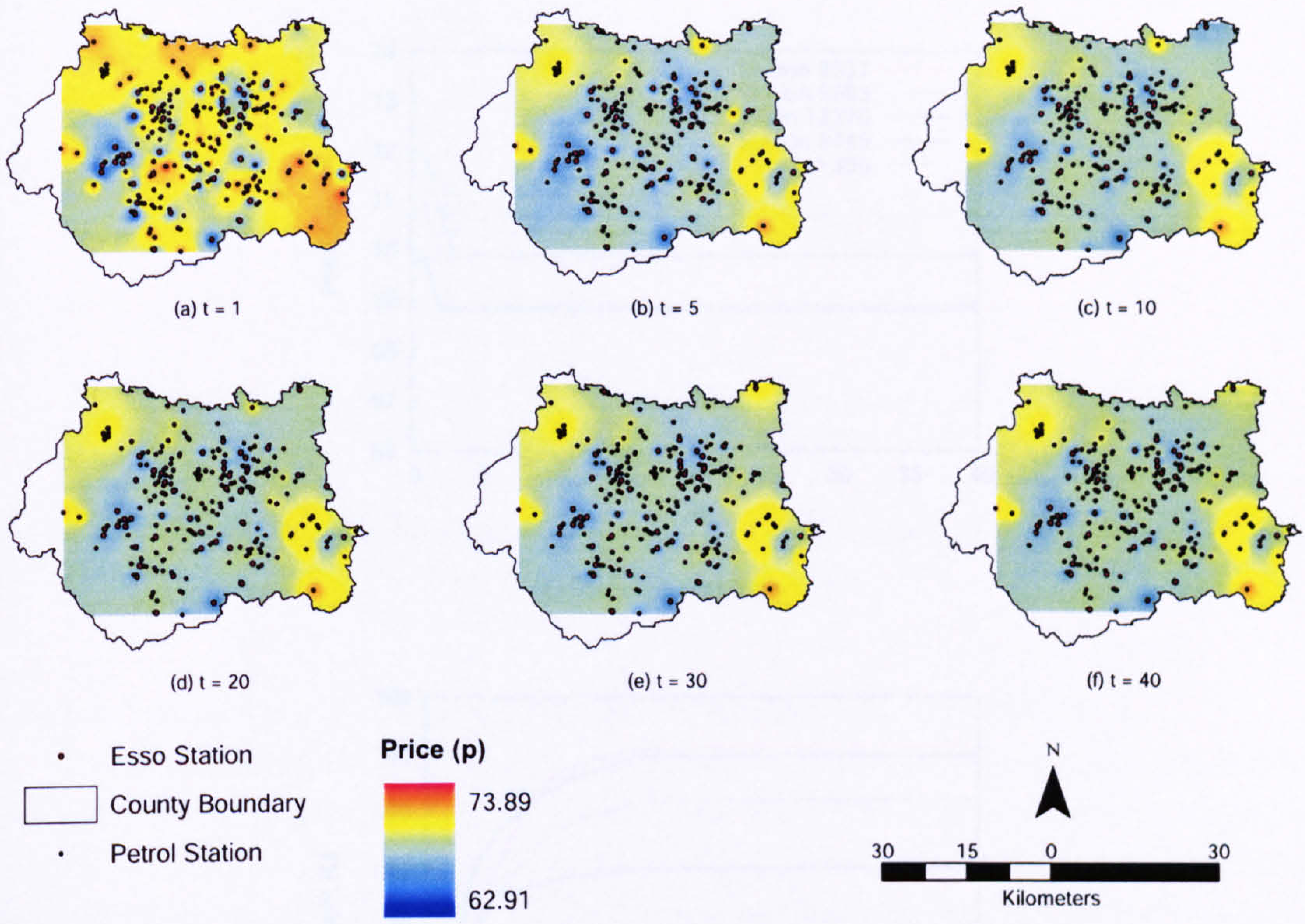


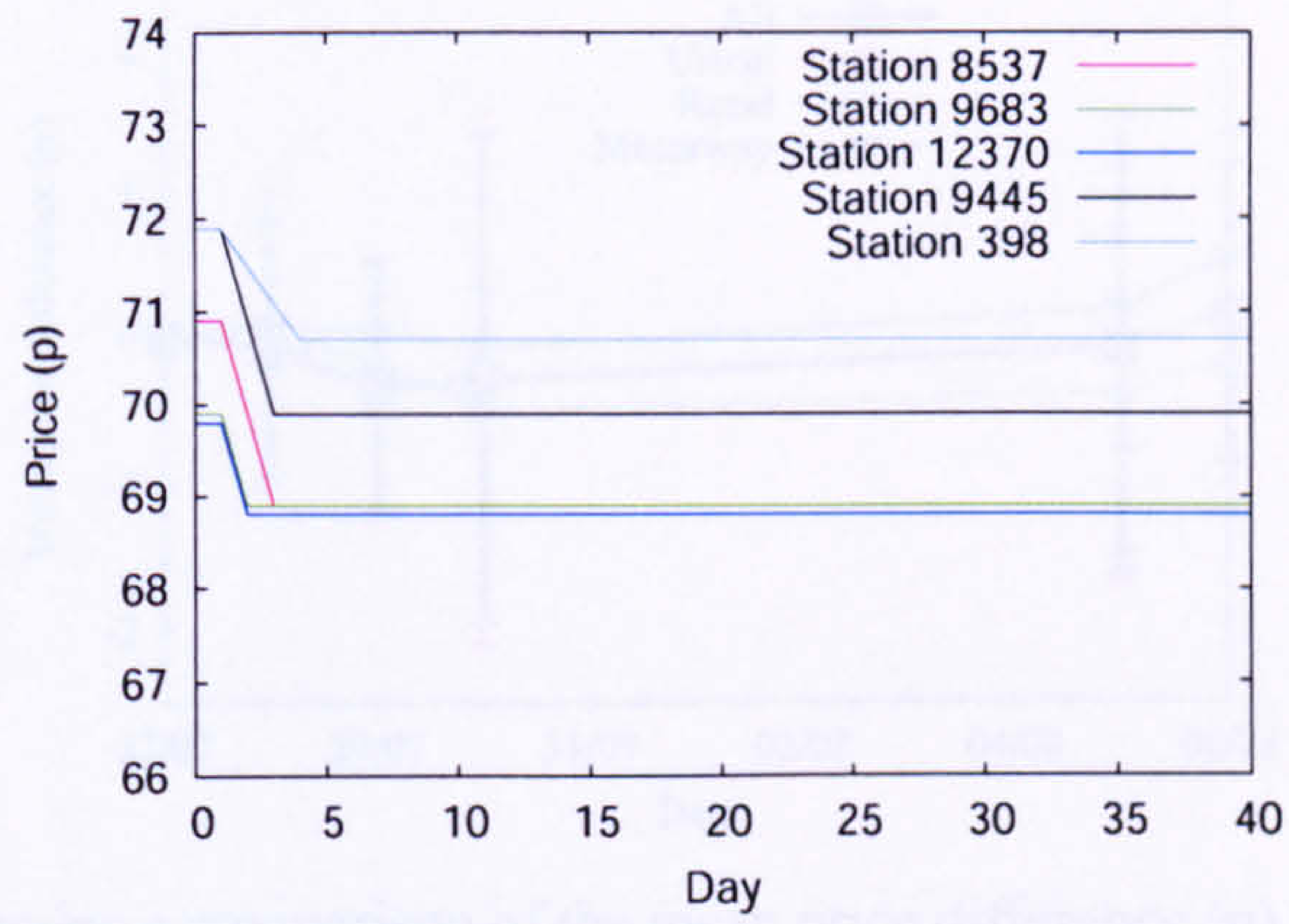
Figure 8.9: Spatial patterns produced by the dropping of all of the Esso stations by 5p on the initialisation day.

though the hetereogeneous distribution of petrol station and of population mean that the diffusion is less regular than the idealised experiments in Chapter 7, the general behaviour is very similar. Experiments in which every station of a particular brand was dropped in price, rather than just one station also showed a similar diffusion behaviour. The model reached a state where most of the prices had dropped far more rapidly because the prices were diffusing out from several different petrol stations. The similar patterns of behaviour observed in all these simulations indicate that the model is robust to changes within the initial conditions and is able to produce sensible results under a variety of different conditions.

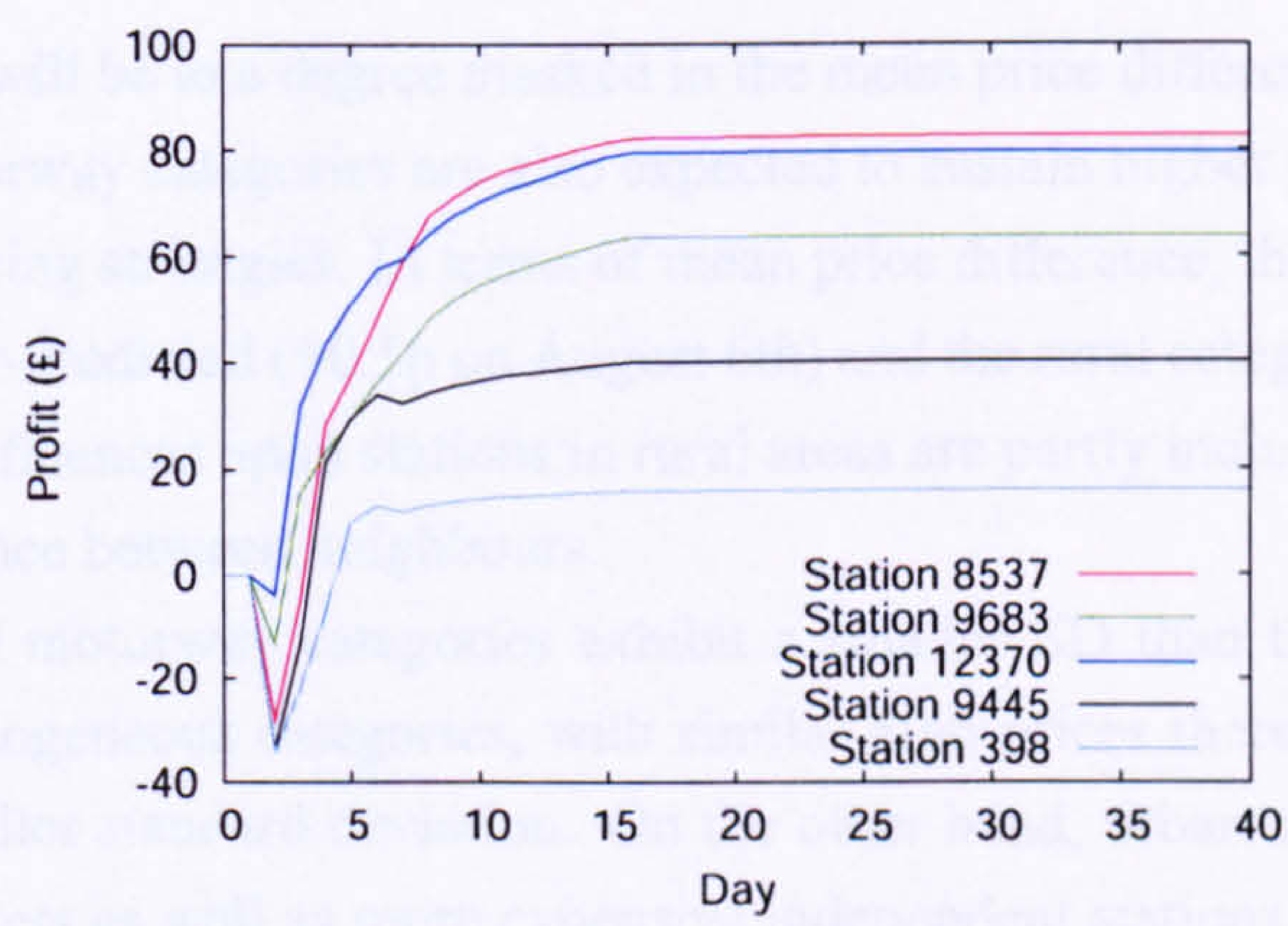
8.5 Analysis with Classifications

In Chapter 4, the real data was divided into geographical and "petrol station type" classifications (see §4.5). Analysis performed using these classifications highlighted price variations within different geographical locations, for example, variations between urban and rural stations. Price differentials were also found within the petrol station type classifications, for example supermarkets were found to be cheaper than multinational and minor stations. By calculating the statistics per classification, we will be able to see how well the network hybrid model performs in each of the individual categories as well as the overall network hybrid model. This may suggest where improvements or adjustments in the rule sets have to be made.

The output analysed in the following sections comes from the network hybrid model initiated



(a) Price (p) over time



(b) Profit (£) over time

Figure 8.10: Change in (a) price and (b) profit over time for the selected stations in the experiment with a 5p initial price drop at all the Esso stations.

with the real data from July 27th for West Yorkshire. The default parameters were used. The mean and SD of the price difference were calculated up to August 6th on days where both real and network hybrid model data were available.

8.5.1 Geographical Classification

In terms of mean price difference, the network hybrid model is performing well in all the classifications, predicting to within ±0.5p. (Figure 8.11). The trend in the prices follows that of the real data. The best performance is found within the urban classification. The prices in this category are being slightly under predicted at the beginning of the simulation (-0.2p on July 30th). However this improves over time and by August 6th the prediction is accurate to +0.1p. This is encouraging as this group accounts for the largest proportion of stations in the study area (84%). The size of the group is also reflected in the size of the standard deviation. By July 30th, this is the largest and remains so until day 10. In a group the size of the urban category, it is expected that there will be

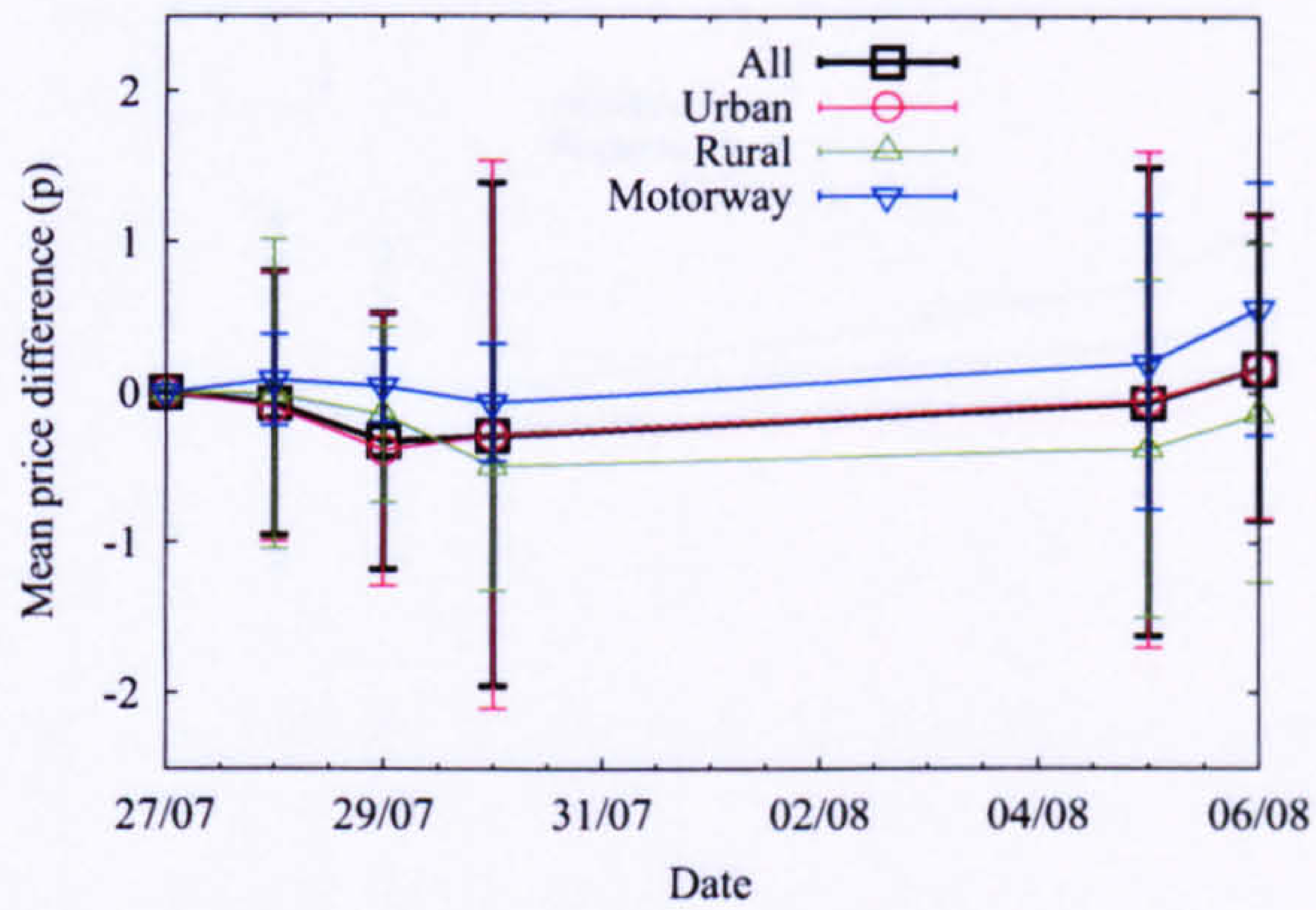


Figure 8.11: Graph showing a comparison of the mean price difference (p) within the geographical classifications over time. SD is represented by the vertical bars.

"extreme" values that will be to a degree masked in the mean price difference.

The rural and motorway categories are also expected to sustain higher prices due to their location and suspected pricing strategies. In terms of mean price difference, the motorway category is consistently being over-predicted (+0.5p on August 6th) and the rural category under priced (-0.2p on August 6th). The influences upon stations in rural areas are partly included through population distributions and distance between neighbours.

Both the rural and motorway categories exhibit a smaller SD than the urban classification. As they are more homogeneous categories, with similar high prices there is less variation in the price and hence a smaller standard deviation. On the other hand, urban areas often contain very competitive supermarkets as well as more expensive independent stations and therefore display a larger range of prices.

8.5.2 Petrol Station Type Classification

The largest category within the petrol classifications are the multinationals. They comprise 265 stations (51%). Figure 8.12 shows that the mean price difference of this category mimics the pattern of the real data with a higher degree of accuracy than the supermarket and minor categories. In §8.5.1, the classification containing the most stations (urban) had the largest SD. This was also found to be the case with the multinational category (-0.3p on July 30th and 0.1 on August 6th). The network hybrid model over-predicted both the supermarket (1.0p on July 30th) and minors (0.3p on July 30th) categories. The SD of the supermarkets is less than that of the minors category. This is probably due to the fact that this category is the smallest: there are 31 supermarkets (6%) and 221 minors (43%). Analysis of the real data in Chapter 4 showed that the supermarkets were the most tightly and competitively priced. The poor prediction of the supermarkets clearly highlights that this behaviour is not being accounted for. Assignment of an individual rule set for the supermarket category may both improve their performance and that of the overall model.

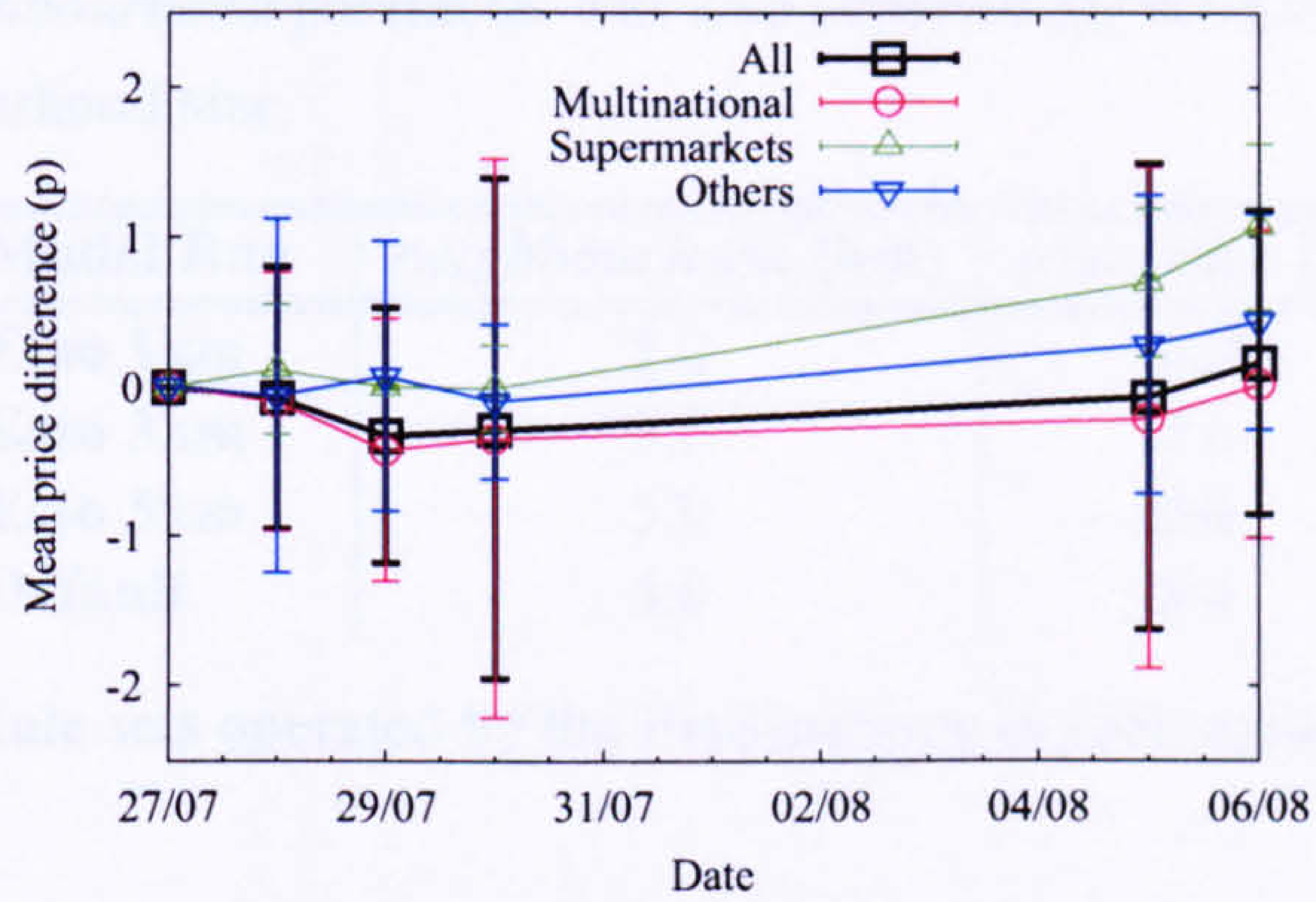


Figure 8.12: Graph showing a comparison of the mean price difference (p) within the different petrol station classifications over time. SD is represented by the vertical bars.

8.6 Further Experimentation with Rule Sets

Geographical and petrol station type classifications were used in the previous sections to examine how the network hybrid model performed in each of the different categories. The motorway and supermarket categories were both identified as performing poorly. It was suggested that they may benefit from individual rule sets. The factors affecting motorway stations are more complicated, as research by McFarland (2003) determined. Their rules are closely linked to the volume of traffic on the motorway and they do not generally compete with the local non-motorway petrol stations (this is with the exception of the Esso stations that were found to match competitors within a 3 mile neighbourhood). To test out whether assigning individual categories unique rule sets will improve the overall network hybrid model performance, experiments will be conducted with separate rule sets for the supermarket and Esso stations.

8.6.1 Esso Price Watch Revisited

In Chapter 5, the basic agent model was run to examine the effectiveness of imposing the Esso Price Watch (see §5.8.2). To re-cap, the Price Watch pledge states that Esso will match any price within the surrounding 3 miles (5km). Having this published rules information provides a useful opportunity to assign an accurate rule set and determine its impact on the model performance. The results with the agent model indicated that the size of the neighbourhood did not make a notable difference to the mean and SD. It was largely inconclusive as to whether this meant that the Price Watch policy was ineffective or the system was insensitive to changes within this parameter.

The network hybrid model was used to re-visit the Esso case-study and reassess the impact of the Price Watch pledge within a more realistic profit maximising model. To enable comparison with earlier simulations, the experiments were run on West Yorkshire. All the stations were initialised with the real data from July 27th. The default parameters (Table 8.2) were assigned to every station with the exception of the *overprice* parameter. Within the Esso stations rule set, this was set to 0.0p to force the Esso stations to drop their price to match the lowest in its neighbourhood. This reflects the Price Watch pledge. The main differences to the default rules are outlined in Table 8.3. The *neighbourhood* parameter was also adjusted for the Esso rules to determine the impact of the neighbourhood size.

Model Run	*neighbourhood* (km)	*overprice* (p)
Esso 1km	1.0	0.0
Esso 3km	3.0	0.0
Esso 5km	5.0	0.0
Default	5.0	5.0

Table 8.3: Rule sets operated by the Esso stations in each model simulation.

Figure 8.13 clearly shows that assigning the Esso stations their own Price Watch rules produced a worse performance than initialising all the stations with the same rule set. The mean price difference and SD for each of the Esso runs increased over time (Figure 8.13 (a)) with each of the Esso simulations being 0.6-0.9p under priced by August 6th. This compares to the 0.1p under price that the default run produced at the same point. Increasing the neighbourhood of the Esso stations produced an increasing error. The experiment using a 5km Esso neighbourhood (the equivalent of the 3 miles stated in the Price Watch Pledge) gave the worst performance along with the Esso 1km experiment, by August 6th it was under predicting by 0.9p.

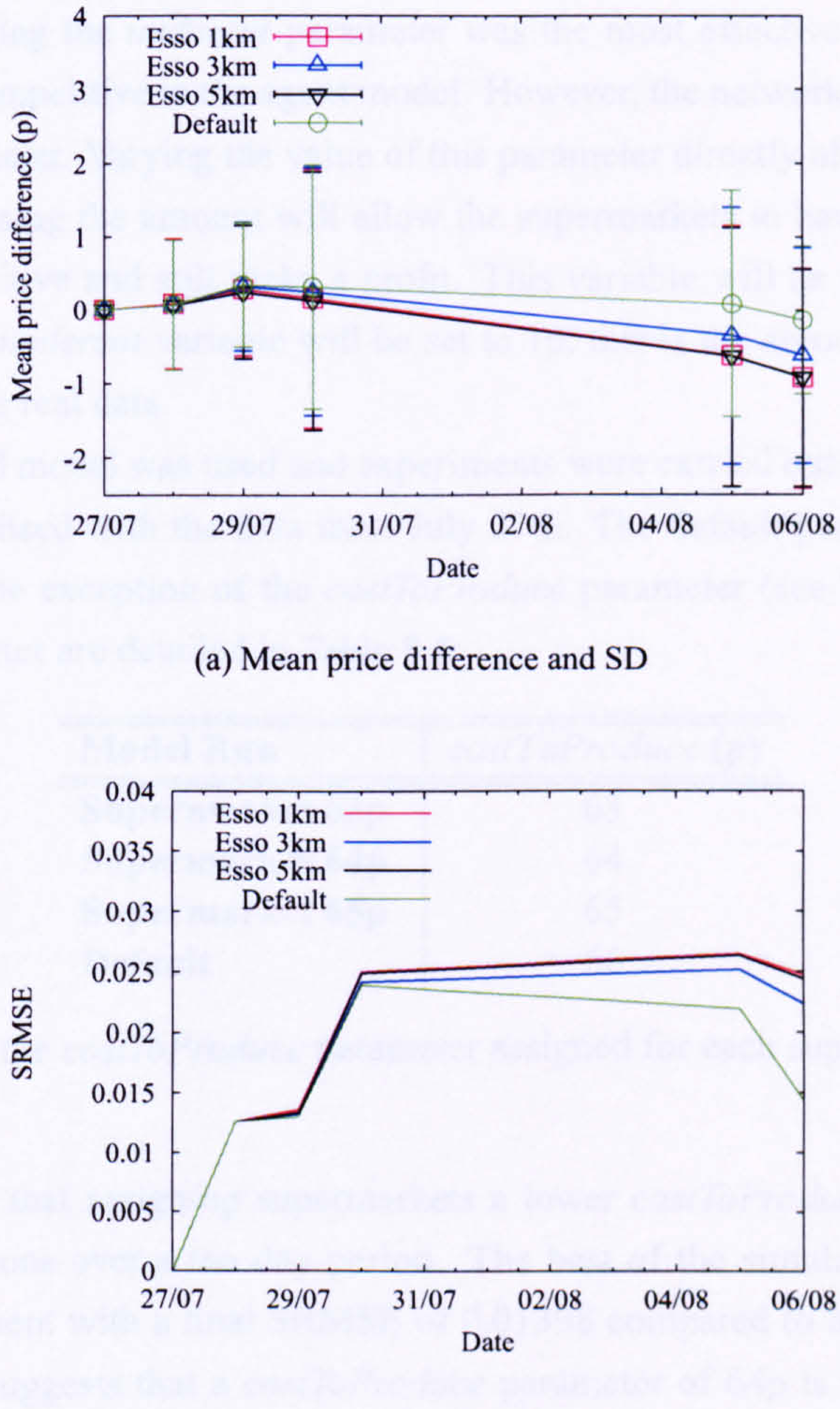
The SRMSE over time (Figure 8.13 (b)) shows that each of the Esso simulations were performing well until July 30th (day 3 of the simulation). After this day, the errors become larger. The effects of the Esso Price Watch would take a few days to impact on the system. By August 6th (day 10), the Esso rule set is exerting an influence on the behaviour within the system. With the *overprice* parameter set to 0.0p, this tends to reduce the price of the Esso stations and hence the mean price. With these rules the system tends to reduce prices too aggressively so the agreement with the real data is less good.

Table 8.4 shows the mean profit for all the Esso and non-Esso stations on day 10. Esso stations make more profit than the non-Esso stations, even in the default simulation. This suggests that the Esso stations are geographically well located. Further discussion of this point is given in §8.9. Since the Esso rules decrease the mean price, they also decrease the overall profitability of all the stations. However, in relative terms the Esso stations makes about 2.7 times more profit than the non-Esso stations under the Price Watch rules (Esso 5km) compared with 1.6 times more for the default rules. This suggests that although an aggressive pricing policy might affect a stations overall profit, it could affect competitors' profit levels more severely.

Mean Model Run	Mean Profit Esso (£)	Profit Non-Esso (£)
Esso 1km	209	63
Esso 3km	183	101
Esso 5km	176	65
Default	285	176

Table 8.4: Profits at Esso and non-Esso stations for simulations using separate Esso rules.

The default parameters selected to run the network hybrid model with were determined in Table 8.2. These were partly based on experimentation with all the stations operating identical



(a) Mean price difference and SD

(b) SRMSE

Figure 8.13: Comparison of (a) the mean price difference and standard deviation (SD is indicated by the vertical bars) and (b) the SRMSE over time between the real and network hybrid model data for experiments with the Esso rule set. The study area is West Yorkshire and the network hybrid model was initialised with data from July 27th.

rules. Assigning the Esso stations a unique rule set introduces new behaviour into the system. It seems reasonable that different default rules may be needed when running the network hybrid model with extra rule sets for particular categories of station. The other parameters may need to be slightly adjusted to account for the new Esso rules. This may partly explain why the simulations here underpredict the price when using the Esso rules.

8.6.2 Supermarkets

In the analysis of the real data in Chapter 4, supermarkets were identified as being both cheaply priced and competitive (see §4.6.2). This category was experimented with in Chapter 5 (§5.8.1) to see whether an individual rule set would improve the overall model performance. The *undercut* parameter was varied to prevent other stations from pricing more competitively. The results showed that assigning the *undercut* parameter values of between 1.5p - 5p improved the performance of the agent model. Varying the *undercut* parameter was the most effective way the supermarkets could be made more competitive in the agent model. However, the network hybrid model contains a *costToProduce* parameter. Varying the value of this parameter directly affects the profit margins of the stations. Decreasing the amount will allow the supermarkets to have a larger scope to cut their prices, be competitive and still make a profit. This variable will be used instead of the *undercut* parameter. The *undercut* variable will be set to 1p, this is the amount calculated in §8.2.4 based on analysis of the real data.

The network hybrid model was used and experiments were carried out on West Yorkshire. All the stations were initialised with the data from July 27th. The default parameters were assigned to every station with the exception of the *costToProduce* parameter (see Table 8.2). The values assigned to this parameter are detailed in Table 8.5:

Model Run	*costToProduce* **(p)**
Supermarket 63p	63
Supermarket 64p	64
Supermarket 65p	65
Default	66

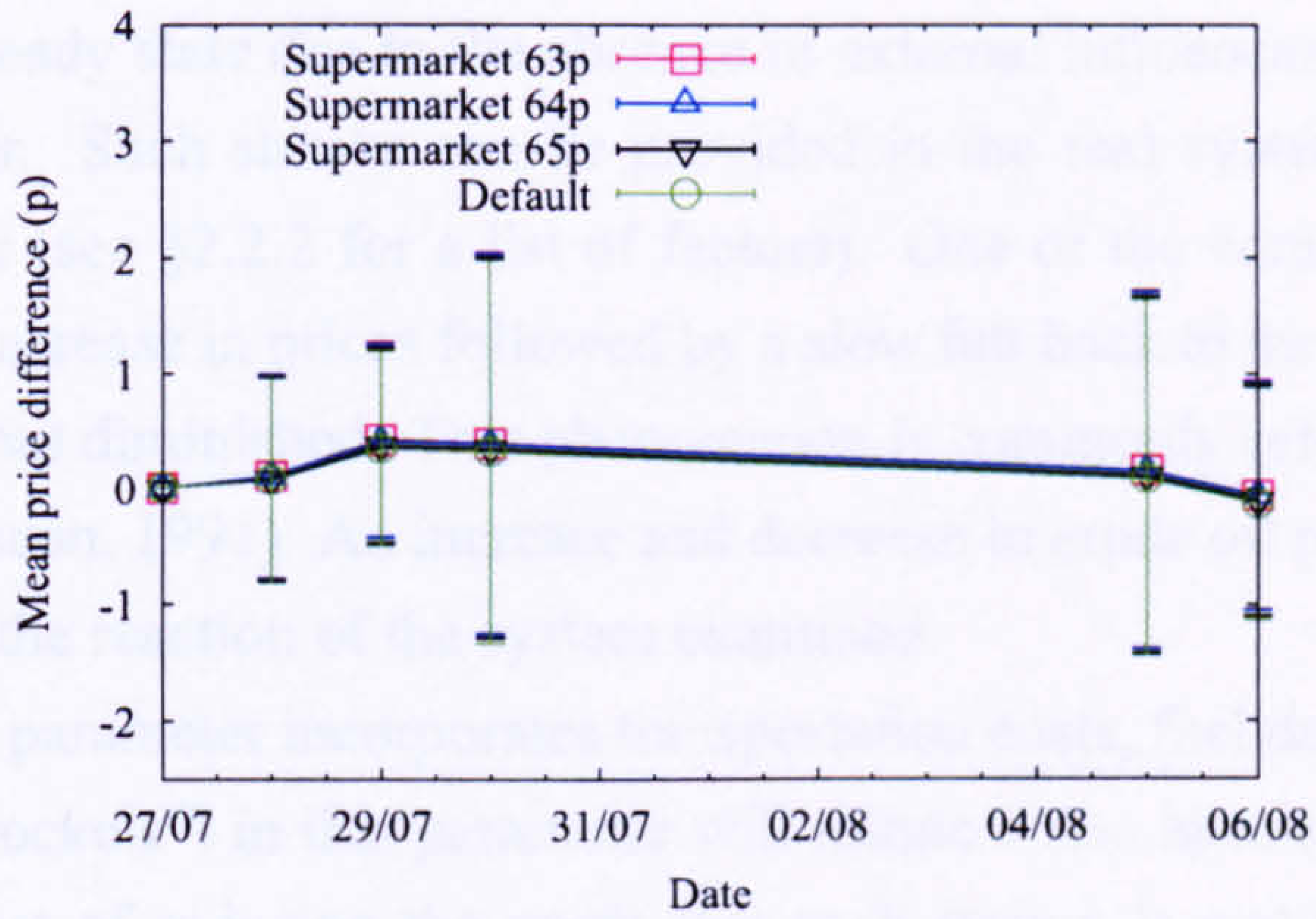
Table 8.5: Values of the *costToProduce* parameter assigned for each supermarket simulation.

Figure 8.14 shows that assigning supermarkets a lower *costToProduce* value has very little impact on the simulations over a ten day period. The best of the simulations was with the supermarket 64p experiment with a final SRMSE of 0.01398 compared to the value of 0.01430 for the default run. This suggests that a *costToProduce* parameter of 64p is the optimal value. The supermarket 63p gave a SRMSE of 0.01403 and the supermarket 65p gave 0.01414. In practice these SRMSE values are so close as to make negligible impact on the overall model performance over ten days. Over longer periods these small differences may become more significant.

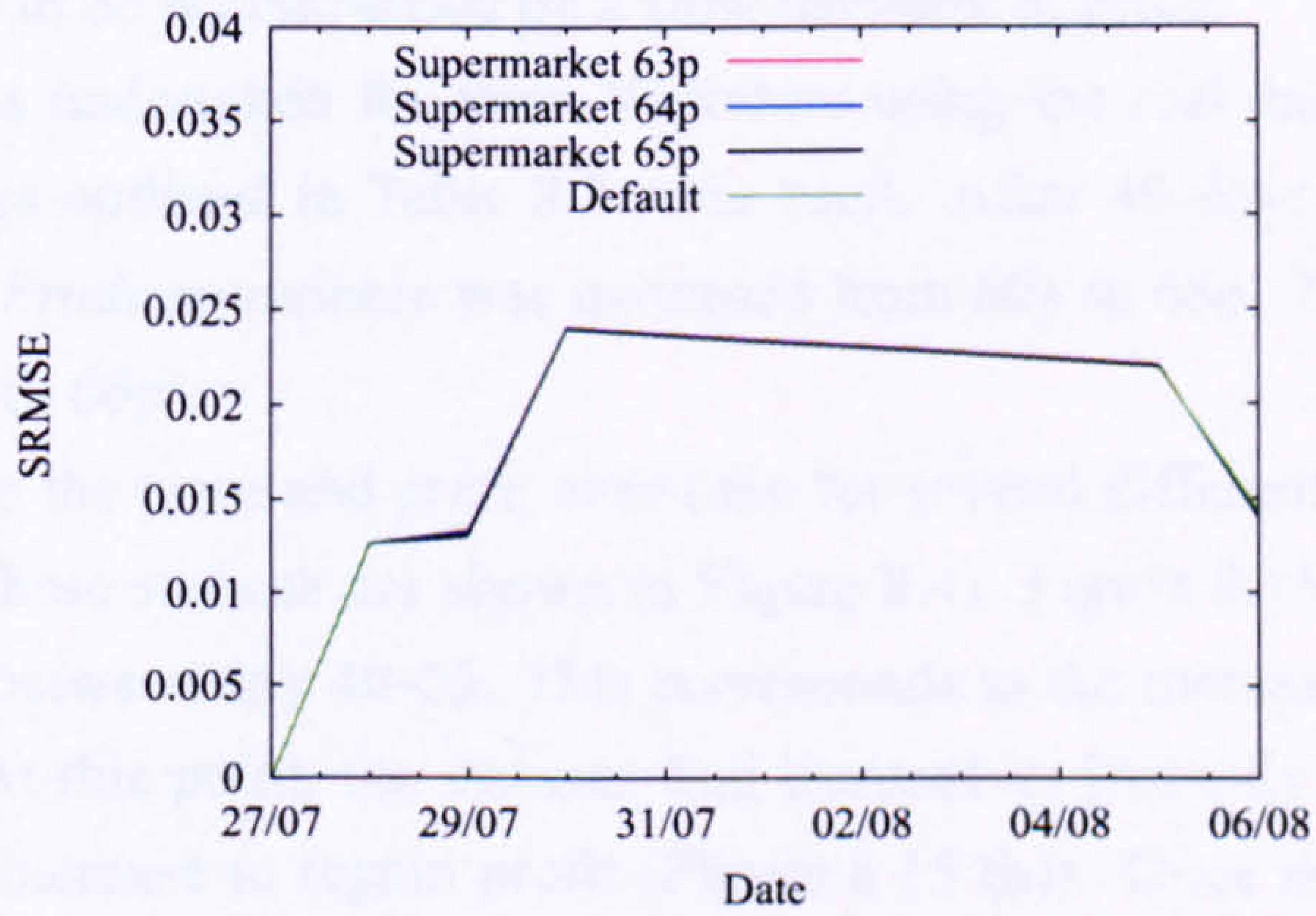
Table 8.6 shows the mean profit for all the supermarket and non-supermarket stations on day 10. The supermarket stations make more profit than the non-supermarket stations even with the default rules. This could be because supermarkets are generally located in urban areas with a large potential customer base nearby. Decreasing the *costToProduce* for supermarkets makes only a small difference to the overall pricing, therefore the non-supermarket stations make a comparable profit to the default rules. The profit levels for the supermarket stations obviously increase because their *costToProduce* values are less.

Mean Model Run	**Mean Profit Supermarket (£)**	**Mean Profit Non-Supermarket (£)**
Supermarket 63p	622	193
Supermarket 64p	521	192
Supermarket 65p	412	188
Default	297	187

Table 8.6: Profits at supermarket and non-supermarket stations for simulations using separate supermarket rules.



(a) Mean price difference and SD



(b) SRMSE

Figure 8.14: Comparison of (a) the mean price difference (SD is indicated by the vertical bars) and (b) the SRMSE over time between the real (day 10) and network hybrid model data for different model runs with the supermarket data. The study area is West Yorkshire (Model initiated on July 27th). The real data differences are plotted for comparison.

The assignment of a unique rule set for supermarkets introduced new behaviour into the system. As with the assignment of an individual rule set for Esso, this did not result in any significant improvement to the network hybrid model performance. The case was put forward with the Esso stations that when introducing new behaviour, new default parameters may have to be derived. A method for obtaining these parameters will be presented in Chapter 9.

8.7 External Influences/Shocks

§8.4 examined the reaction of the system to a sudden price decrease in one or more stations in several initialisation scenarios. The outcome of these experiments was a general decrease in price as the effects of the new, lower price diffused across the study area. The system reached a steady solution and remained in this state with minor fluctuations. It was concluded that the system did not deviate from this steady state due to the absence of external influences to "shock" the system into different behaviour. Such shocks can be provided in the real system by increases in fuel duty or crude oil prices (see §2.2.2 for a list of factors). One of the common reactions to these influences is the rapid increase in prices followed by a slow fall back to the original price after the effect of the influence has diminished. This phenomenon is commonly referred to as the "rockets and feathers" effect (Bacon, 1991). An increase and decrease in crude oil prices will be mimicked within this section and the reaction of the system examined.

The *costToProduce* parameter incorporates transportation costs, fuel duty and crude oil prices. Forcing an increase ("rockets") in this parameter will mimic a rise in one or all of these factors. This will have the effect of reducing the profit that each station is making, precipitating price adjustments in order to maintain profits. Decreasing the *costToProduce* ("feathers") will allow the stations to begin to act more competitively i.e. undercutting each other to maximise their profit levels. This is expected to be accompanied by a slow decrease in price.

The experiment was undertaken for West Yorkshire using the real data from July 27th. The default parameter values outlined in Table 8.2 were used. After 40 days (as the model reached equilibrium) the *costToProduce* variable was increased from 66p to 68p. This was then decreased after 80 days from 68p to 66p.

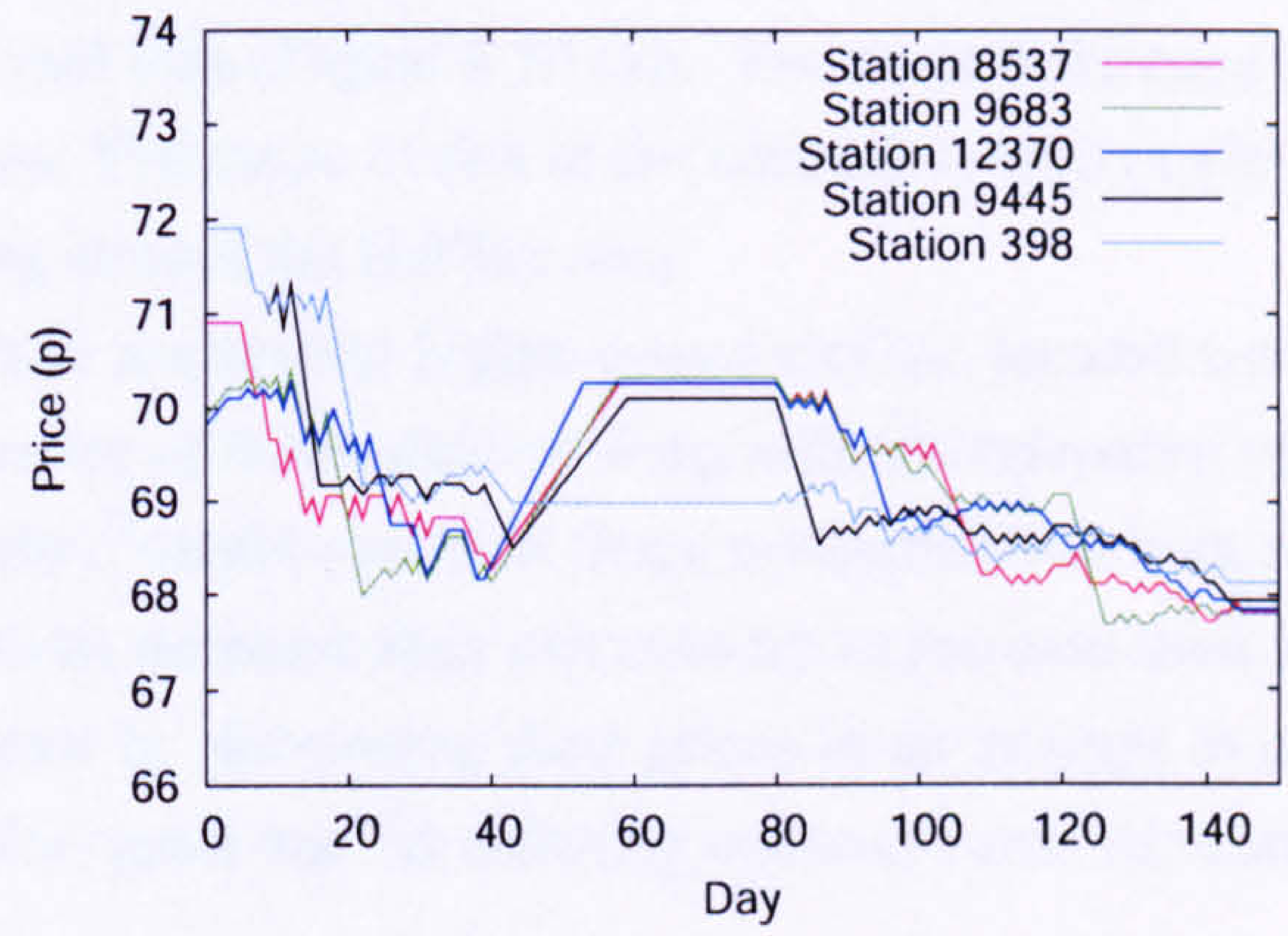
Figure 8.15 presents the price and profit over time for several different stations in West Yorkshire (the positions of these stations are shown in Figure 8.4). Figure 8.15 (a) shows the increase in price at the stations between day 40-60. This corresponds to the increase in the *costToProduce* parameter at day 40. At this point, the stations find themselves instantly making less profit and take action via a price increase to regain profit (Figure 8.15 (b)). Once this course of action has been taken by all of the stations and profit is being made, alternative strategies (decreasing the price) are implemented to maximise this amount. This behaviour is gradually transmitted across the study area and continues until an equilibrium is reached (days 60 - 80).

On day 80, the *costToProduce* parameter is decreased by 2p to 66p. This results in a sharp increase in profit (Figure 8.15 (b)) followed by a gradual decrease in price (days 80 - 120). This decrease in price is a result of the stations employing competitive measures to attract more consumers, i.e. undercutting each other. The profit levels slowly decrease with the price, but the price remains above the *costToProduce* value and so the stations remain in profit. The price finally settles to approximately 68p - 69p. This is roughly the same variation in price that the system showed before the increase in *costToProduce*, on day 39 the price varied between 68 - 69.5p.

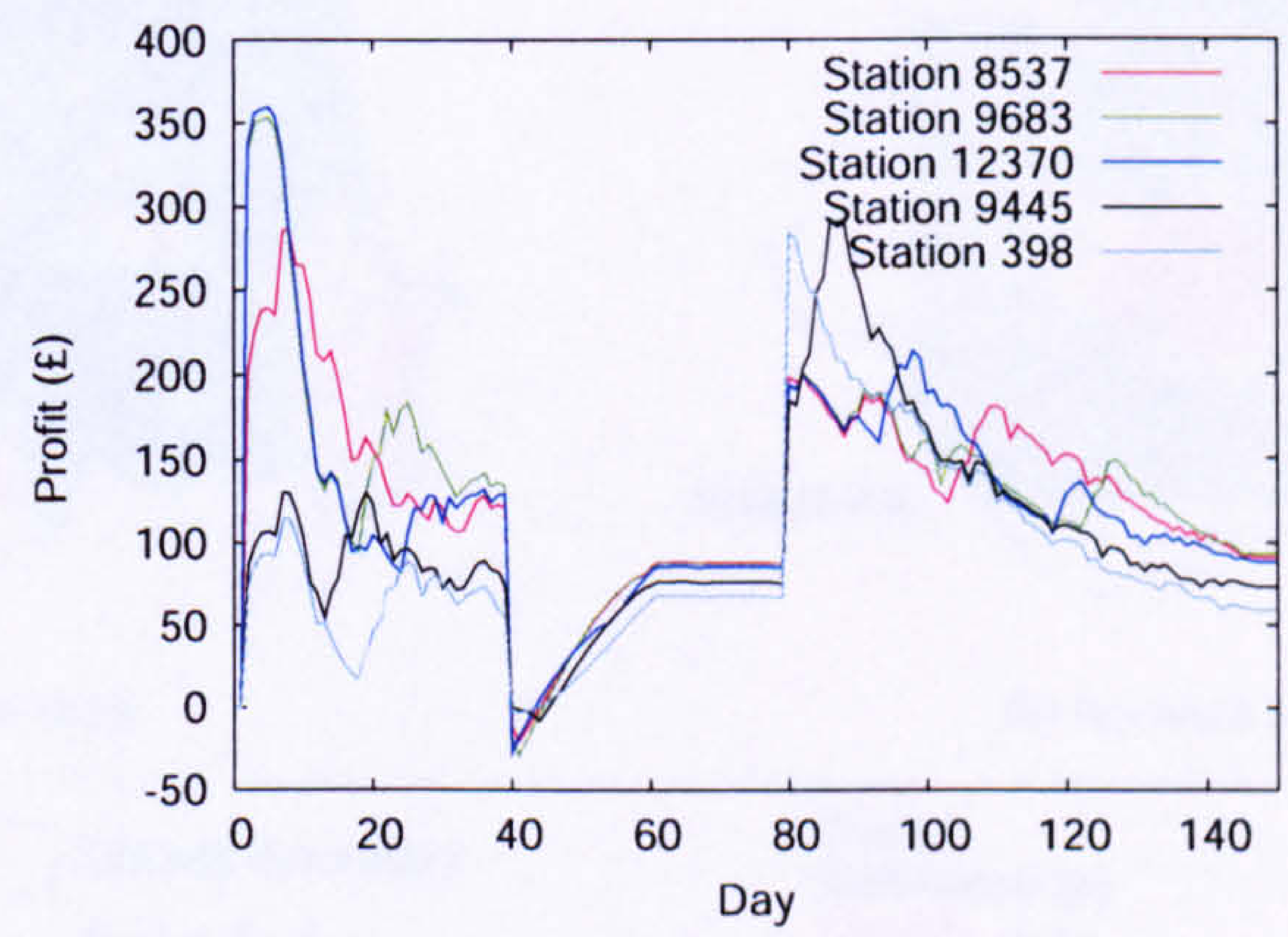
8.8 Further Assessment of Model Performance

Experimentation and analysis using the network hybrid model has to date, been undertaken by using the West Yorkshire study area initialised with the prices from July 27th. This is not a comprehensive test of the robustness of the model. Initialising the model with the second half of the data set and using a different region provides an opportunity to assess how well the model performs under different conditions.

In Chapter 4, the data was described and divided into two clear temporal sections (July 27th



(a) Price over time



(b) Profit over time

Figure 8.15: Mean price (a) and profit (b) plotted against time for simulations of the "rockets and feathers" effect using the West Yorkshire data.

and August 19th data sets) and two geographical regions (West Yorkshire and the Yorkshire region). §8.8.1 will investigate how well the model performs using data from August 19th and §8.8.2 will assess the performance on the whole Yorkshire region rather than just West Yorkshire.

8.8.1 19th August Data Set

The network hybrid model was initialised with the real data from 19th August for West Yorkshire and run using the default model parameters (Table 8.2). The 19th August data spans a slightly longer period than the 27th July data (12 days instead of 10) and is on average, 2p higher. This is due to an increase in the price of crude oil between 6th August and 19th August. This price difference will be accounted for by increasing the *costToProduce* variable from 66p to 68p. The network hybrid model output will be compared against the real data on day 12 (September 1st). This choice is determined by availability of data and provides some consistency with the analysis methods used for 27th July.

Figure 8.16 (b) shows that in general, the model is re-creating similar rural-urban patterns to those seen within the real data (Figure 8.16 (a)). The main difference is within the magnitude of the price differentiation. The cause of this in the network hybrid model data is what appears to be a "price war" occurring around the Halifax area.

In the real data, there are several higher priced stations located next to stations that are a few pence lower. The presence of these stations along with the operation of the rule set and strategies within the network hybrid model creates a fierce competition to both retain and maximise profit. The higher priced stations decrease their prices to try to increase their sales and profit. The lower priced stations then react by decreasing their prices in an attempt to stop their profit losses. By day 12, the effect of this "price war" is diffusing out to the area surrounding Halifax.

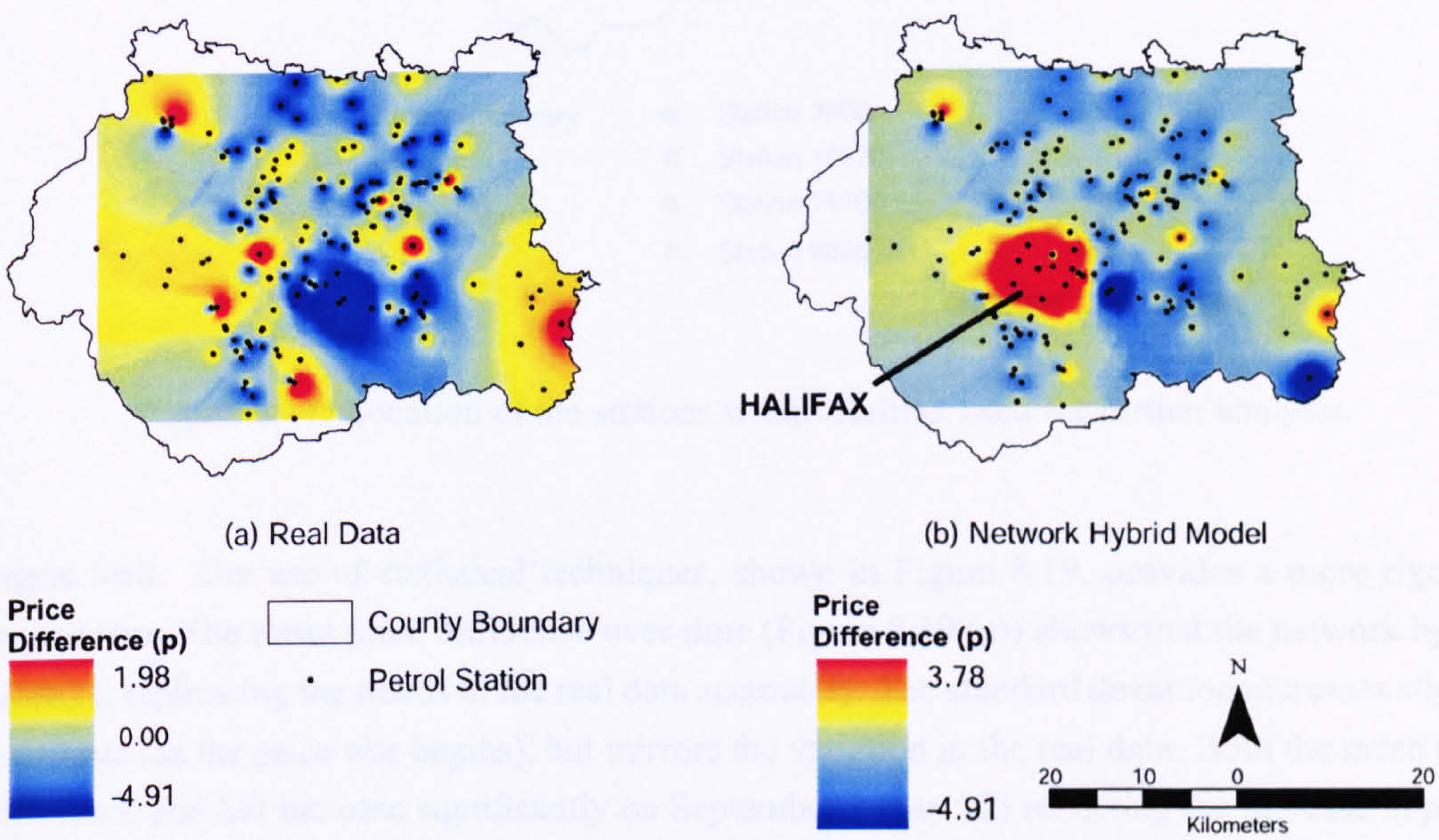


Figure 8.16: Price difference between the real (day 10) and model data for the network hybrid model (b) for West Yorkshire. The difference between the real data (August 19th and September 1) (a) is included for comparison.

What is causing this "price war"? A set of 4 adjacent stations forming a transect across Halifax were examined more closely (Figure 8.17). Figure 8.18 shows price and profit levels at these stations over the first 20 days of the simulation. Station 16070 initially has a high price but low profit. In an attempt to increase its profit, it begins to decrease its price (Figure 8.18 (a)). This ploy works as over the next 6 days the station sees an increase in profit of over £1000. At this point, station 16070 has affected the surrounding stations sufficiently that they react by dropping their prices in an attempt to compete. This results in an increase in their profit and a corresponding decrease in profit at station 16070 (Figure 8.18 (b)). Eventually, all the stations get close to the *costToProduce* value (66p) and therefore cannot reduce their prices further without going into negative profit. This produces a high mean price difference between the real and model data.

Figure 8.16 is a qualitative assessment of the performance of the network hybrid model on one day. Based on this evidence alone, the network hybrid model is not seemingly predicting the

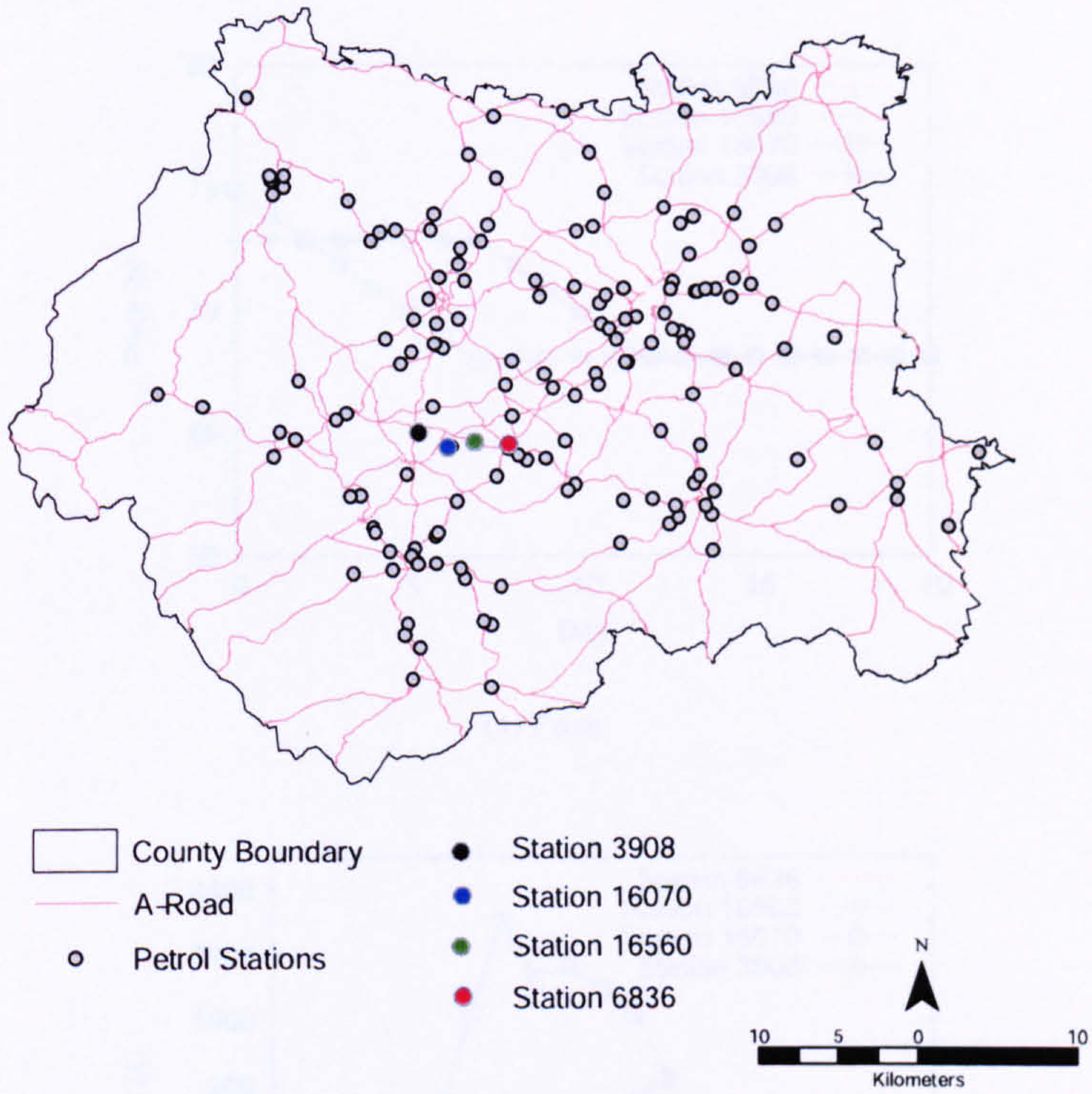
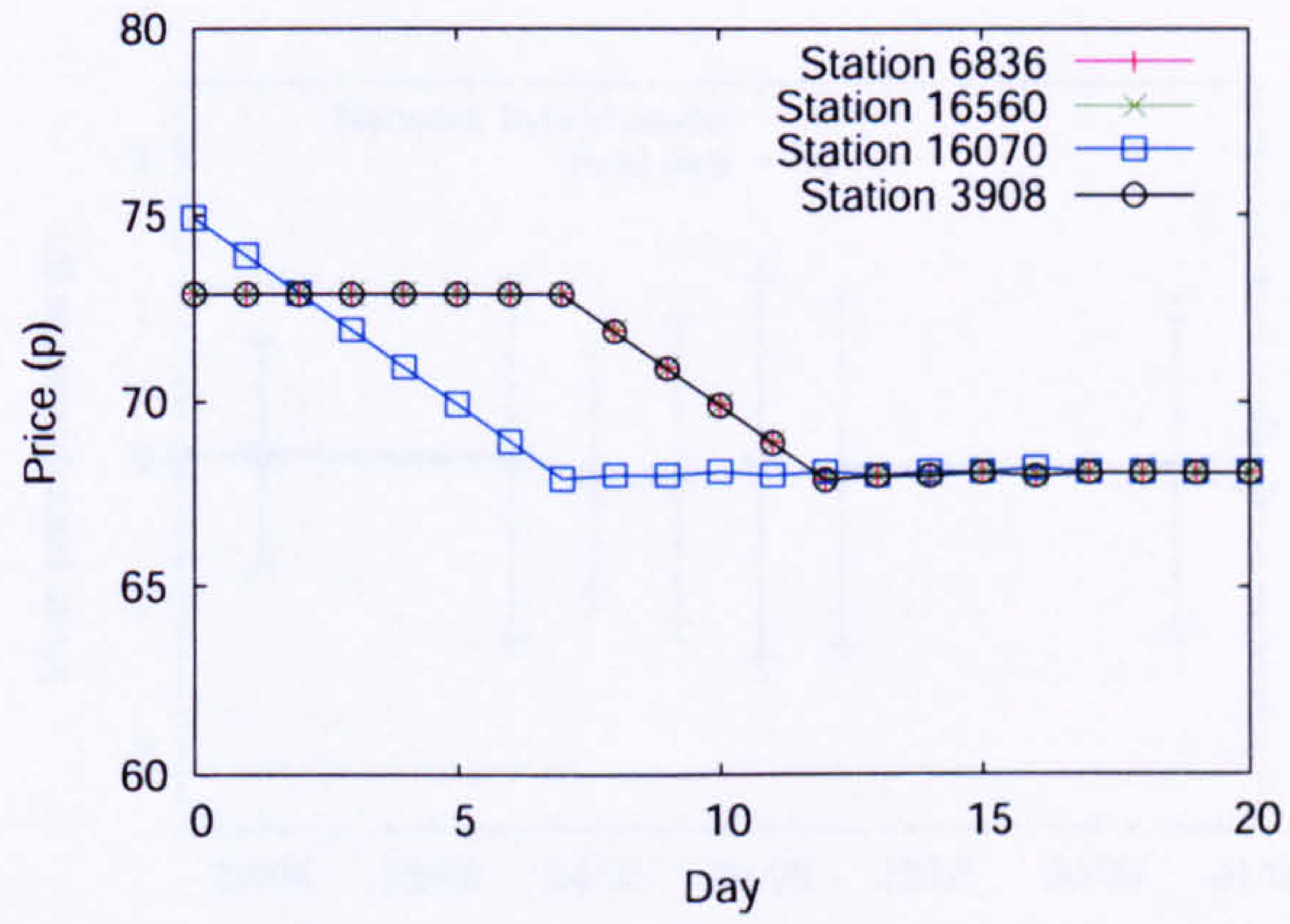


Figure 8.17: Location of the stations within Halifax used for further analysis.

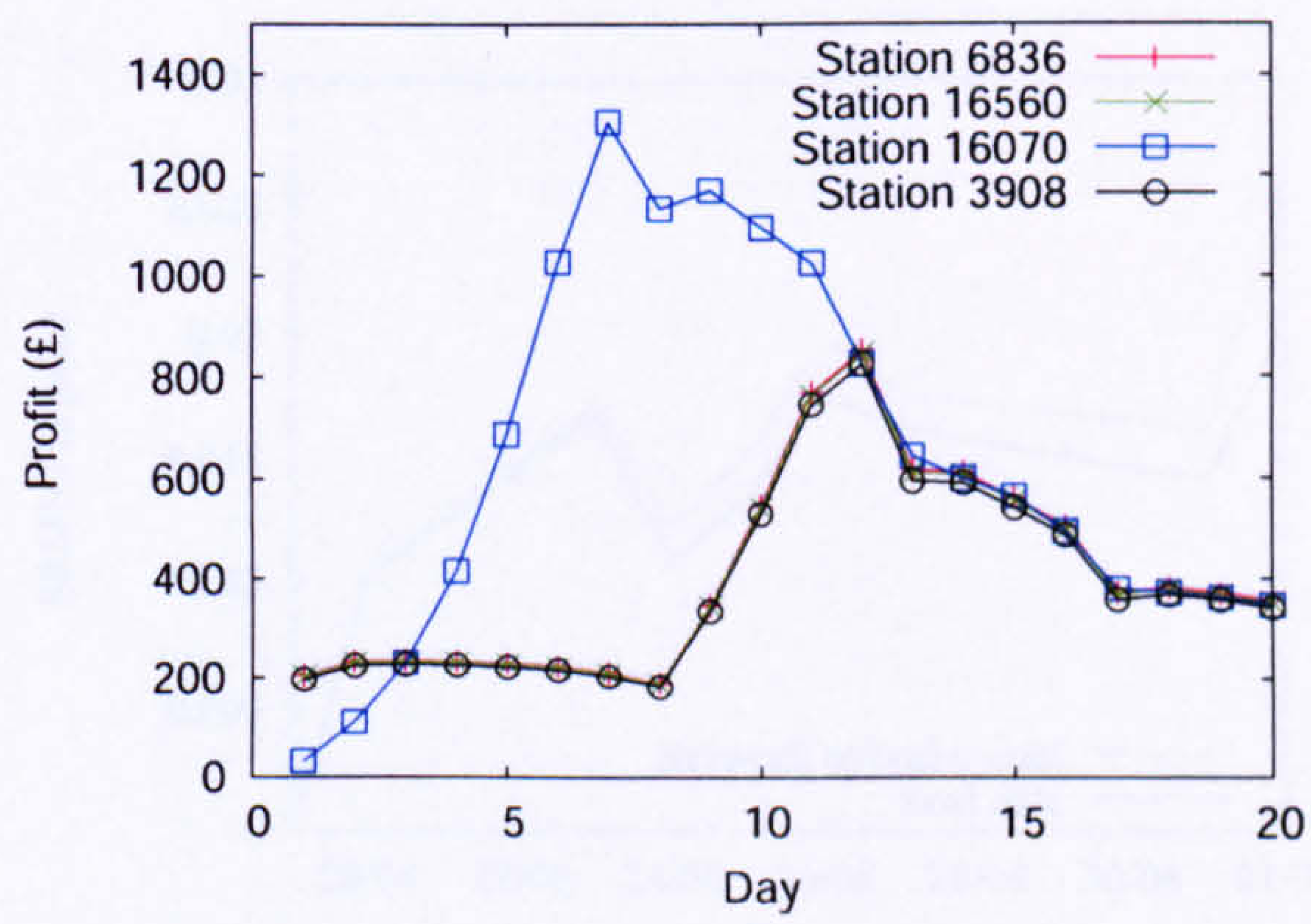
prices well. The use of statistical techniques, shown in Figure 8.19, provides a more rigorous assessment. The mean price difference over time (Figure 8.19 (a)) shows that the network hybrid model is replicating the trends of the real data accurately. The standard deviation increases slightly over time (as the price war begins), but mirrors the variation in the real data. Both the mean price difference and SD increase significantly on September 1 (day 12) reflecting the increase in prices around the Halifax area.

Similar trends are also found within the SRMSE (Figure 8.19 (b)). Up to the 30th August, the SRMSE varies between 0.01 and 0.02 with both the model and constant data displaying similar results. On the 1st September both the constant and model SRMSE suddenly increase, this change in both models indicates movement in price within the real data. This is evident by assessing the decrease in mean price of the real data between August 30 - September 1 (Figure 8.19 (c)). Figure 8.19 (c) also shows that the network hybrid model continually over-predicts the real data price changes throughout the simulation.

The second half of the data set has been used to test the performance of the network hybrid model under slightly different conditions to those used to tune it. The only difference in parameters was an increase in the *costToProduce* to account for a real increase in costs over this period. The network hybrid model showed a good performance, reproducing the spatial variations existing within the real data. In terms of mean price difference and SD (price variation), the results were also promising with the network hybrid model predicting accurately to within ± 0.1p. Although the time period was slightly different, this case still had the same underlying spatial distribution of petrol stations as the July 27th case.



(a) Price



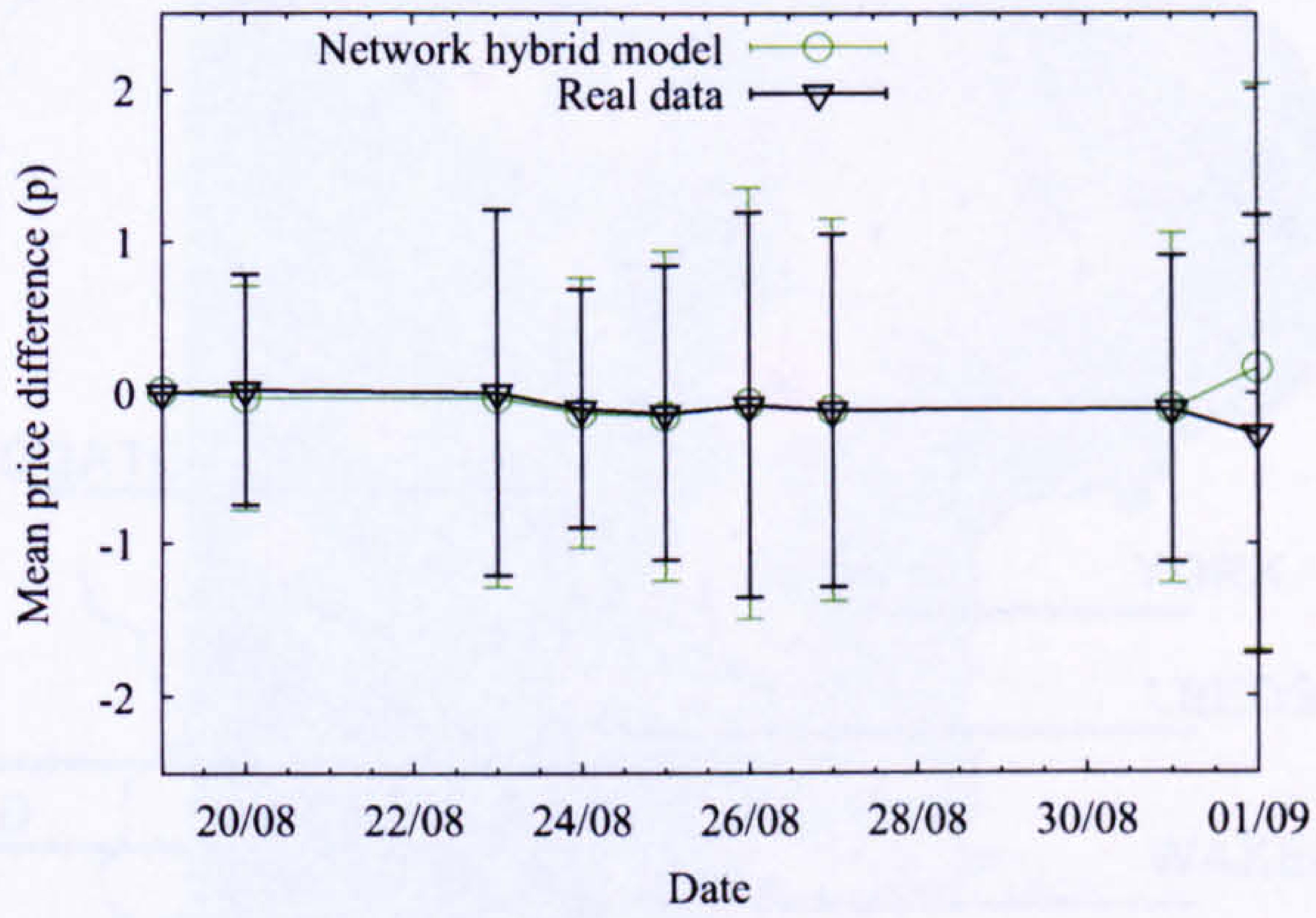
(b) Profit

Figure 8.18: Comparison of the (a) price and (b) profit over time at stations in the price war area around Halifax.

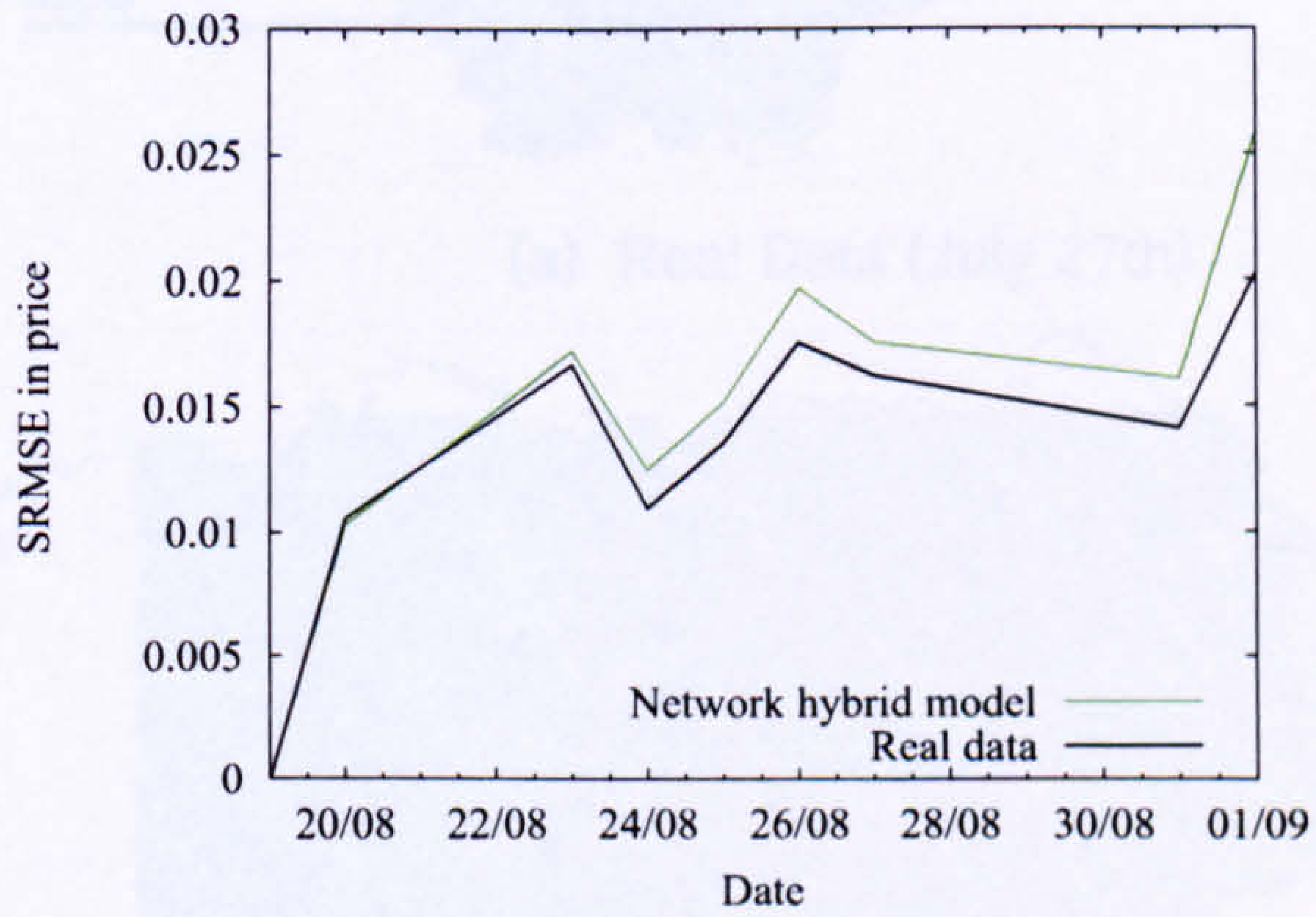
8.8.2 Yorkshire Region

Testing the ability of the network hybrid model to generate rather than preserve variations within the data was first used in §8.3.2 on West Yorkshire. In this section, the Yorkshire region will be used. This will test the ability of the network hybrid model to recreate rural-urban patterns on a much larger geographical scale. The Yorkshire region contains a mixed variety of geographical characteristics (see Figure 8.20(a)). West Yorkshire is an urban conurbation surrounded by rural areas whilst North Yorkshire is a much larger, predominately rural county. South Yorkshire offers a mixture of rural and urban characteristics. This will provide an interesting test for the network hybrid model.

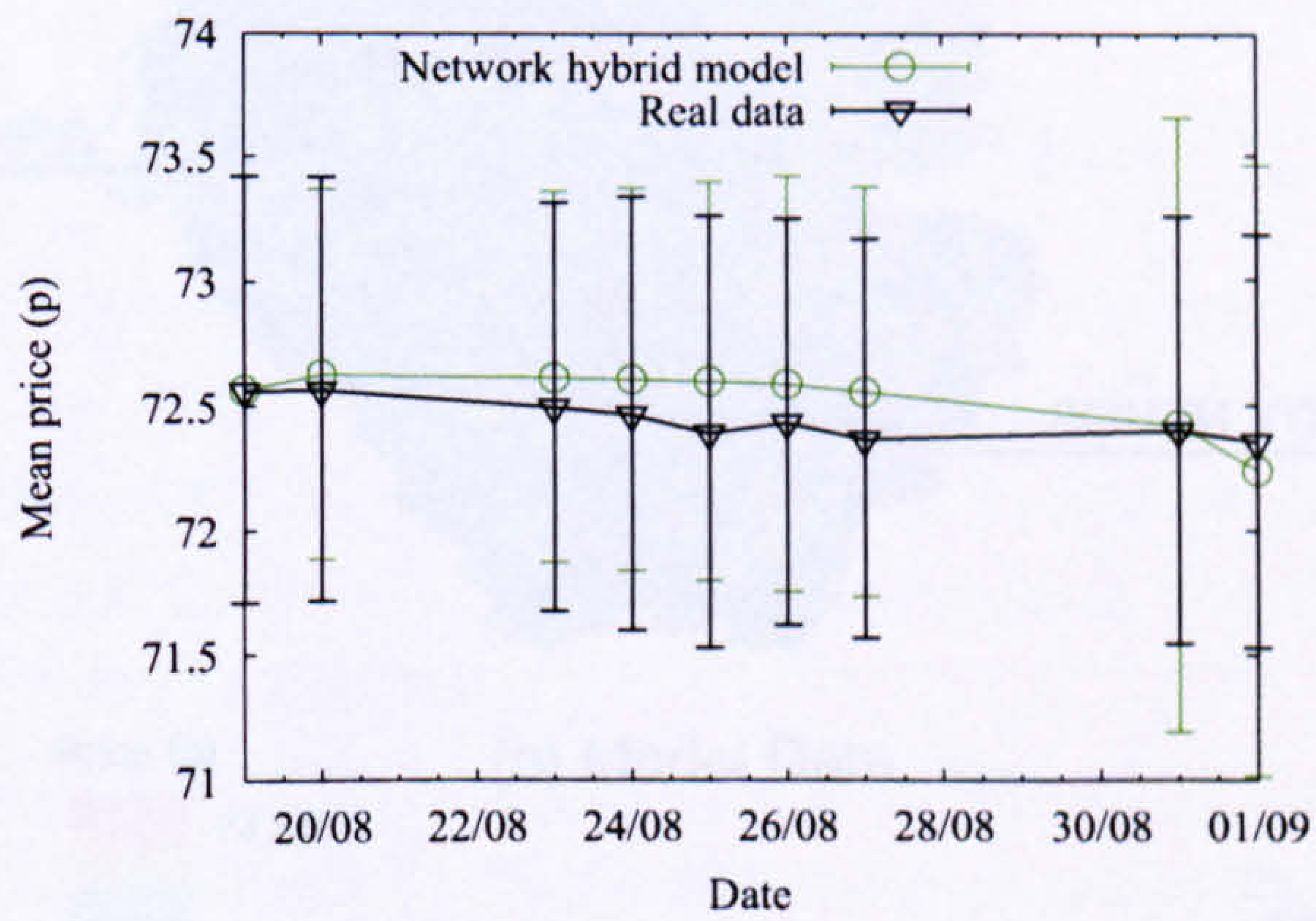
The Yorkshire region petrol stations were initialised at a constant price of 71p. The default parameter values summarised in Table 8.2 were used and the network hybrid model was run to equilibrium. One of the first observations that can be made from Figure 8.20(b) is that the network hybrid model has captured the main rural-urban trends within the region. The rural area of North



(a) Mean Price Difference and SD



(b) SRMSE



(c) Mean Price

Figure 8.19: Comparison of the (a) mean price difference (b) SRMSE and (c) mean price over time between the real and network hybrid model data. The model was initialised with data from August 19th for West Yorkshire. SDs are represented by vertical bars.

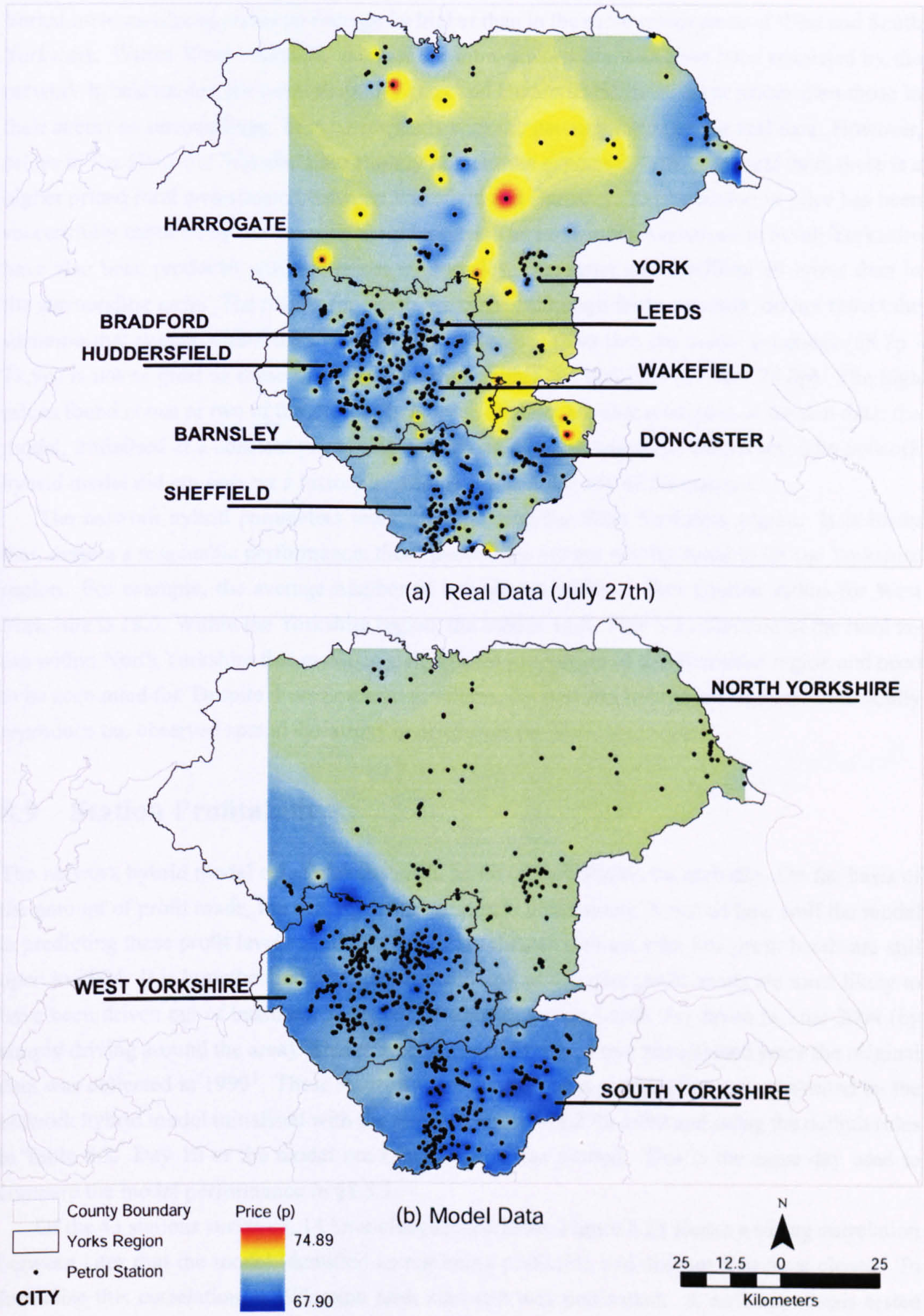


Figure 8.20: Comparison of the (b) network hybrid model's performance on the Yorkshire Region with the real data (a) for July 27th.

Yorkshire is sustaining prices on average 2p higher than in the more urban areas of West and South Yorkshire. Within West Yorkshire, most of the intra-urban variations have been recreated by the network hybrid model, for example both Leeds and Huddersfield have lower prices than those in their suburban surroundings. This corresponds with the patterns found in the real data. However, prices in Bradford and Wakefield are slightly higher than expected. Within the real data, there is a higher priced rural area situated between Wakefield and Barnsley. This variation in price has been successfully captured by the network hybrid model. The intra-urban variations in South Yorkshire have also been produced with the prices in Barnsley, Doncaster and Sheffield all lower than in the surrounding areas. The results for North Yorkshire, although fairly accurate, do not reflect the variation that occurs within the real data. The range of prices that the model produces (68.7p - 71.9p) is not as great as those naturally occurring within the real data (67.9p - 74.8p). The high prices found at one or two of the stations in North Yorkshire are characteristics of the real data; the model, initialised at a constant price, was not expected to recreate these anomalies. The network hybrid model did not possess a history of pricing patterns for any of the stations.

The network hybrid parameters were derived using the West Yorkshire region. It is likely that, despite a reasonable performance, these parameters are not wholly suitable for the Yorkshire region. For example, the average number of neighbours within a 5km circular radius for West Yorkshire is 18.0. Within the Yorkshire region, the total is 14.8. This is a reflection of the rural areas within North Yorkshire that make up a significant proportion of the Yorkshire region and need to be accounted for. Despite these potential problems, the network hybrid model does successfully reproduce the observed spatial variations in price over the Yorkshire region.

8.9 Station Profitability

The network hybrid model calculates the profit made at each station on each day. On the basis of the amount of profit made, the strategy of the station is determined. A test of how well the model is predicting these profit levels would be to assess whether stations with low profit levels are still open in 2004. It is hypothesised that stations with low or negative profit levels are most likely to have been driven out of business. A survey of stations within Leeds was taken in June 2004 (by simply driving around the area). This provided a list of stations that have closed since the original data was collected in 1999[1]. These stations were overlaid on a profitability map generated by the network hybrid model initialised with the real data from July 27th 1999 and using the default rules in Table 8.2. Day 10 of the model run (August 6th) was plotted. This is the same day used to compare the model performance in §8.3.1.

Of the 43 stations surveyed, 14 have closed since 1999. Figure 8.21 shows a strong correlation between sites that the model identified as not being profitable and stations that have closed. To formalise this correlation, a Wilcoxon rank sum test was performed. A null hypothesis tested was "there is no difference between the mean profits of stations that are open and closed". The Wilcoxon test gave a result of $p=0.000812$. This means that there is a smaller than 0.1% chance of the null hypothesis being true. The null hypothesis can therefore be rejected. As clearly evident in Figure 8.21 there is a definite difference in the profit between stations that are open and closed.

[1] 43 stations were surveyed out of a possible 51.

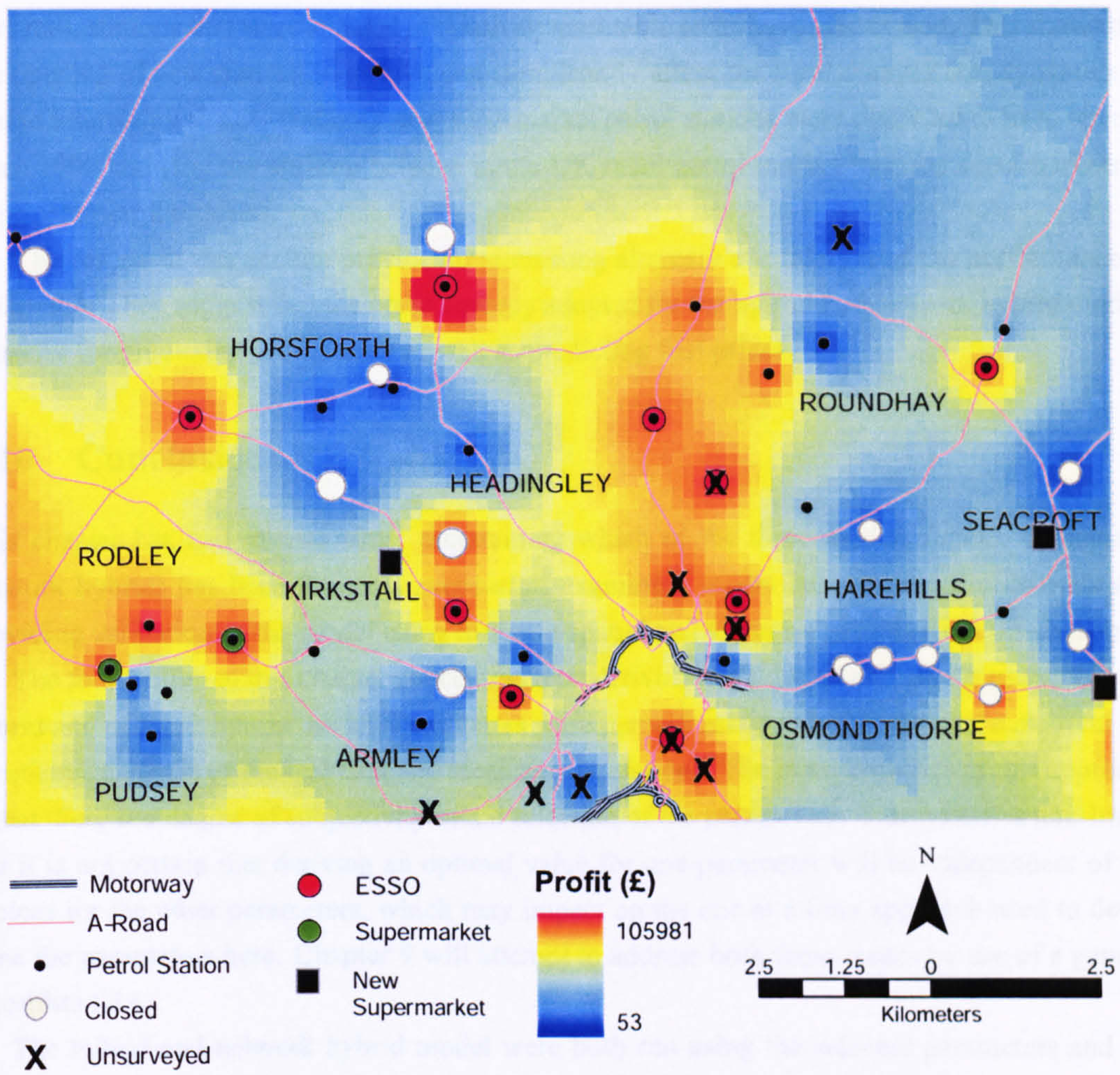


Figure 8.21: Map showing the profitability of stations in Leeds using the network hybrid model with data from 1999. Also shown for comparison are the status (open or closed) of the stations in 2004.

In §8.6.1, the assertion was made that Esso petrol stations are geographically well located in terms of maximising their profit. The results from Figure show that this is overwhelmingly the case. Each of the Esso stations is located in an area of high profitability. Whether this is due to initial site research by the company or as a result of the success of the Price Watch policy is difficult to conclude.

A more confident assertion can be made about supermarket stations. Figure 8.21 shows that these stations are also located at the high points of the profitability map. The original data pre-dates the appearance of the new supermarket stations at Kirkstall, Seacroft and near Osmondthorpe. Since the appearance of these stations a large number of their closest competitors have closed. This can be particularly seen when examining the sites surrounding the new supermarket stations at Seacroft and near Osmondthorpe. These now closed stations were identified by the model in 1999 as sustaining low profits. With the introduction of the supermarket stations with their aggressive pricing, the marginal profits of these stations would have plummeted thus resulting in their closure. This effect is not solely reserved for the low profit stations. At Osmondthorpe and

Kirkstall, stations that were situated in relatively profitable areas have also closed. This shows that the opening of new stations in an area can significantly affect the local competition dynamics. It is also interesting to note that only new supermarket petrol stations were observed to have opened since 1999 showing the shifting balance in the UK retail petrol market towards supermarkets as important petrol retailers.

The results in this section provide an interesting alternative validation of the performance of the model. They suggest that the profit strategy adopted for the hybrid and network hybrid models is useful for predicting profitability as well as predicting fuel prices.

8.10 Conclusions

This chapter has had several aims; ascertaining which of the three models (agent, hybrid and network hybrid) has been the most successful; examining model behaviour and sensitivity and providing validation of the model using different spatial and temporal periods from the data set.

The first section of this chapter concentrated on deriving suitable parameters for use within the hybrid and network hybrid model. Parameters were derived using a combination of numerical experiments, analysis of the real data and idealised simulations. The main criticism of this approach is that there is a degree of subjectivity in the selection of the parameters. The system is non-linear and it is not certain that deriving an optimal value for one parameter will be independent of the choices for the other parameters, which may impact on the one at a time approach used to determine the parameters here. Chapter 9 will attempt to address both these issues by use of a genetic algorithm (GA).

The hybrid and network hybrid model were both run using the selected parameters and the results compared with each other and with those from the agent model. Through visual and statistical comparison with the real data, the network hybrid model was found to be the most successful. The inclusion of the consumer and network models has resulted in a marked improvement on the original agent model performance. Attempts were made to further improve the performance of the model by assigning individual rules to the supermarket and Esso brands. However, this was not conclusive, suggesting that perhaps the introduction of new brand rules into the system would require the default parameters to be adjusted. This will be re-examined in Chapter 9 where the parameters will be objectively selected using a GA.

The behaviour, sensitivity and robustness of the model were tested via a series of diffusion experiments. Similar patterns of behaviour were recorded for each of the experiments. This led to the conclusion that the model was robust to changes in the initial conditions and was able to reproduce sensible results under a variety of environments. Further examination of the model behaviour was undertaken by increasing and decreasing the *costToProduce* parameter. This produced results that were analogous to the "rockets and feathers" phenomena first documented in Chapter 2.

Finally, attempts were made at validating the performance of the model against other real pricing data. This was accomplished by running the model on different areas and for different time periods to those used in calibrating the model. A comparison was also made of the profitability of stations generated by the model for 1999 with a list of stations that were known to have closed between 1999 and 2004. This work showed that a high number of stations determined to be poor earners in 1999 had closed over the period to 2004. Several factors are hypothesised to contribute to the poor performance of these stations including poor location, the increase of supermarket stations and the Esso Price Watch policy.

In summary, the work within this Chapter has shown that the network hybrid model produces sensible results that reproduces the behaviour seen in the real petrol market. However, one of the weaknesses of this model is the subjectivity in the choice of the parameters. A GA will be introduced in Chapter 9 that will objectively evolve a set of optimal parameters for a given data set.

Chapter 9

Optimisation using a Genetic Algorithm

9.1 Introduction

In Chapter 8, the network hybrid model was run with a set of parameters derived from the analysis of the real data in Chapter 4 and the idealised numerical simulations performed in Chapter 7. The performance of the model was tested with both idealised and real data. The results showed the model to be both robust and accurate at replicating processes within the real system. However, the method of selecting individual parameters is not entirely objective. The system is also non-linear and experimenting with one parameter at a time may not result in an optimal set of parameters. Initial investigations showed the system to be complex with multiple local minima suggesting that there are many different solutions available for the problem.

Graphically, this can be represented by a series of mountains and valleys (see Figure 9.1). Finding the lowest point in a particular valley is relatively easy. A simple progression down the slope will generally give the local minimum. However, locating the global minimum is more difficult since there is no certain way of knowing which valley it is located within, or even how many valleys there are, without fully searching the whole of the parameter space. There are many techniques available that can be used for finding this global minima ranging from more conventional search techniques, for example hill-climbing and annealing to artificial intelligence techniques such as fuzzy logic and evolutionary algorithms.

There are several issues to be considered when reviewing appropriate optimisation techniques. In this case, the research problem is complex, the system is non-linear and, as evident in Chapters 7 and 4, changes in the SRMSE are not smooth since abrupt changes in price may occur. These features mean the function is not differentiable or even smooth and so it is not possible to calculate the derivatives (i.e. how the error changes as the parameters are changed). This invalidates several methods that rely on derivatives, for example conjugate gradient and variable metric methods (see Press *et al.*, 1992, for a discussion).

Techniques such as hill-climbing and annealing are often incapable of optimising non-linear, multi-modal functions (see Pham and Karaboga, 2000, for a detailed discussion). In such cases, a random method may be required. However, undirected search techniques are extremely inefficient for large domains.

Evolutionary algorithms (EAs) are search methods that take their inspiration from natural

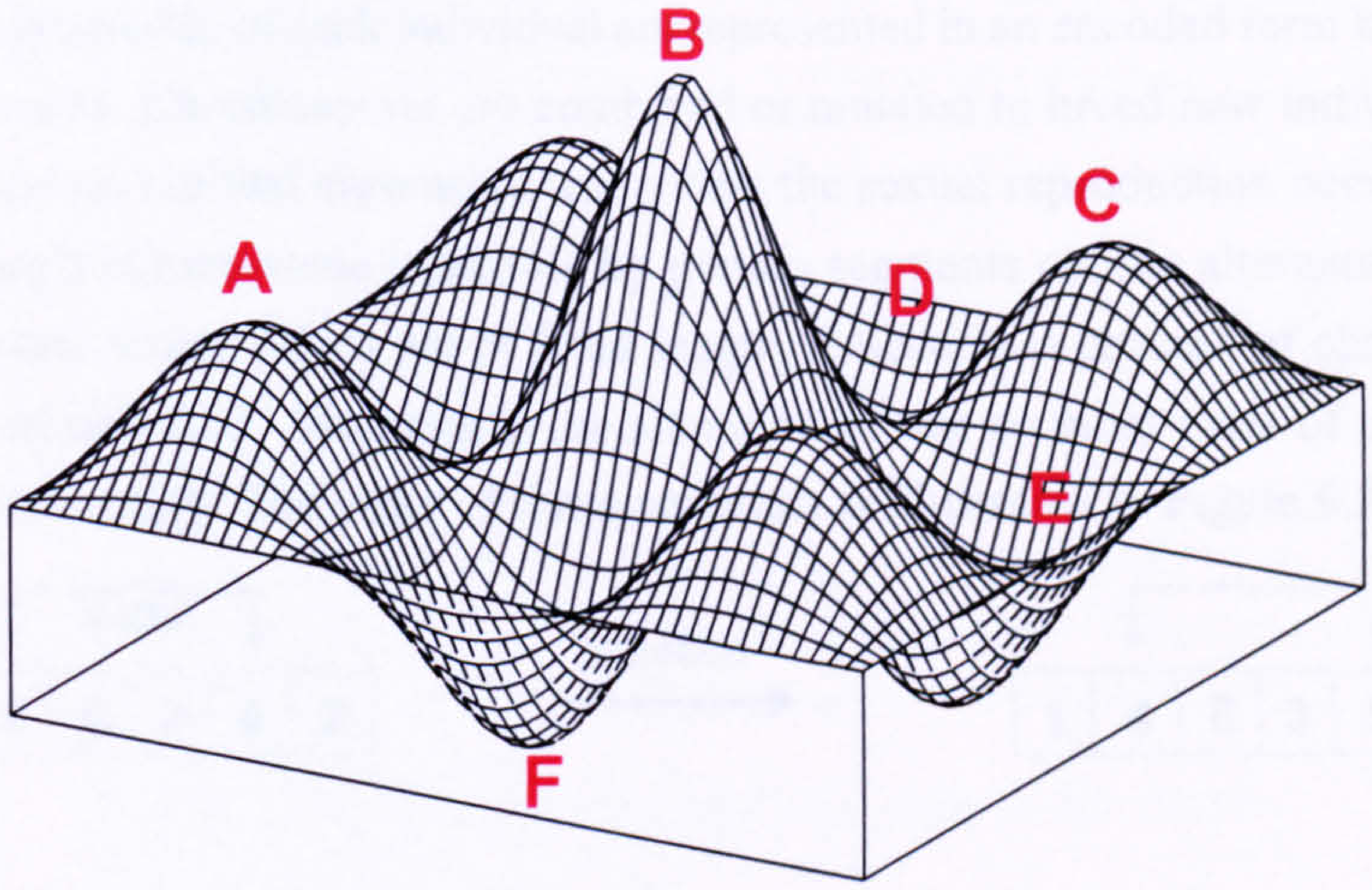


Figure 9.1: Example of a function containing multiple maxima and minima to demonstrate the difference between local and global extrema. A and C represent local maxima; D and E are local minima; B is the global maximum and F is the global minimum.

selection and survival of the fittest in the biological world. These algorithms can find the global optimal solution in complex multi-dimensional space. Different main schools of evolutionary algorithms have evolved during the last 30 years: genetic algorithms, mainly developed in the USA by Holland (1975), evolutionary strategies, developed in Germany by Rechenberg (1973) and Schwefel (1977) and evolutionary programming by Fogel *et al.* (1966). Each of these constitutes a different approach, however, they are inspired by the same principles of natural evolution.

The specific type of EA that will be used in this thesis are genetic algorithms (GAs). GAs are useful for multidimensional optimisation problems in which the chromosome (set of rules) can encode the values for the different variables being optimised. This is of particular use for the research within this thesis. This chapter will be divided as follows. A brief literature review will provide an explanation of GAs and highlight the advantages of using this approach over traditional techniques. Following this, the construction of the GA will be detailed along with discussion of which genetic operator methods are to be used. This is followed by experimentation with optimal space parameters for the GA. The final sections of the chapter involve comparing the performance of the model using the best GA parameter values and the parameters devised in Chapter 8. This will focus on how well the GA derived rules replicate the trends in the real data and how well they perform on other parts of the real data, for example the Yorkshire region. Finally the Esso and supermarket case-studies will be revisited to answer the question of whether the GA can produce optimal rule sets for both these categories and thus improve the overall model performance.

9.2 Genetic Algorithms

Genetic Algorithms (GAs) are modelled on natural evolution, specifically, the operators it employs are inspired by natural evolutionary process. These processes, known as genetic operators, manipulate individuals in a population over several generations to improve their fitness. A detailed introductory survey can be found in Reeves and Rowe (2003).

In a GA, the properties of each individual are represented in an encoded form known as a *chromosome* (or *genome*). Chromosomes are combined or mutated to breed new individuals. *Recombination* (or *crossover*) of two chromosomes models the sexual reproduction occurring in nature. Here, an offspring's chromosome is created by joining segments chosen alternately from each of two parents' chromosomes which are of fixed length. Selection is process of choosing the chromosomes to be recombined. Mutation is the alteration of one or more parts of the chromosome with a random probability. The order of these operators is illustrated in Figure 9.2.

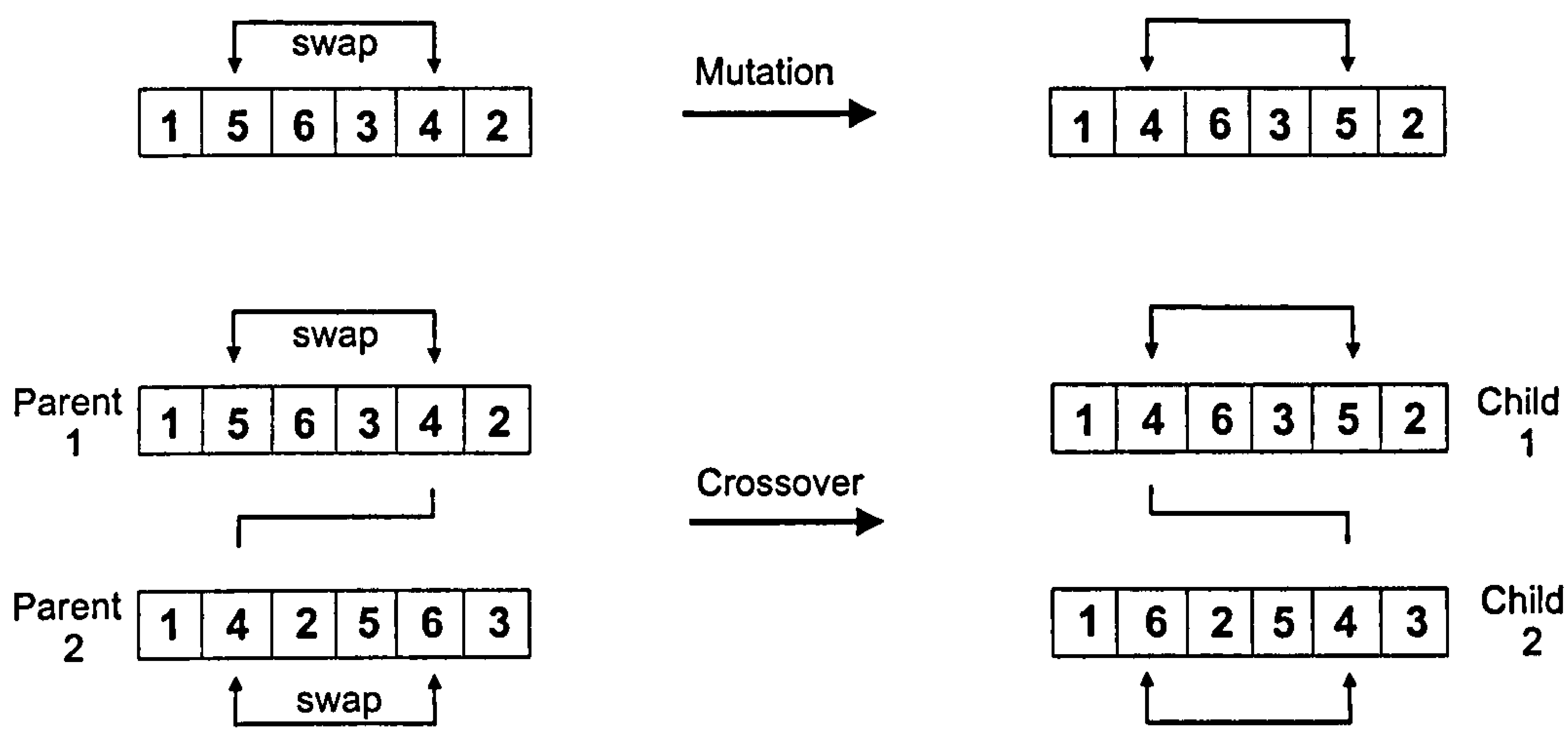


Figure 9.2: The crossover and mutation operators applied to candidate solutions of a combinatorial optimisation problem (after Flake, 2001).

Figure 9.3 presents a simple schematic of how a GA operates. The algorithm is very simple, with the main functions contained in the innermost loop. These include the process of selection, recombination and mutation.

```
Initialise the population P
Repeat for some length of time
        Create an empty population, P1
        Repeat until P1 is full;
    SELECT two individuals from P based on some fitness criteria
    Optionally mate and RECOMBINE with offspring
    Optionally MUTATE the individuals
    Add the two individuals to P1
    Let P now be equal to P1
```

Figure 9.3: Basic structure of an EA (after Flake, 2001).

The terminology commonly associated with GAs are defined in Table 9.1. These terms will be used throughout this chapter.

Term	Explanation
Gene	The parameter that is being optimised.
Chromosome/Genome	The combination of several genes representing an individual.
Fitness	A measure of the success or failure of the gene.
Selection	The process of choosing the chromosomes to be recombined.
Recombination/Crossover	The exchange of "biological information" to produce offspring by the juxtaposition of the parents.
Mutation	The alteration of one or more of the genes in a chromosome with a random probability.

Table 9.1: Explanation of terminology used in this chapter.

9.3 Control Parameters

The successful implementation of a GA is down to several factors. These include the initial size of population and the methods used for selection, recombination and mutation. These conditions vary due to the problem. A poor selection of control parameters can result in a corresponding poor performance in the GA. Several researchers have examined the effect of these parameters on the performance of a GA (Schaffer *et al.*, 1989; Grefenstette, 1986; Fogarty, 1989). The main conclusions were:

- A large population gives the simultaneous handling of many solutions and increases the computation time per iteration. However, as many samples from the search space are used, the probability of convergence to a global optimal solution is higher than when using a small population size.
- Low crossover/recombination rates decrease the speed of convergence. High rates result in concentration around one solution.
- High mutation rates introduce high diversity in the population and may cause instability. However, if the rate is too low, it can be very difficult for the GA to find a global optimal solution.

There are a number of different selection, mutation and recombination conditions that are available. These will be reviewed within the following sections.

9.3.1 Initial Population

At the start of an optimisation, a GA requires a group of initial solutions. There are two ways of forming this initial population. First, randomly produced solutions can be generated. This is the preferred method for problems about which no a priori knowledge exists or for assessing the performance of an algorithm. The second method uses a priori knowledge about the problem. Using this knowledge, a set of requirements are collected to form an initial population. In these cases, the GA starts the optimisation with a set of approximately known solutions and therefore converges to an optimal solution in less time than with the first method.

9.3.2 Representation

Most of the problems suitable for GAs involve identification of a set of parameters (whether for optimisation, combinatorial or other problems), which need to be represented to allow evolutionary operators to be effectively applied. As GAs are robust, there is little need to rigorously identify the "best" representation for a particular problem (Goldberg, 1989). There are two broad methods that can be used for representation; binary alphabets (Holland, 1975) and real numbers (Davis, 1991; Beasley *et al.*, 1993; Janikow and Michalewicz, 1991; Michaelewicz, 1992).

There is no single "correct" coding method for encoding a problem, the mode of representation is dependent on the problem. However, the coding sequence must adequately represent the problem to ensure that the optimal solution is available to the algorithm. For example, if the optimal solution contains a value of 2.5, there is little use in developing a representation of a real number in the range ±2.0.

9.3.3 Fitness and Selection

In order to evolve better performing solutions, the fittest members of the population are selected and randomly exposed to mutation and recombination. This produces offspring for the next generation. The least fit solutions die out through natural selection as they are replaced by new recombined, fitter, individuals. Evaluation of the fitness of chromosomes involves some form of comparison between observed data and the results for a particular solution, or test to see if a particular solution meets certain criteria or constraints.

There are number of possible ways for selection to take place. The following parental selection schemes that recur within the literature[1]:

- **Ranking Selection:** The population is sorted from best to worst. The number of copies that an individual receives is given by an assignment function and is proportional to the rank assignment of an individual.

- **Tournament Selection:** A random number of individuals are selected from the population. The best individual from this group is chosen as a parent for the next generation. This process is repeated until the mating pool is filled.

- **Roulette Wheel Selection:** Individuals are mapped to contiguous segments of a line, such that each individual's segment is equal in size to its fitness. A random number is generated and the individual whose segment spans the random number is selected. This process is repeated until the desired number of individuals is obtained.

- **Truncation Selection:** Truncation sorts individuals according to their fitness (from best to worst). Only the best individuals are selected to be parents.

Along with the selection method, the selective pressure parameter is critical (Hancock, 1994). This parameter measures the probability of the best individual being selected compared to the average probability of selection. Essentially, this parameter drives the algorithm towards a solution;

[1]There are a variety of other selection methods that are not listed here. These include stochastic remainder and stochastic universal selection. A comparative analysis of parental selection methods can be found in Goldberg and Deb (1991).

too large a value and the search will converge rapidly onto a solution (not necessarily optimal); too small and progress will be very slow.

9.3.4 Recombination

The main reproductive genetic operator is recombination (also known as crossover). This is the process by which new individuals are produced by combining the information from two parents (chromosomes). The resulting offspring inherits components from both parents. This allows the GA to explore new areas in the search space. Without recombination, the offspring are simply duplicates of the parents. This does not give any opportunity for improving the fitness of the population.

There are several methods of recombination available, the suitability of the method is dependent on the type of gene (variable) stored on the chromosome. Only methods that can be applied to real value data will be detailed here[2]. These are the intermediate, line and extended line recombination methods.

In intermediate recombination, the variable values of the offspring are randomly chosen from between the values of the the parents (see Figure 9.4(a)). Values of normally up to 25% outside this range can be used. The value of 25% is chosen to ensure that statistically a space covered by the recombinations does not decrease in size with time leading to a loss in diversity. The position of the variable chosen on the line determines how much each parent contributes to the offspring and is chosen uniformly at random for each gene. Line recombination is similar to intermediate recombination except that the same random number is used for selecting the value of every gene in a chromosome (see Figure 9.4(b)). Extended line recombination is different from the above techniques in that the variable range is not limited to a range around the parents. The probability of any particular value being taken is not uniform but varies with a high probability near the parents and a low probability far away from the parents. The probability distribution can also be chosen to favour the fitter parent. The value controlling the amount of the parent used is generated randomly and then used for selecting the value of subsequent genes (see Figure 9.4(c)).

9.3.5 Mutation

Mutation occurs after recombination has taken place. By mutating individual genes, the GA can exploit existing areas to find a near optimal solution. There are two parts to selecting the mutation method; the probability of mutating and the step-size.

Several papers were reviewed for the optimal mutation rate used (De Jong, 1975; Schaffer *et al.*, 1989; Grefenstette, 1986). The most commonly used mutation rate under a wide variety of test functions was $1/n$ (n: number of variables of an individual). This means, that for every mutation only one gene per individual is on average mutated. The more genes one individual has, the smaller the mutation probability for an individual gene. The mutation rate is therefore independent of the size of the chromosome.

---

[2]The main omission is binary recombination. The reader is directed to Reeves and Rowe (2003) for further details of this technique.

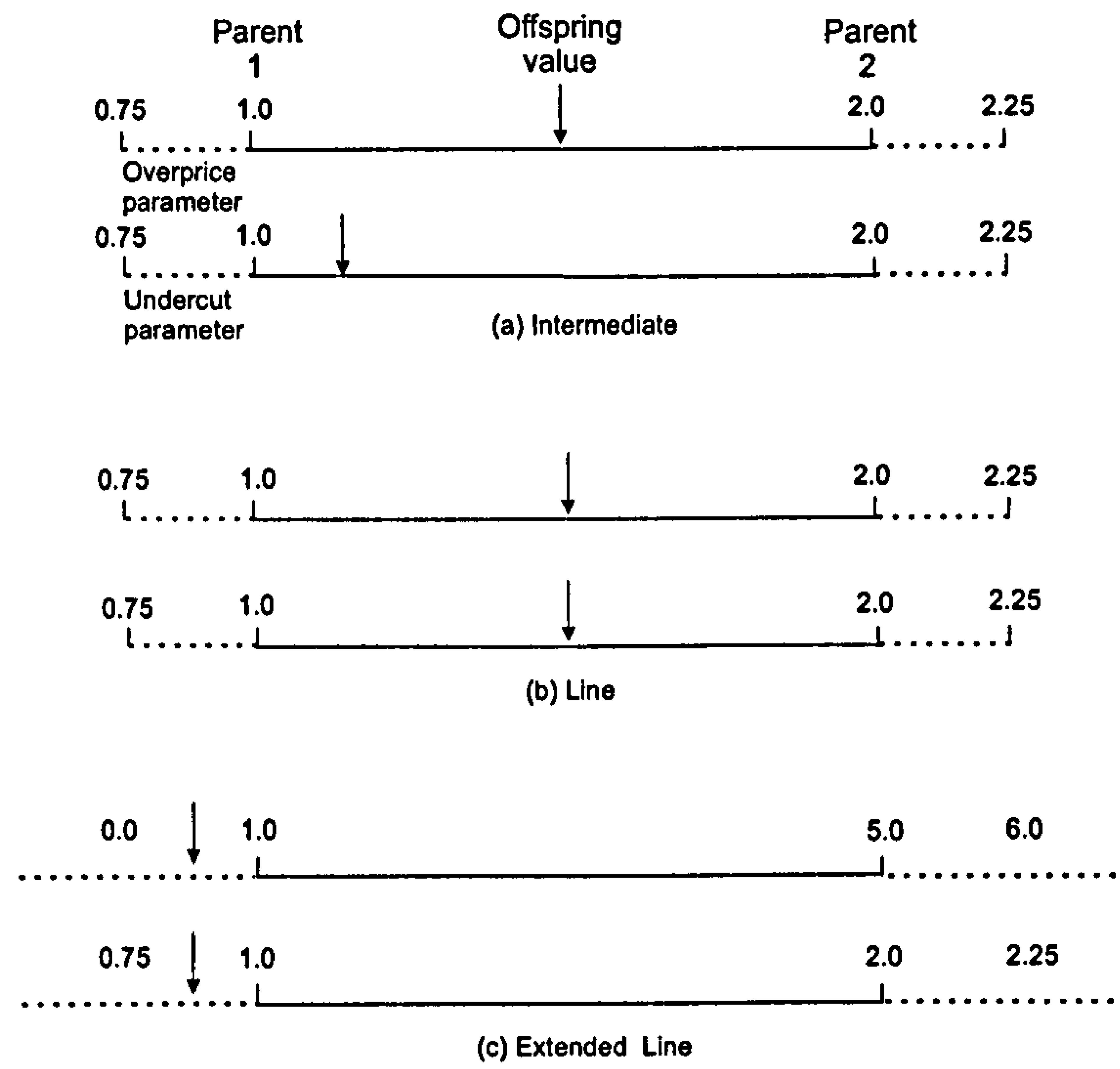


Figure 9.4: Illustration of (a) intermediate, (b) line and (c) extended line recombination.

However, these recommendations for the mutation rate are only correct for separable functions. Most real world functions are not fully separable. For these functions no recommendations for the mutation rate were given, instead a mutation rate of $1/n$ was suggested in the absence of further information.

The literature offers no strict guidelines for the selection of the size of the mutation step. The optimal step-size depends on the research problem and may even vary during the optimisation process. Small mutation steps are acknowledged in the literature as being successful, especially when the individual is already well adapted. However, large mutation steps can when successful, produce good results very quickly. A good mutation operator should therefore produce small step-sizes with a high probability and large step-sizes with a low probability.

9.4 Using GAs for Optimisation

There are several advantages to using GAs over traditional techniques for optimisation (Pham and Karaboga, 2000). One of the most significant is parallelism. GAs are capable of considering many points at once during the search process. This reduces the chance of converging to local optima. During the search process, GAs use probabilistic rules, not deterministic. This allows them to outperform conventional optimisation techniques on difficult, discontinuous and multi-modal functions (Reeves and Rowe, 2003). Additionally, GAs do not require derivative information or other auxiliary knowledge; only the objective function and corresponding fitness levels influence the directions of search.

One of the strengths of using GAs is the ability to hybridise with other techniques. For example, GAs have been extensively used to optimise both the connection weights and the actual connectivity of a neural network (Whitley, 1989; Harp and Samad, 1991) whilst See (1999) demonstrated the use of GAs to optimise fuzzy logic models. An increasing amount of research has been directed at using GAs to optimise multi-agent systems (MAS). Examples can be found in a range of different applications. Choi *et al.* (2004) used a hybrid GA-MAS model to improve scheduling systems for supply chains. Dosi *et al.* (1999) showed how a GA can be successfully used to model learning in industrial environments, whilst Wu *et al.* (2003) used GAs to optimise agents' initial positions in land combat simulations. Finally, Laureano-Cruces *et al.* (2004) took a different perspective using agents to improve the GA, specifically the updating of sub-populations, thereby avoiding the problem of local minima.

No examples were found in the literature of GAs being used to optimise the rule set of a MAS for determining petrol prices either as single model or as part of a hybridised application.

Despite their unique and adaptive search capabilities, there are no guarantees that GAs will find the global solution; however, they can often find an acceptable one quickly (Goldberg and Deb, 1991). GA can provide a number of potential solutions to a given problem with the final choice being left to the user.

9.5 Summary

The purpose of this brief literature review is to present an overview of GAs to provide a basic understanding of their concepts and mechanisms. The importance of selecting suitable control parameters is of vital importance for successful implementation of a GA. The following sections will detail the selection of these parameters as well as outlining the construction of the GA.

9.6 Construction of the GA

In total, there are 8 genes (parameters) that will be optimised. These are β, λ, *costToProduce*, *neighbourhood*, *fixedCosts*, *overprice*, *changeInProfit* and *undercut* (see Chapter 6 and Chapter 7 for definitions). All of these genes exert some influence over the performance of the model and have been previously experimented with in Chapters 5, 7 and 8. The values associated with each parameter are continuous, and a range of allowed values will be assigned to each gene based on values obtained through prior experimentation. As highlighted in §9.3.2, the ranges have to be sensible if the GA is to operate successfully (populations will be initialised and reproduce using the values within these ranges). The range of values that will be used are summarised in Table 9.2.

9.6.1 Operation of the GA

Two new classes were built and linked to the network hybrid model to run the GA. The Genetic class contains the code to run the GA (i.e. it controls the order in which the operators are applied) and the Chromosome class contains the details of each chromosome, for example the values of the genes. Several new parameters were created, these are presented in Table 9.3. A flow chart illustrating the operation of the GA is given in Figure 9.5.

Gene	Range
β	0.000003 - 0.003
λ	0.001 - 1.5
fixedCosts	100p - 10000.0p
costToProduce	60.0p - 70.0p
changeInProfit	2000p - 5000p
undercut	0.1p - 5.0p
overprice	0.1p - 5.0p
neighbourhood	1000m - 10000m

Table 9.2: Range of allowed values assigned to each gene in the GA.

Parameter	Explanation
numChromosomes	Number of chromosomes (population size) in each generation.
numKeep	Number of chromosomes kept at the end of each generation (the fittest chromosomes).
numGenes	Number of parameters that can be changed in each chromosome.
gmax, gmin	Continuous range of values that are associated with each gene. Initial values are randomly assigned in this range.
minfit	If the fitness is less than x amount, the simulation will stop.
maxiter	Maximum number of generations.
gmut	The initial maximum size of the mutation.
converge	*gmut* is multiplied by this value to determine the maximum size of the mutation at the current generation. *converge* becomes smaller with each generation.
cfact	Amount by which *converge* is multiplied each generation. Must be < 1 to ensure that *converge* gets smaller with time.

Table 9.3: Explanation of new parameters created for the GA.

The model will terminate if one of two conditions are satisfied; if the number of generations set are reached or if the SRMSE value matches the *minfit* value. The *converge* and *gmut* variables will be further explained in §9.7.4.

The full source code and a compiled version of the network hybrid model with the GA extension can be found on the CD accompanying this thesis.

9.7 Selecting Control Parameters

Several studies have concentrated on developing optimal parameter settings for GAs (see De Jong, 1975; Schaffer *et al.*, 1989; Grefenstette, 1986). Each study presents a different set of optimal parameters showing that these optimal values vary for the problem under consideration. For example, suggested values for population size range from 50-100 (De Jong, 1975), 20-30 (Schaffer *et al.*, 1989) and 30 (Grefenstette, 1986).

§9.3 showed that there are several alternative methods that can be used for each of the genetic operators. The following sections will briefly outline each of the options available before selecting the method that is the most appropriate for the application within this thesis.

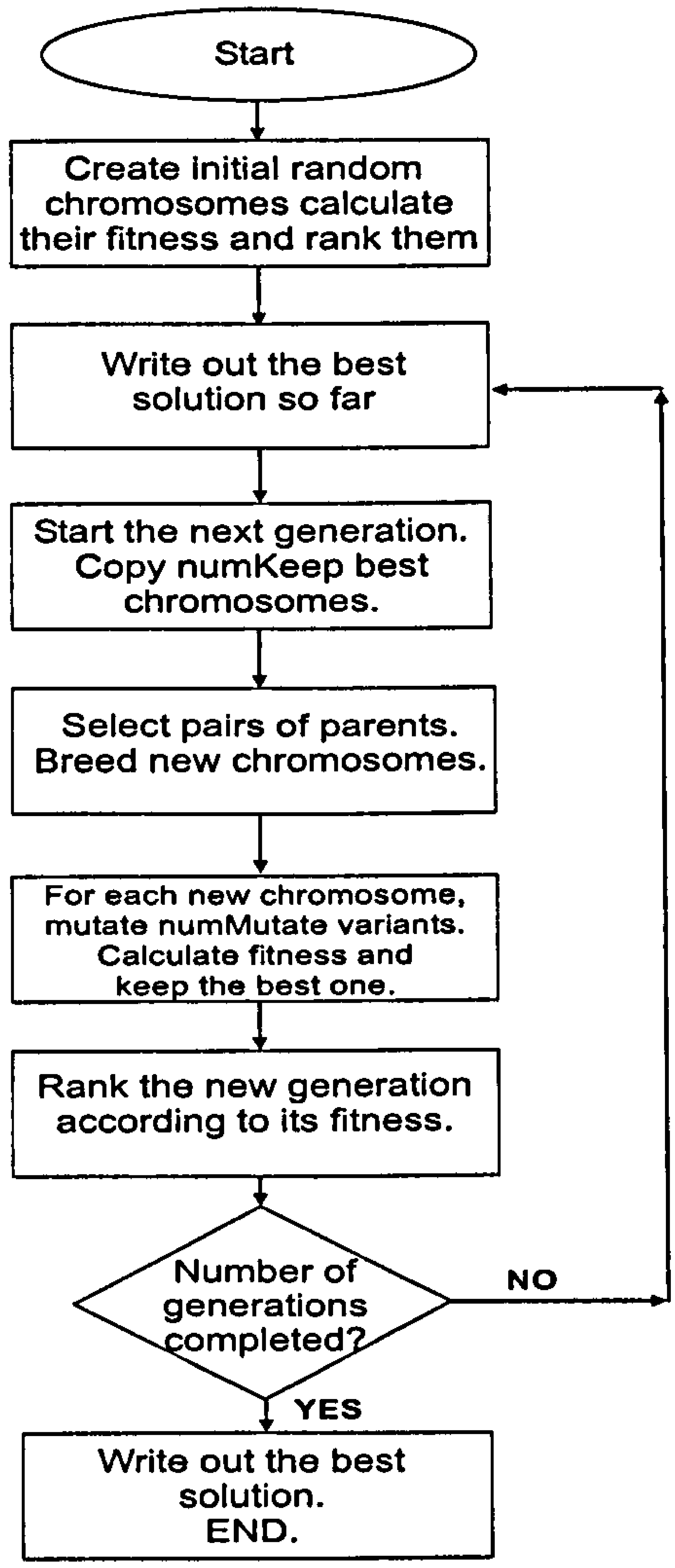


Figure 9.5: Flow chart summarising the operation of the GA.

9.7.1 Population Size and Initialisation

There are no set guidelines as to what the optimal population size should be for a successful GA. However, the literature does suggest that a larger population is advantageous as this creates a bigger search space and thus more potential solutions (Reeves and Rowe, 2003). This is particularly useful as the population will be randomly initialised. Random initialisation does not guarantee that the full genetic space will be represented, however, using a large population will increase the likelihood of this occurring.

A common problem with GAs is premature convergence to non-optimal solutions. One of the ways to avoid this problem is to explore several different areas of the search space simultaneously by evolving different sub-populations of solutions in parallel (Pham and Karaboga, 2000). Within the GA, subpopulations will be created after the parents have mated (recombined) to produce an offspring. Each of the other children produced is a mutation of this initial recombination. The fitness of each of these new chromosomes is worked out and only the best one is kept to be a parent in the next generation. This prevents any one particular group of mutations dominating the population at the next generation.

9.7.2 Selection Method

Holland (1975) commented that "the very essence of good GA design is retention of diversity". Selection, however, can have the effect of reducing diversity. Several methods of selecting parents were outlined in §9.3.3. Of these techniques, the ranking method will be used. Unlike approaches such as truncation, ranking gives any of the population the chance of becoming parents. The fitter the chromosomes, the more likely the chromosomes are to be parents (determined by a probabilistic distribution dependant on the ranking). Linear ranking is used. This means that the probability P of a chromosome being selected is linearly related to its ranking, k, (which is determined by its fitness) i.e.

$$P(k) = \frac{2-S+2(S-1)(k-1)/(N-1)}{N} \tag{9.1}$$

where N is the number of chromosomes, k is the position in the ranking of the chromosome ($k=1$ is the least fit, $k=N$ is the most fit) and S is the selection pressure. The selection pressure is the probability of the best chromosome being selected compared to the average probability of a chromosome being selected. For the linear ranking used here S must be in the range $[1,2]$. A value of 2 ensures the maximum chance of the fittest chromosomes being selected.

9.7.3 Recombination Method

Of the techniques reviewed in §9.3.4, intermediate recombination will be used. In this method, the variable values of the offspring are randomly chosen from between the values of the parents. This prevents "super genes" from dominating i.e. one gene rapidly dominating the search space, thus producing a sub-optimal solution. Therefore, in producing any generation, the best *numKeep* solutions from the previous generations are copied across and the rest of the population is generated from random recombinations of the previous generations. For each gene X the value of the offspring, X_{off} is

$$X_{off} = aX_1 + (1-a)X_2 \tag{9.2}$$

where X_1 and X_2 are the gene values of the parents and a is uniformly distributed in the range $[-d : 1+d]$. This means the offspring value for each gene is a linear combination of the two parent genes. A value of $d=0$ would ensure that the offspring value was somewhere between the two parents values however over time this results in a shrinkage of the range of values covered by the various chromosomes. A value of $d=0.25$ ensures (statistically) that the variable area of the offspring is the same as the variable area spanned by the variables of the parents and so this is the value used in this model.

9.7.4 Mutation Method

Mutations will be introduced by specifying an initial range of mutations for the gene. At each generation, the value of the gene will be multiplied by the convergence value. This will result in the mutations decreasing with each generation.

The *gmut* parameter determines the amount that the population is initially mutated by (i.e. the maximum size of the mutation). The actual mutation is assigned randomly within this range. The size of the mutation is checked to ensure the gene value does not exceed the specified range of values. If it does then the mutation is truncated to keep it in range. If convergence at these boundary values appears early on, this suggests that the range of values for a parameter needs to be larger. The size of *gmut* for each gene is chosen to span the range of possible values so that initially mutations can cover the whole search space. The *converge* parameter is multiplied by *gmut* to determine the size of the mutation at each generation. The parameter *gmut* is kept constant. The *converge* parameter is initially set to 1.0 and at each subsequent generation it is multiplied by the *cfact* parameter. The value of *cfact* has to be less than one to ensure that *converge* (the relative size of the mutations) gets smaller as the number of generations increases. The nearer *cfact* is to one, the more slowly the mutations will decrease in size. A greater value, for example 0.99 would be used for large populations or running for many generations to prevent the mutation size shrinking too rapidly. This is the value that will be used. This means that after 100 generations, the mutation size will have decreased to 37% of the initial mutation size.

The chance of each gene being mutated is $1/n$, this value is chosen from recommendations made within the literature in §9.3.5. At each generation, several different mutated versions of each chromosome are created and their fitness calculated. Only the best mutation from each chromosome is kept. The mutations allow for the possibility of each chromosome being improved. Since only the best mutation for each chromosome is kept, this algorithm prevents too rapid a loss of diversity within the population resulting from several similar genes becoming dominant.

9.7.5 Statistical Measure of Fitness

The standardised root mean square error (SRMSE) will be used to determine the fitness of each chromosome. This technique has been used throughout this thesis to assess model performance with various experiments. Use of the SRMSE will therefore allow a ready comparison with the results produced from previous chapters.

9.7.6 Summary

The methods and values (where appropriate) that will used in the initial experimentation with the GA are summarised in Table 9.4.

Parameter	Method
Population initialisation	Random.
Population size	Large. To be determined through experimentation.
Selection	Linear ranking.
Selection Pressure	2.0.
Recombination	Intermediate recombination.
Recombination Value	0.25.
Mutation	Chance of mutation $1/n$.
Fitness	SRMSE.

Table 9.4: Summary of methods and parameters chosen for the GA.

9.8 Optimal Space for Solution

One of the key factors in the success of GAs is finding the correct balance between the amount of exploration and exploitation needed (Flake, 2001). The following sections detail experimentation carried out with the different control parameters of population size (*numChromosomes*), number of mutations (*numMutate*) and number of generations (*maxiter*). The network hybrid model is not used within this section due to the additional computational time that would be required for the redistribution of the population. This omission will not have any significant effect on determining the best parameters for the GA. In §9.10 the GA will be run with the network hybrid model to determine optimal parameters for the network hybrid model.

9.8.1 Default Benchmark Run

In order to provide a comparison for future experiments, benchmark runs were undertaken with a set of default values outlined in Table 9.5. This provides a method of comparing the impact of altering the GA parameters.

Parameter	Value
Number of generations (*maxiter*)	100
Number of chromosomes (*numChromosomes*)	100
Number to keep between generations (*numKeep*)	1
Number of mutations to generate (*numMutate*)	3
Convergence factor (*cfact*)	0.99
Selection pressure (*sp*)	2

Table 9.5: Parameters used for the benchmark (default) GA runs.

Figure 9.6 shows the time series of the SRMSE for several different runs using these default parameters. There are several observations that can be made about Figure 9.6. All of the solutions rapidly converge after between 10 - 20 generations followed by a period of very slow improvement up to the 60th generation. After this point, there is very little further improvement indicating that the model has converged to a minimum. The exception to this is run 1f which experiences a sudden improvement between 50 - 55 generations. However, the figure also shows that there is a wide spread of values in the solutions. This indicates that there is more than one local minimum. Further experimentation will hopefully determine the presence of a global minimum.

9.8.2 Experimentation with Population

To investigate the effect of varying the population size, the GA was run using the default parameters (Table 9.5) except for a differing number of chromosomes. Population sizes of 100, 200, 300 and 500 were used. Each experiment was repeated several times. The purpose was to determine whether a larger population (and therefore a greater genetic search space) would improve the performance of the GA.

Figure 9.7 shows that there are no significant differences in the final solutions. A population of 100 chromosomes produces similar results to a population of 300 or 500. The variation be-

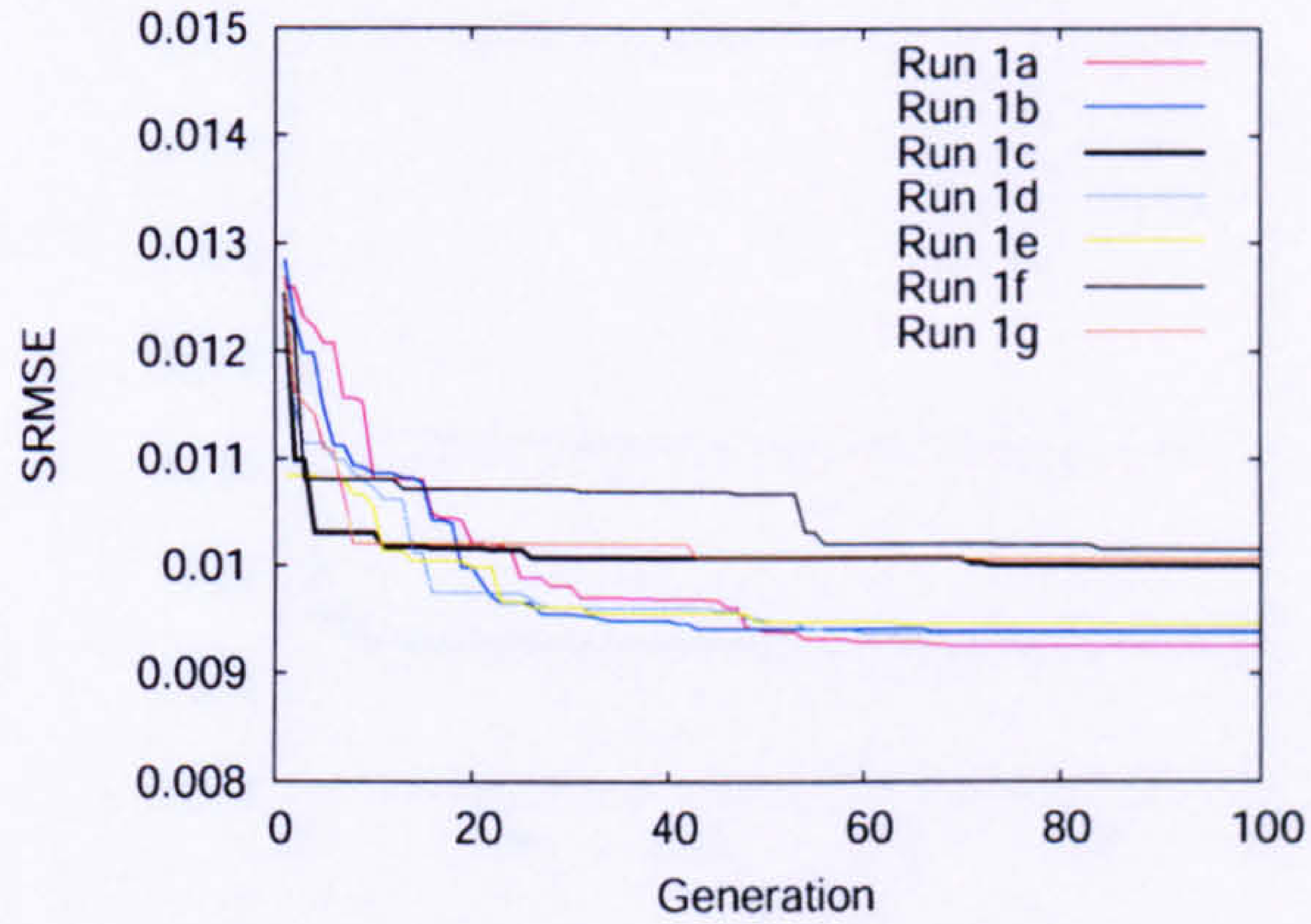


Figure 9.6: Graph of the SRMSE against time for several runs using the benchmark GA parameters.

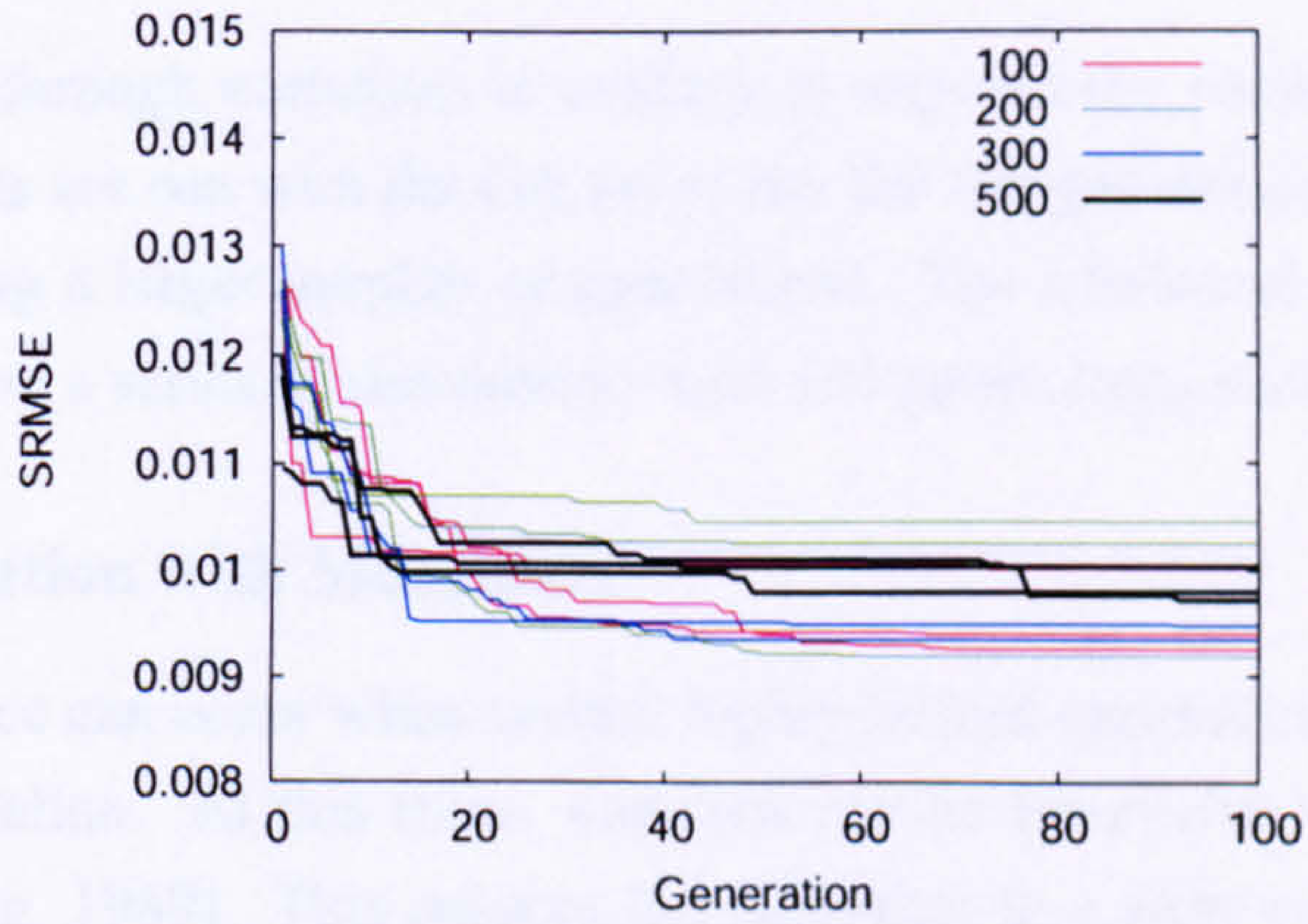


Figure 9.7: Graph of the SRMSE against time for several runs using the benchmark GA parameters with different population sizes.

tween runs with the same number of chromosomes is as great as the variation between those with different population sizes. There appears to be no strong benefit to using larger population sizes; therefore in future experimentation the GA will be set to run with 100 chromosomes.

9.8.3 Experimentation with Generations

To determine how many generations are necessary for the GA to reach a local minimum and whether running the GA for more generations will lead to an improvement on this minimum, the model was run for 500 generations using the default parameters in Table 9.5. This run was repeated 5 times to assess the variation in the final solutions for the optimal parameters. The model was not run for more generations than this due to the length of time for computation, for example, 500 iterations took approximately 23 hours (running on a 2.4GHz PC).

Figure 9.8 shows that the GA converges quickly; after 50 iterations it is close to its final solution. From this point onwards, improvement is slow. Once a solution has been homed in on, most of the genes will occupy a similar region of the search space and the introduction of

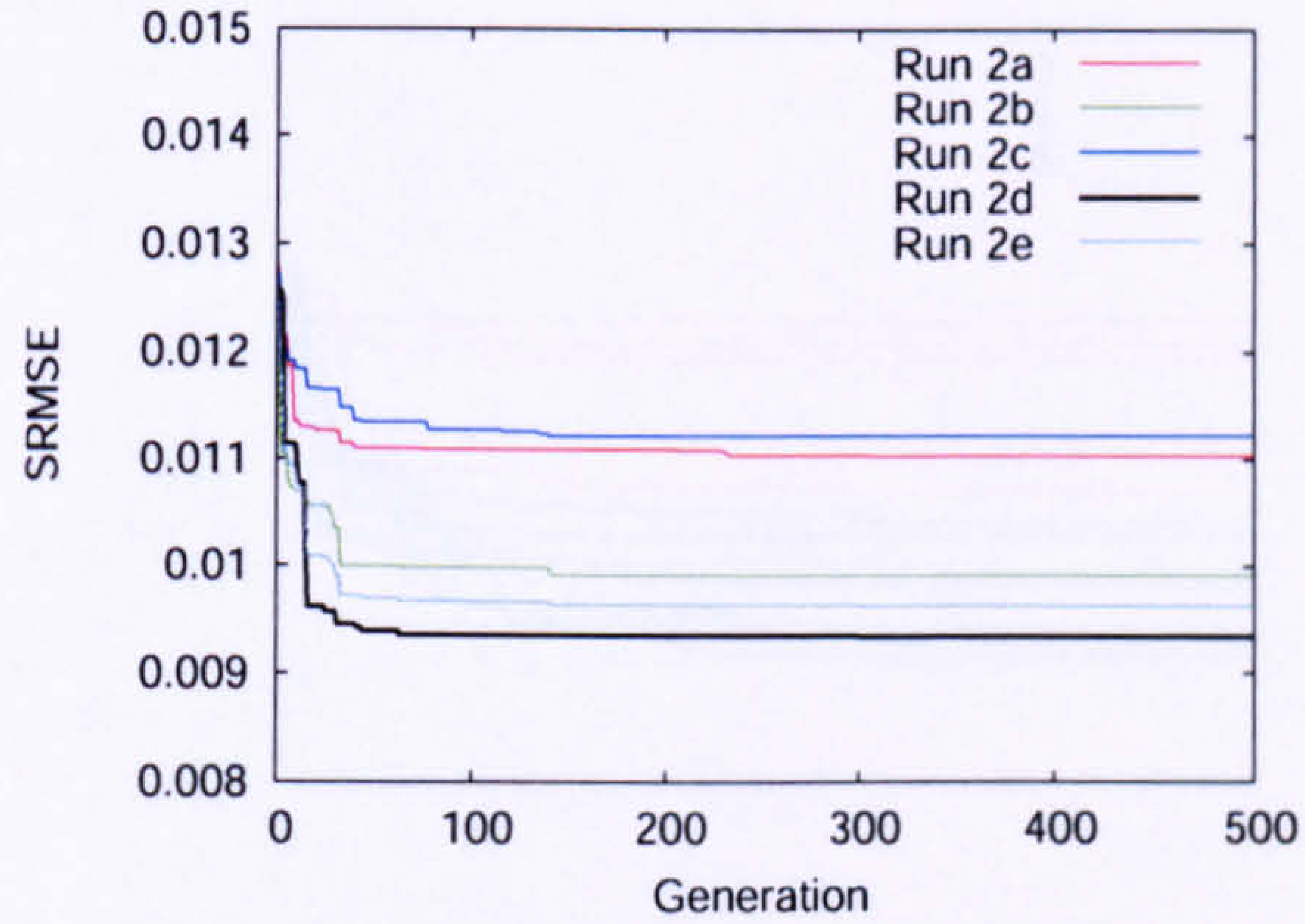


Figure 9.8: Graph of the SRMSE against time for several runs using the benchmark GA parameters but run for 500 generations.

new genetic material (through mutation) is unlikely to improve the results. On the basis of this, subsequent experiments are run with the GA set to run for 100 generations since there appears to be little benefit in using a larger number of generations. The additional computational power is better employed running a series of simulations with 100 generations each.

9.8.4 Experimentation with Mutations

Suboptimal convergence can occur when several highly fit (not optimal) individuals rapidly come to dominate the population. At this stage, mutation can be beneficial by widening the genetic search space (Goldberg, 1989). This reduces the algorithm to a slow random search. To avoid suboptimal convergence, the best solution can be kept at each generation with the remainder of the population generated by recombining the parents. This helps to retain genetic diversity. The selection of the best mutation from each chromosome at each generation rather than simply taking the best overall mutations also helps retain the diversity. Simulations within this section investigate the effect of varying the number of mutations of each chromosome tried at each generation.

Figure 9.9 shows that mutation improves the performance of the GA. The simulations with no mutations (indicated by the red line) produce a poorer performance than those with 3 or 5 mutated chromosomes. There is almost no difference in the performance of the GA with 3 or 5 mutated chromosomes. Some level of mutation is required to ensure continual diversity in the population even at later stages in the simulation and hence produce a better solution. The difference between simulations with 3 and 5 mutations is smaller than the variation between repeated runs with fixed initial GA parameters. There appears to be no significant advantage in using 5 mutations and since this increases the computation all subsequent experiments will only use 3 different mutations of each chromosome.

9.9 Model Parameter Variation

The analysis to date has focused on the SRMSE value produced by the GA. In this section, the GA predicted parameter values for the hybrid model will be examined in more detail. Results from the

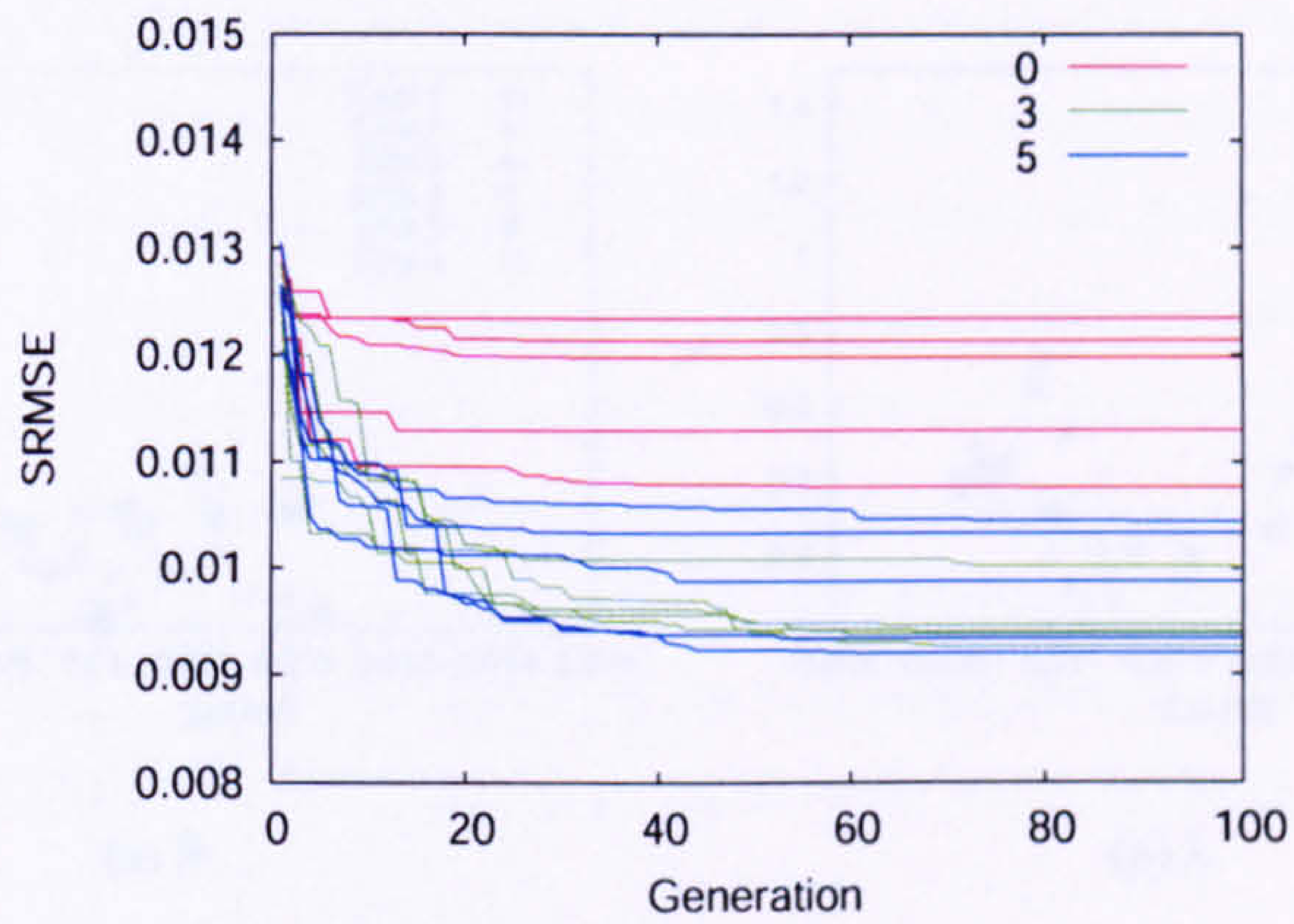


Figure 9.9: Graph of the SRMSE against time for several runs using the benchmark GA parameters but with different numbers of mutations following each recombination.

series of GA runs using both the default GA parameters and variations on these parameters will be used. Many of these experiments have already been discussed in §9.8; however a summary of the details is given in Table 9.6. Each experiment was repeated at least 5 times to give an idea of the variability in the solution.

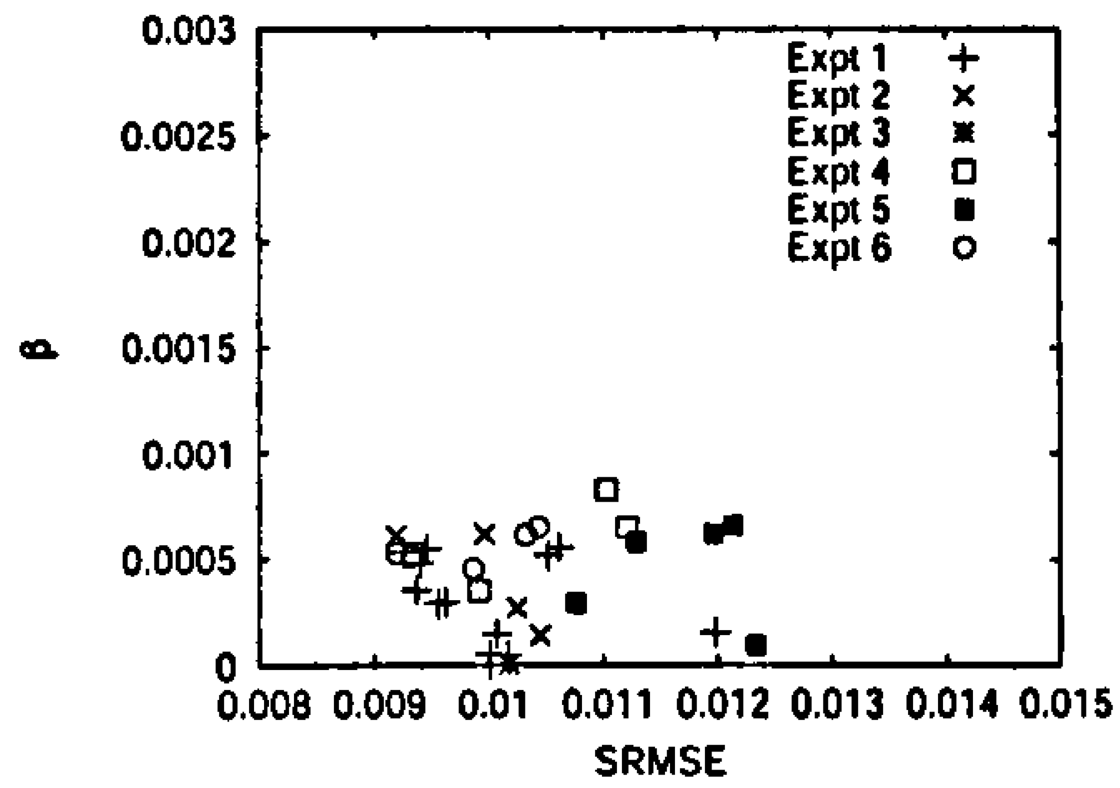
Experiment	Variation
1	Default
2	200 chromosomes
3	50 generations
4	500 generations
5	0 mutations
6	3 mutations

Table 9.6: Summary of the various experiments conducted with the GA. In each case the default benchmark values are used except for the variation shown.

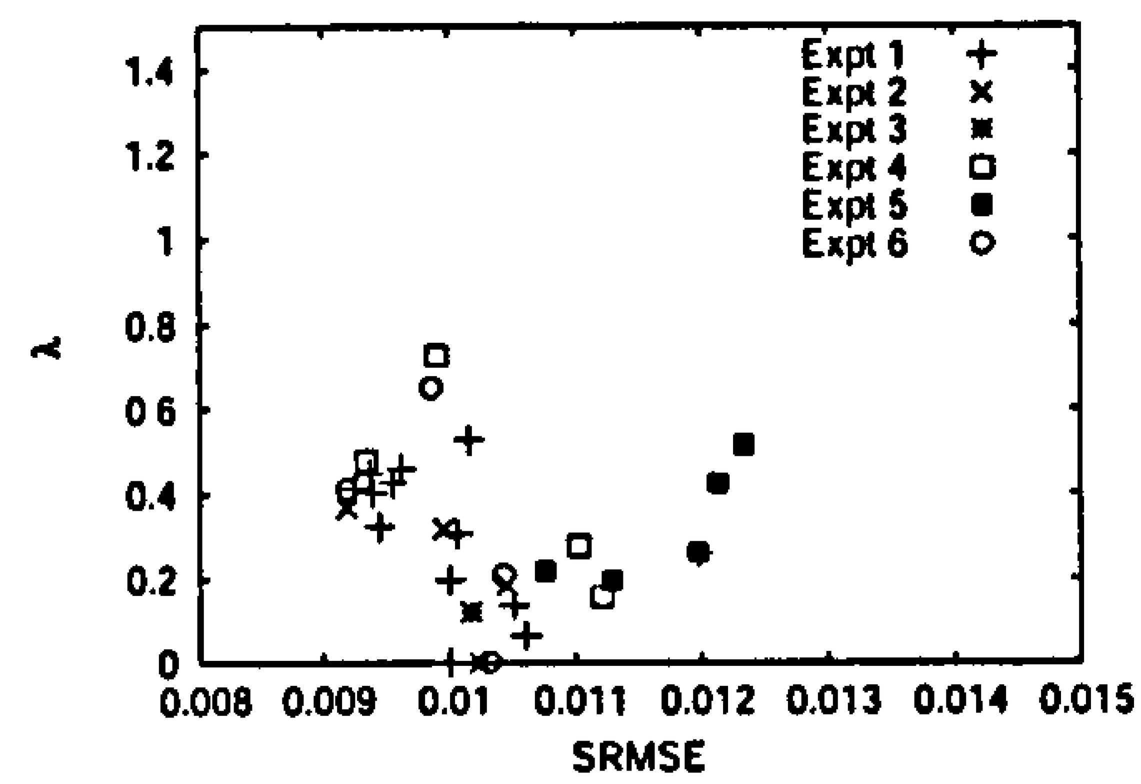
Figure 9.10 shows each parameter value as a function of the SRMSE. For each parameter there is a wide range of possible solutions that correspond to different local minima (see §9.1 for discussion on local and global minima). In addition, each parameter value has a cluster of results situated around the lowest value of the SRMSE. This clustering around one particular value for each parameter suggests that there is indeed a global minimum for this problem. The number of points in the cluster also suggests that provided an experiment with a given set of GA parameters is repeated 5 - 10 times, then there is a high probability that a good approximation to this minimum will be found. In general these clusters are not located near the edges of the allowed parameter range. This suggests that the specified ranges are wide enough so as not to interfere with the GA.

9.10 Comparison of Optimal Model Parameters

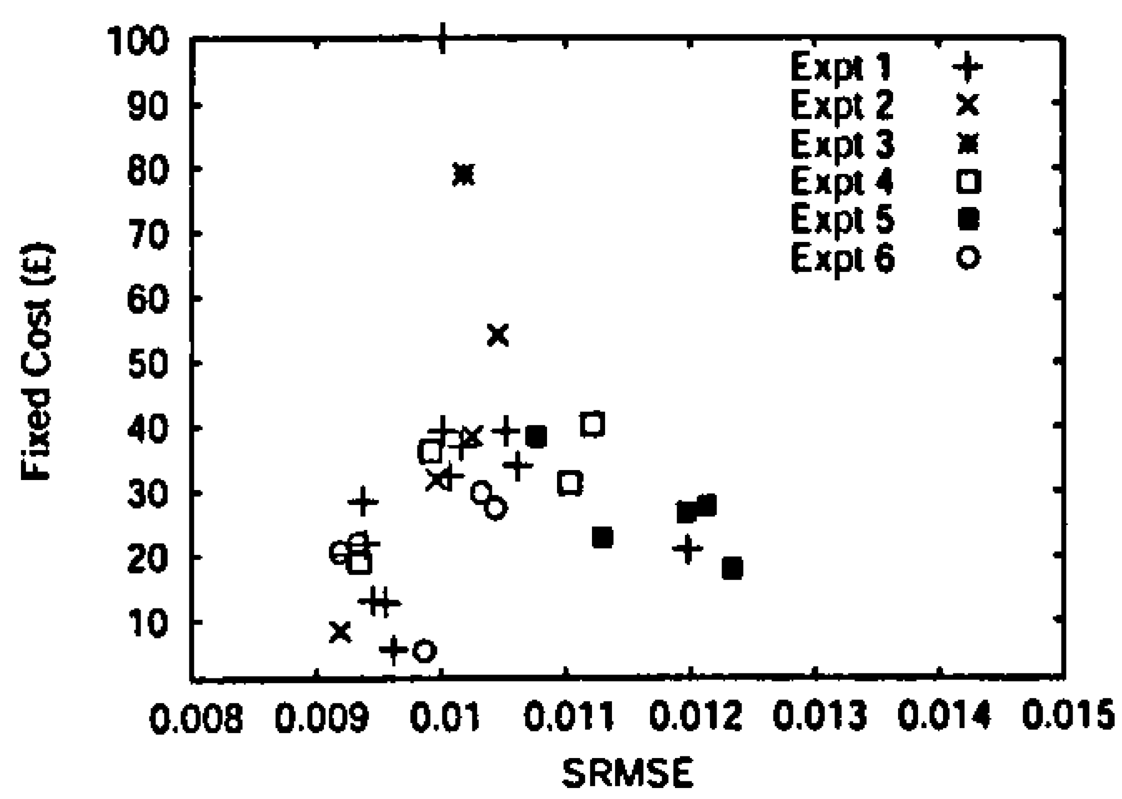
Research within this chapter has so far used the hybrid model with the GA. This was due to the additional computational time required to implement the network part of the model. (For a short



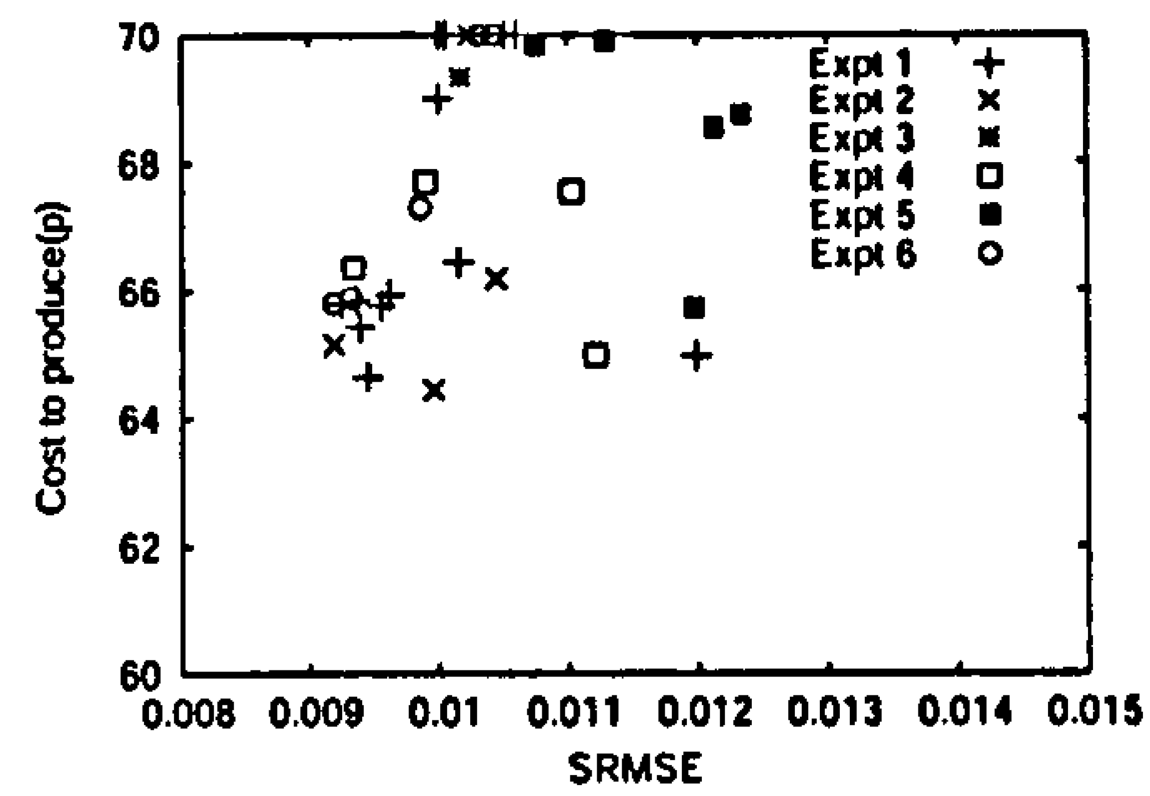
(a) β



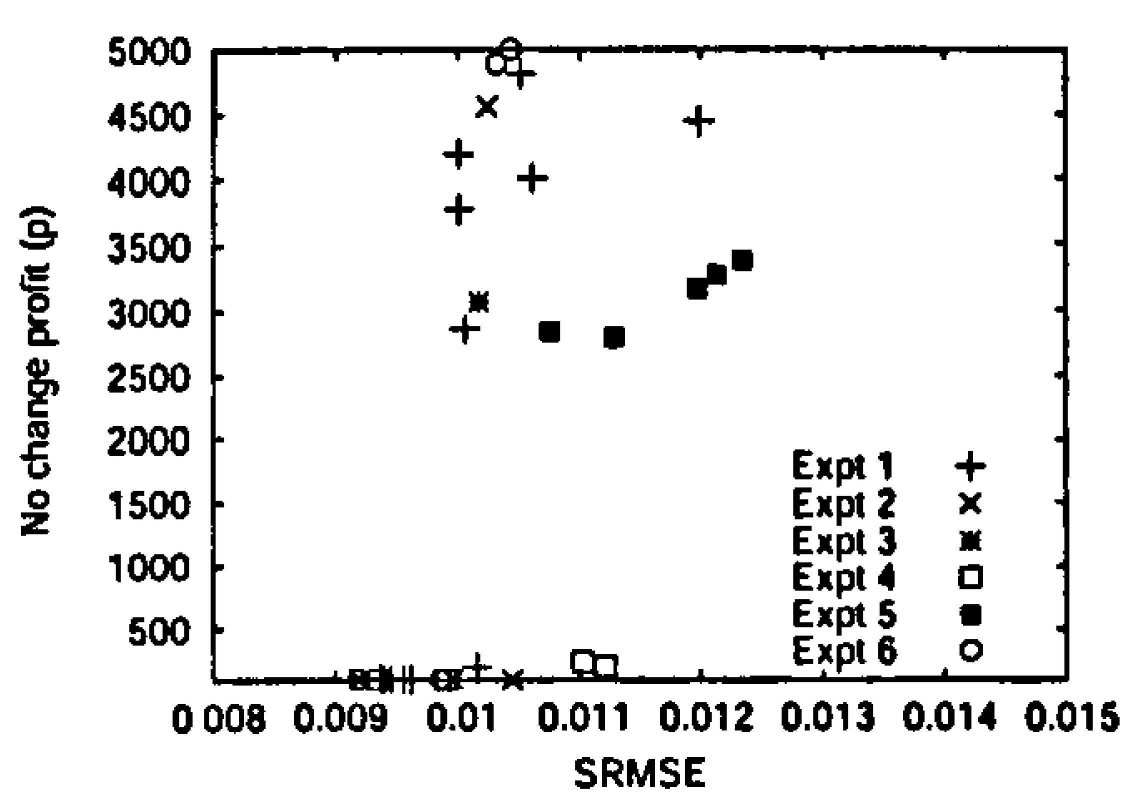
(b) λ



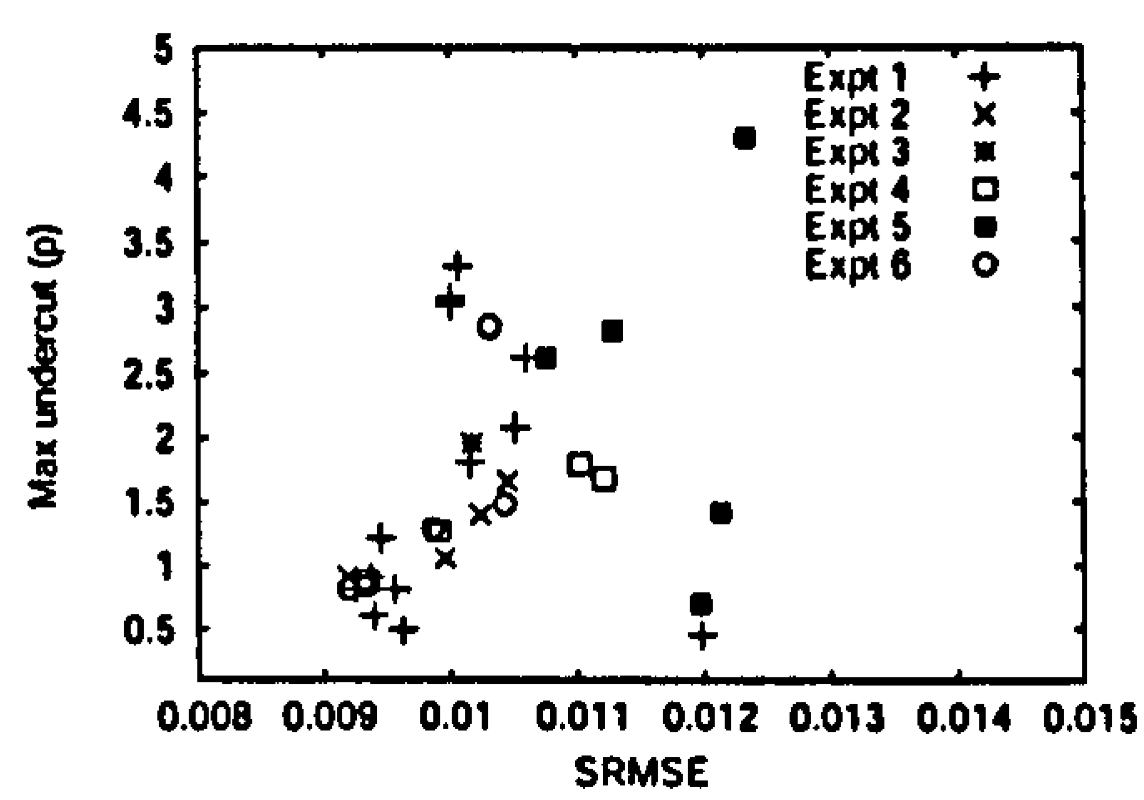
(c) *fixedCosts*



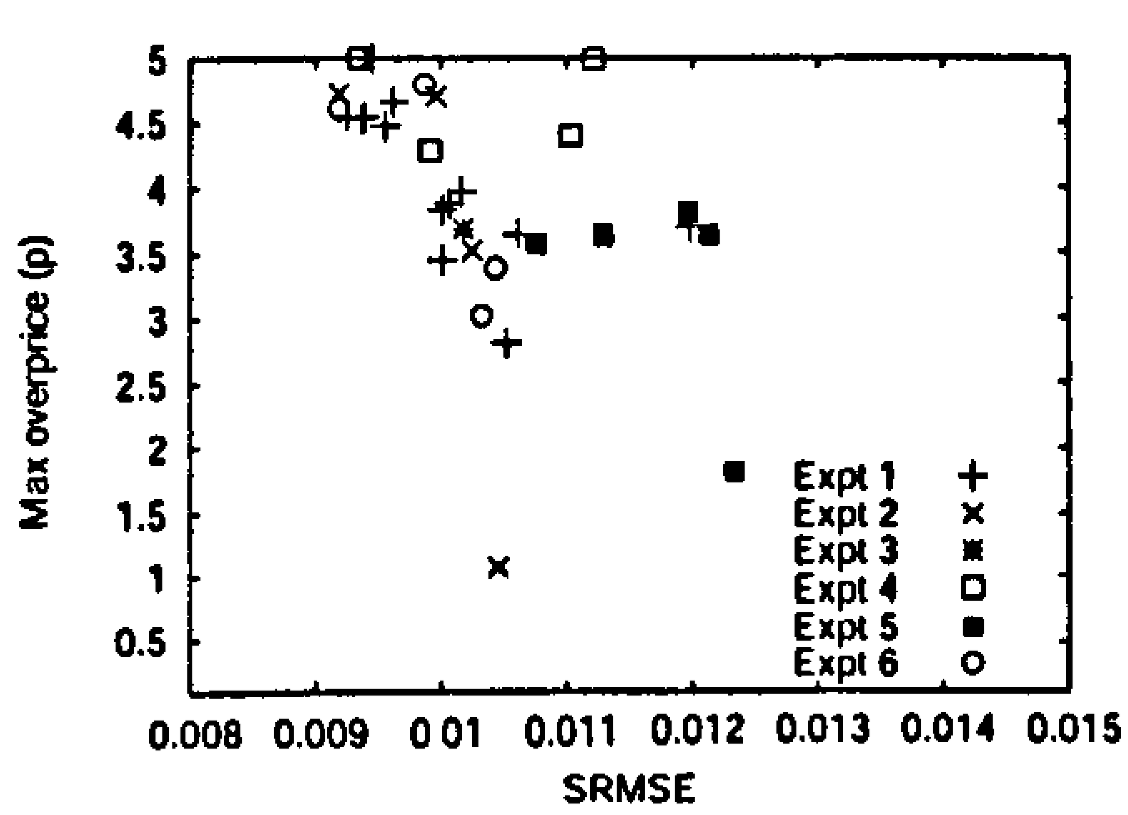
(d) *costToProduce*



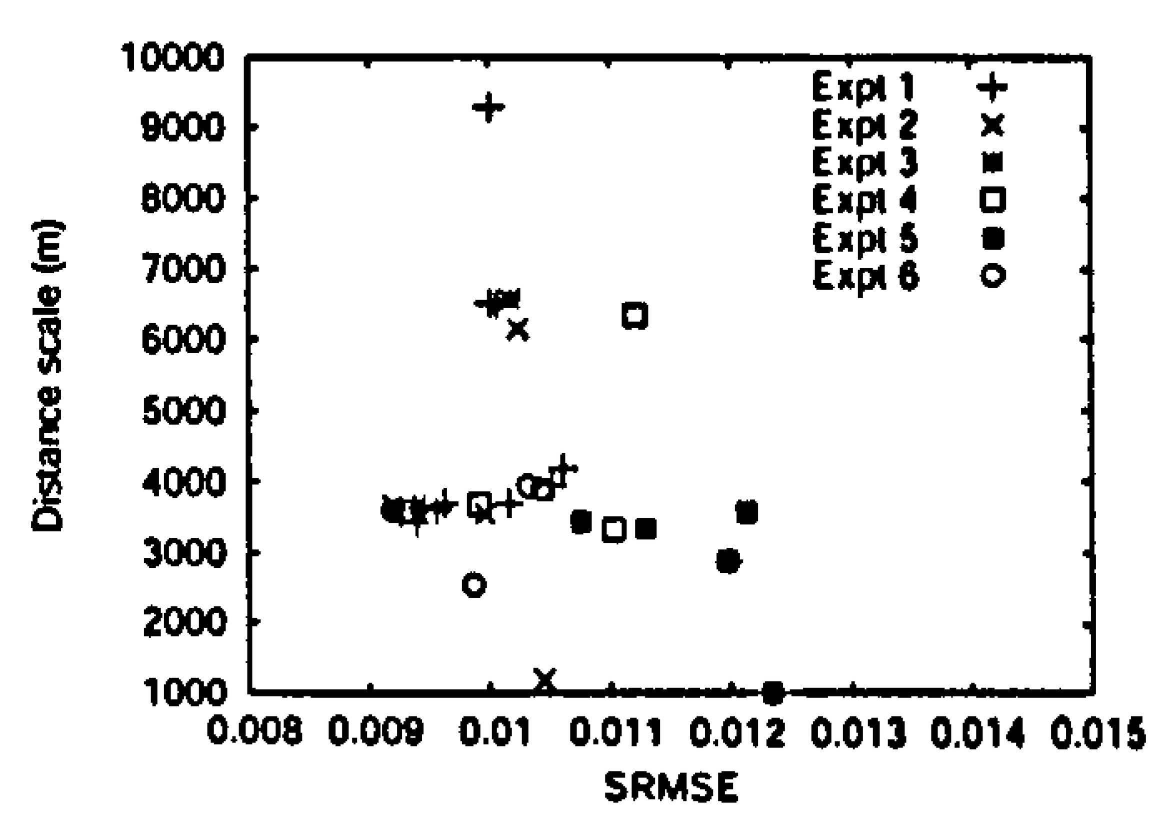
(e) *changeInProfit*



(f) *undercut*



(g) *overprice*



(h) *neighbourhood*

Figure 9.10: Variation in model parameters plotted against SRMSE for simulations with different GA control parameters. The various experiments are explained in Table 9.6.

run of 10 days the construction of the network routes can take well over half the computational time. A future enhancement to the model would be to alter the model so this routing was only done once at the start of each GA run.) In Chapter 8, the performance of the hybrid and network hybrid models was compared (see §8.3.1). Statistically, the network hybrid model performed well and it reproduced the trends in the real data to a higher degree of accuracy than the hybrid model. In this section, the model parameters derived in Chapter 8 for the network hybrid model (henceforth referred to as the Chapter 8 parameter values) will be compared with the optimal parameters values for the network hybrid model derived here using the GA (see below). The purpose is to determine whether the optimal parameter values produced by the GA are similar to the optimal parameters derived in Chapter 8.

The GA model was run 5 times and the best of these parameters were taken as the optimal parameters for the model. Figure 9.11 shows that the model parameters from the 5 simulations using the network hybrid model were very similar to those obtained using the hybrid model (see Figure 9.6).

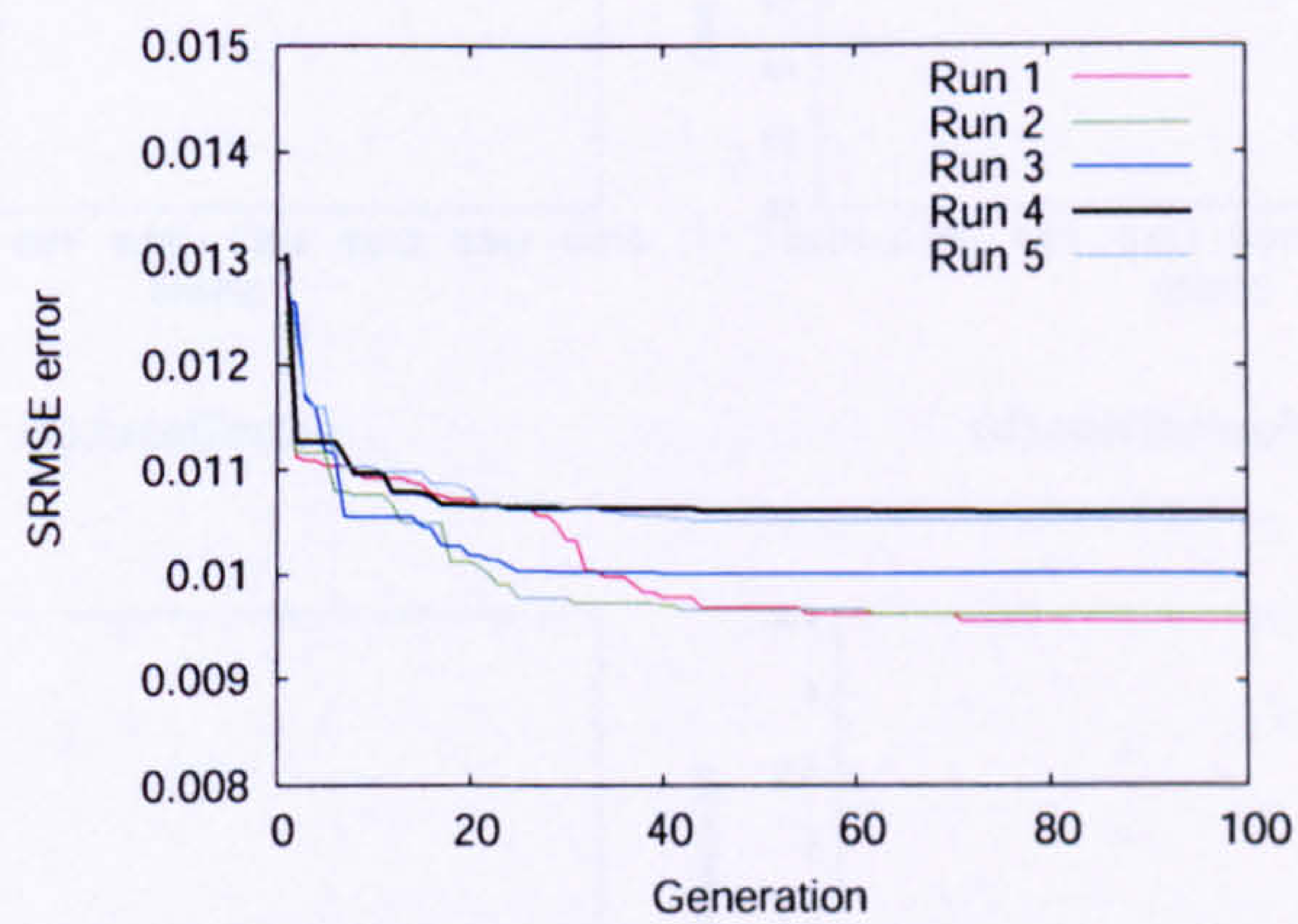
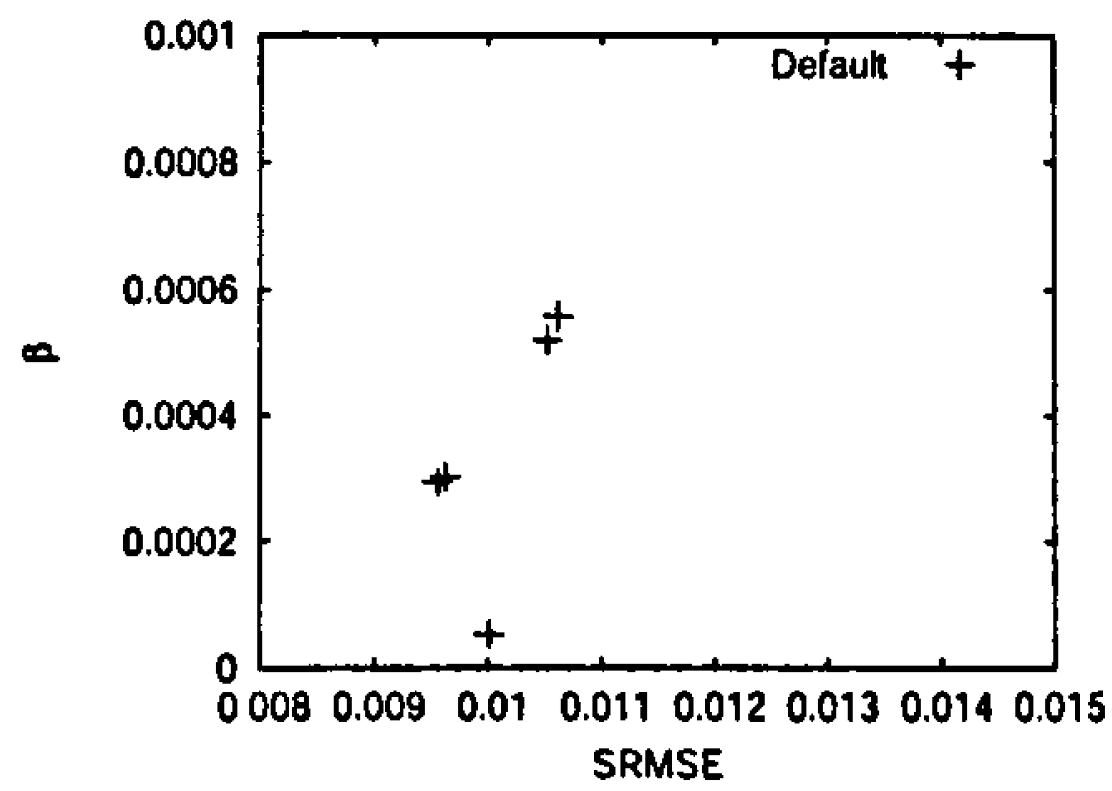


Figure 9.11: Graph of the SRMSE against time for several runs with the benchmark GA parameters and using the network hybrid model.

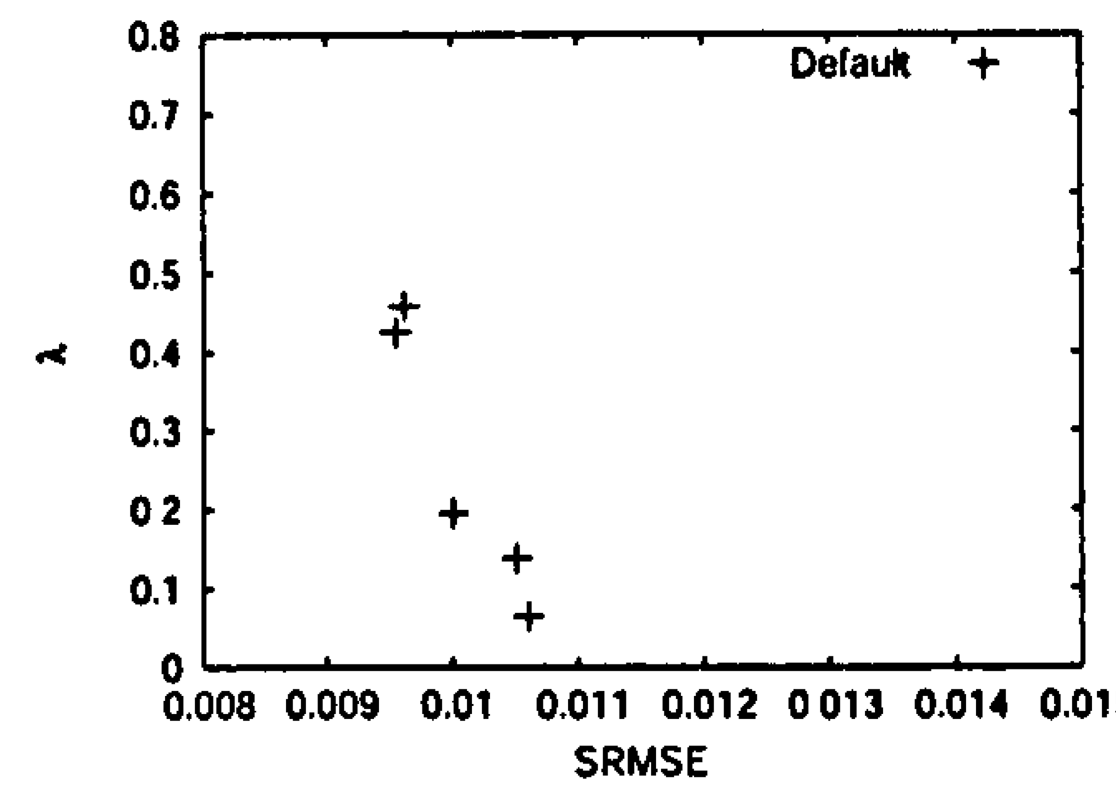
Table 9.7 details the optimal parameters derived in Chapter 8 (the same parameter values were used for the hybrid and network hybrid model) and the optimal parameters suggested by the GA. Also shown is the SD of the 5 GA runs to give an indication of the spread of values obtained using the GA. Figure 9.12 shows the spread of values for each parameters as a function of the SRMSE.

The GA values for β and λ were both reasonably close to the parameters derived in Chapter 8. As evident in Figure 9.12(a) and (b), neither of these values produced definite clusters for the hybrid model. This means that reasonable solutions can be produced from a wide range of values. Similar results are seen in Figure 9.10 for the hybrid model. The standard deviation values in Table 9.7 also support this conclusion. The SD for β is the same size as the optimal value. For λ the SD is smaller, but still 40% of the optimal value. The optimal values for β and λ can be converted into distance and price scales using Equations (7.1) and (7.2). This gives scales of 3.5km and 1.7p over which the sales predicted by the SI model be reduced by a half.

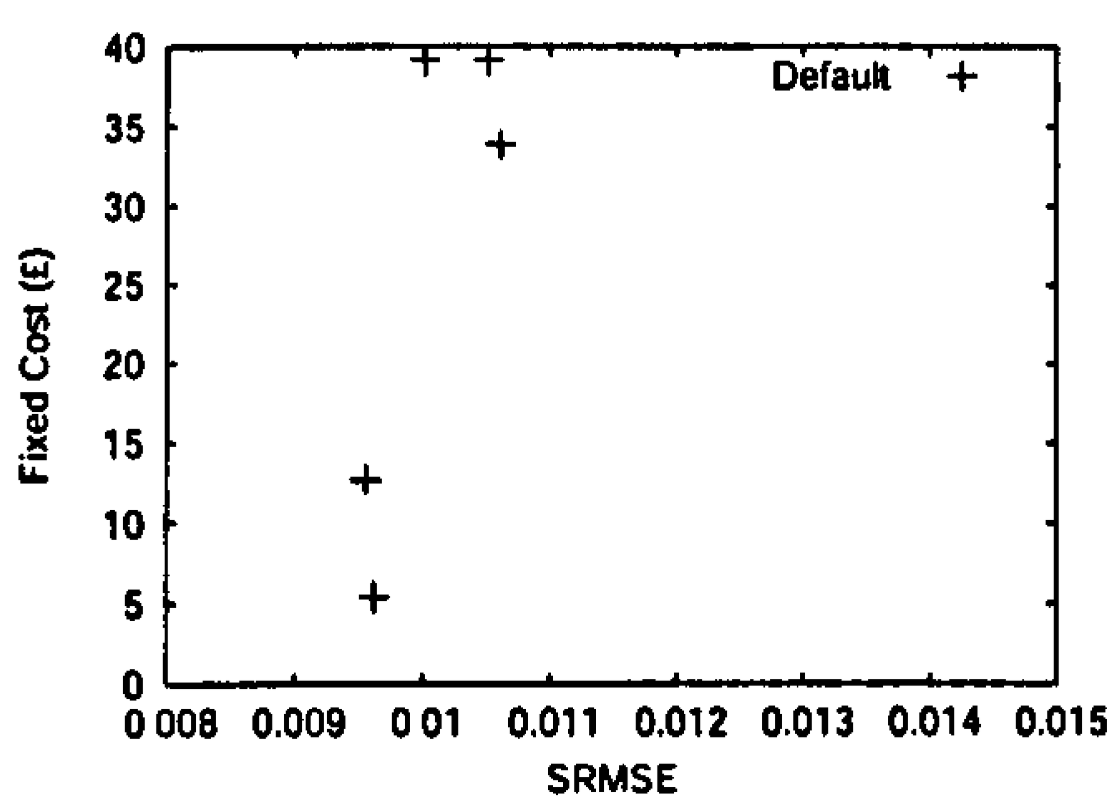
The GA gave the *neighbourhood* parameter a slightly smaller value than the Chapter 8 values. The values are also slightly more clustered than those for β and λ, with a SD of 1.2km and an



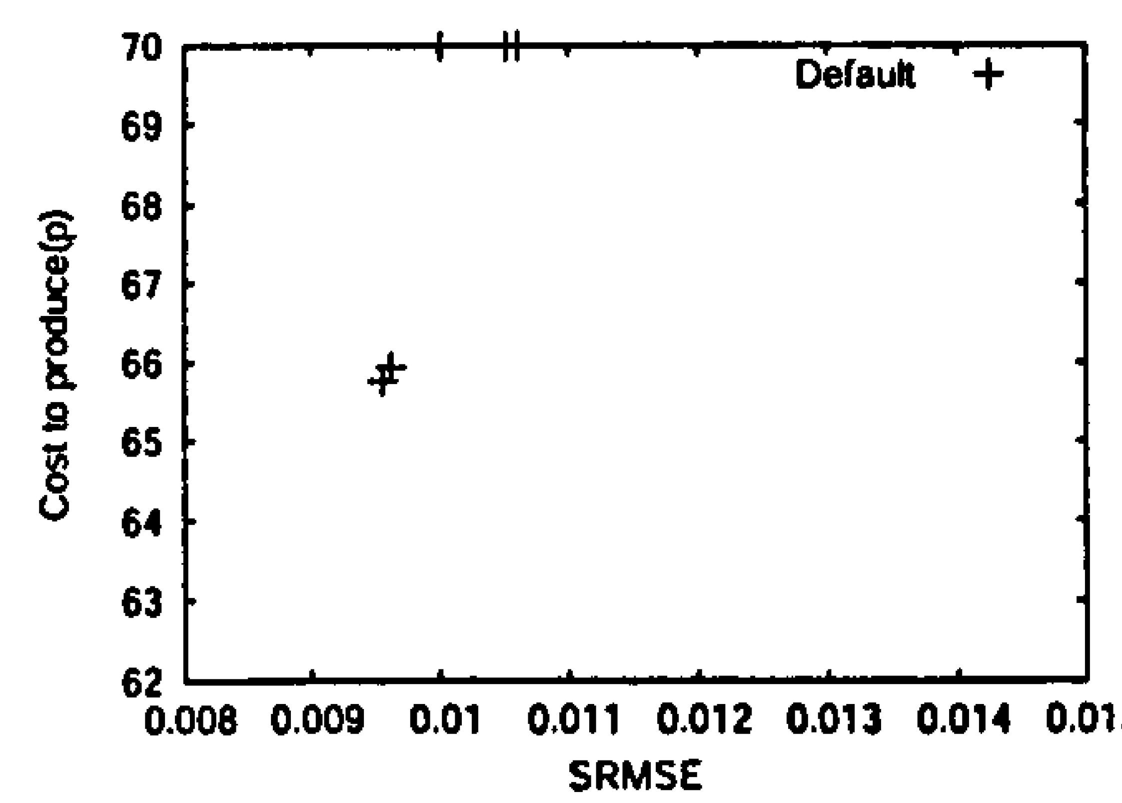
(a) β



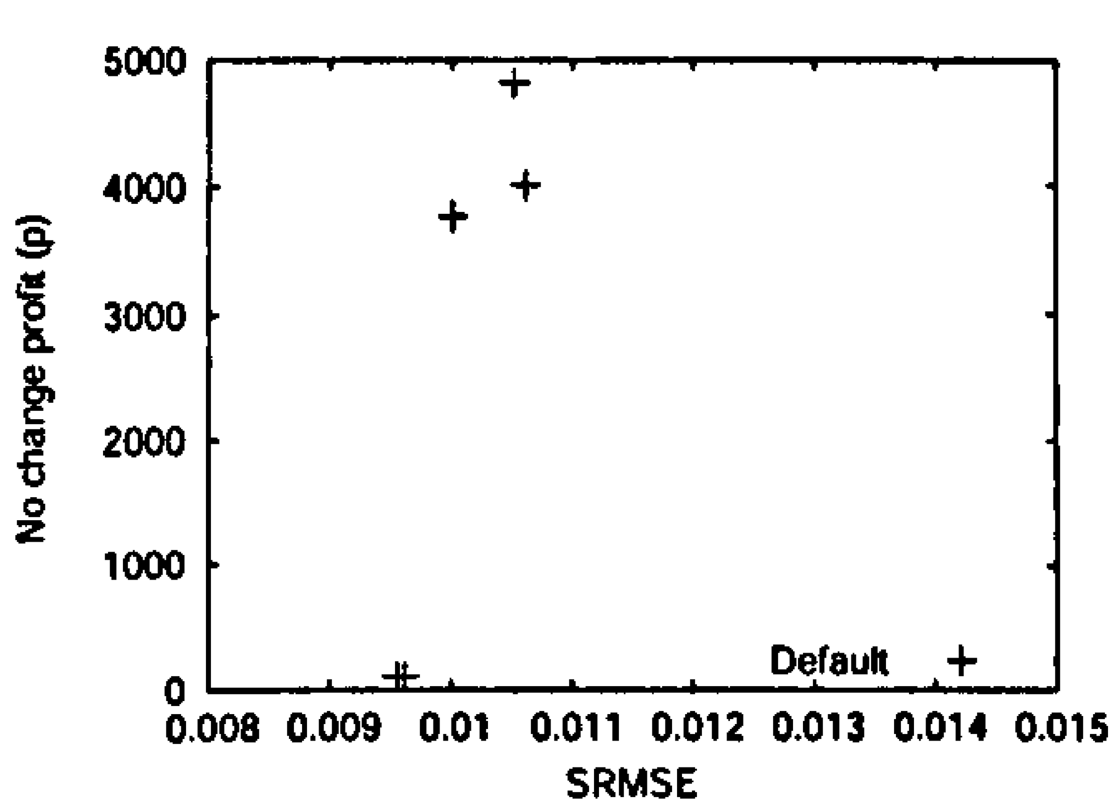
(b) λ



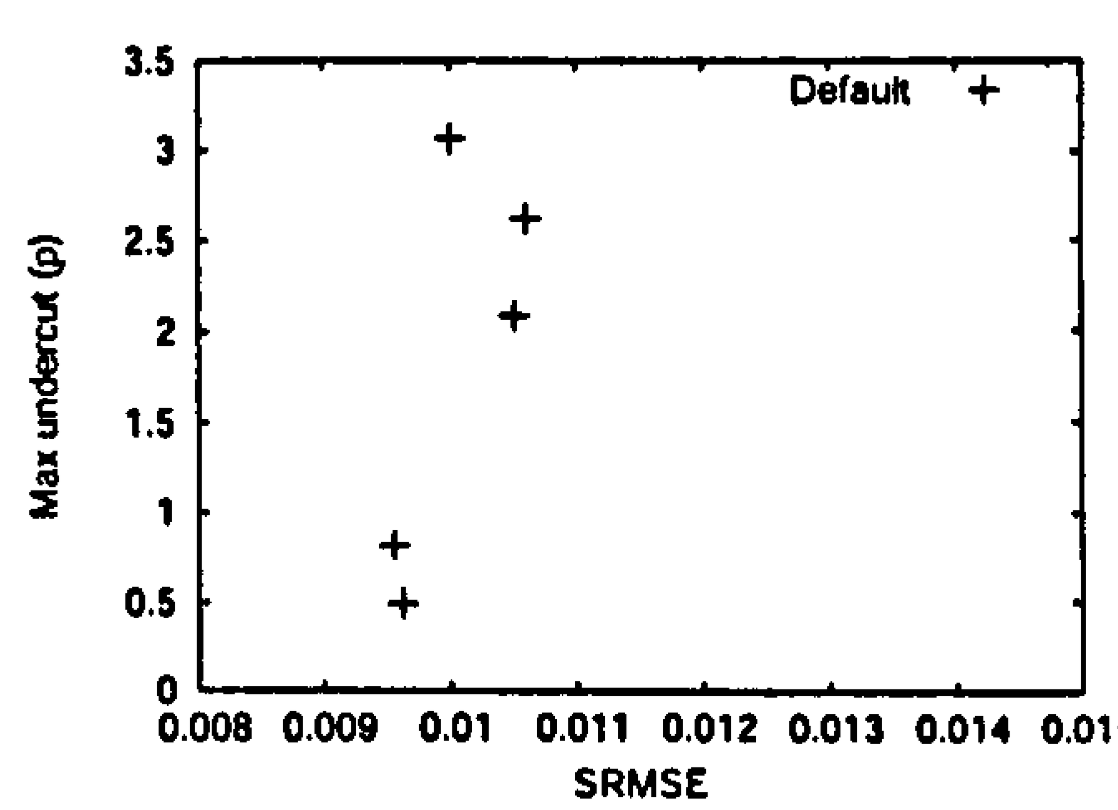
(c) *fixedCosts*



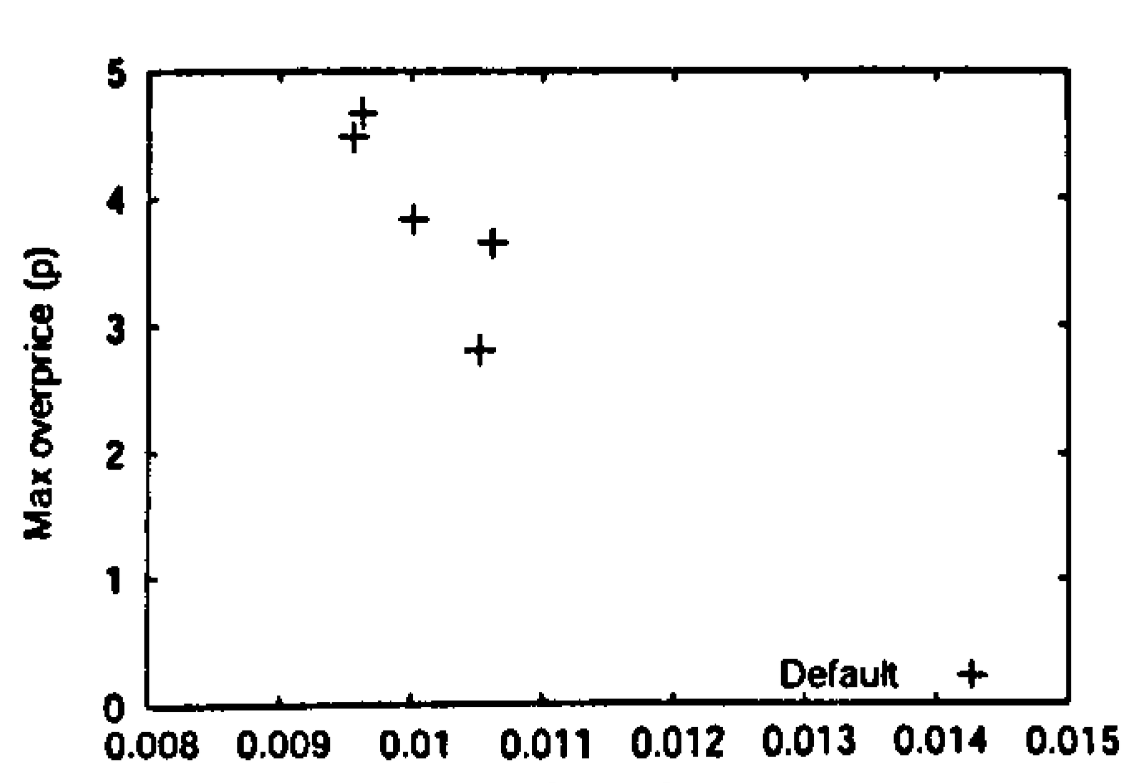
(d) *costToProduce*



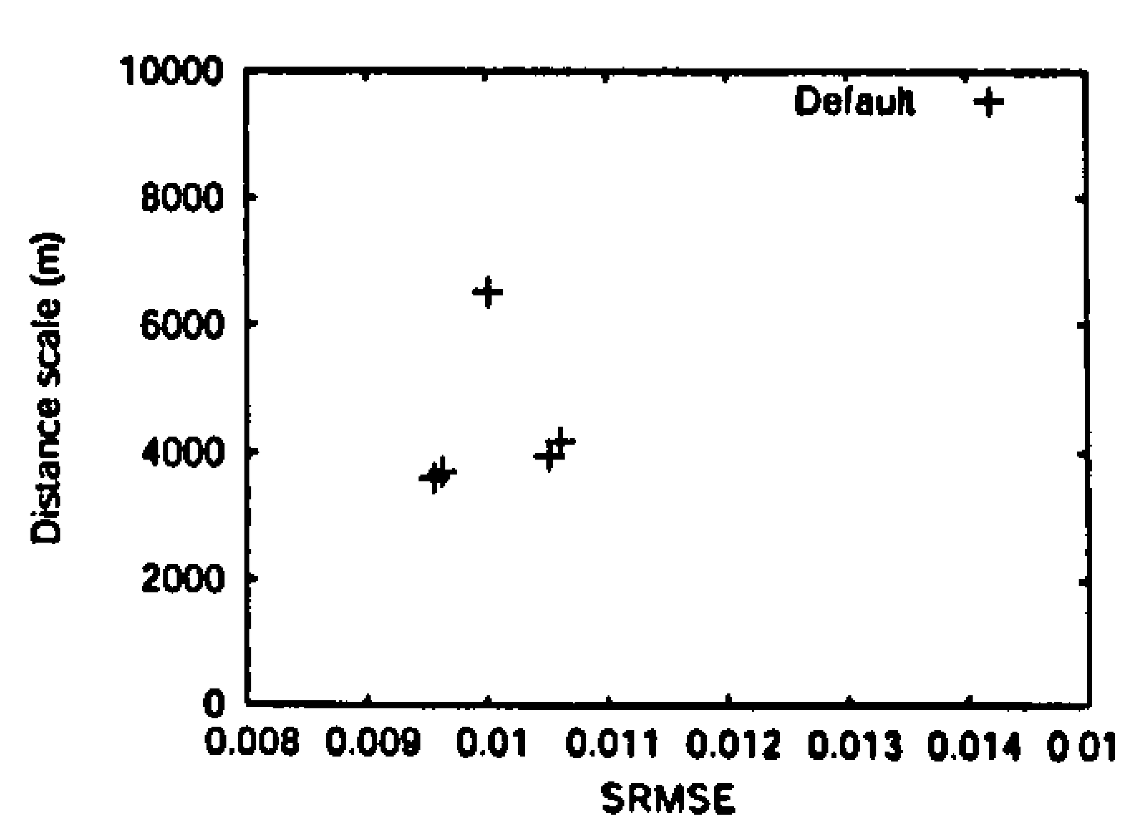
(e) *changeInProfit*



(f) *undercut*



(g) *overprice*



(h) *neighbourhood*

Figure 9.12: Variation in model parameters plotted against SRMSE for simulations with different GA control parameters. The experiments (described in §9.10) are using the network hybrid model to develop rules.

Parameter	Chapter 8 Value	GA Value	SD (5 GA runs)
β	0.0003	0.0002	2.04×10^{-4}
λ	0.7	0.4	0.176
fixedCosts **(£)**	80.00	12.73	15.9
costToProduce **(p)**	66	65.75	2.27
changeInProfit **(£)**	40	1	22.8
overprice **(p)**	5	4.4	0.73
undercut **(p)**	1	0.8	1.12
neighbourhood **(km)**	5.0	3.6	1.21

Table 9.7: Comparison of optimal model parameters from Chapter 8 and from the GA.

optimal value of 3.6km (Figure 9.12(h)). The *overprice* parameter was also relatively well clustered and in good agreement with the Chapter 8 value (Figure 9.12(g)). With the hybrid model the value for *undercut* (Figure 9.10(h)) was also quite tightly clustered, but with the network hybrid model the clustering was less clear (Figure 9.12(f)). This may be a result of the relatively small sample size (5 runs) with the network hybrid model. The optimal value of 0.8p is still close to the value of 1p used in Chapter 8. The optimal *undercut* parameter is much smaller than the *overprice* parameter, as suggested in Chapter 8.

There is a large difference between the GA predicted value (£1) and the Chapter 8 value (£40) for the *changeInProfit* (Figure 9.12(e)). In the experimentation in Chapter 8, the aim was to select a large enough value for the *changeInProfit* to enable the solution to reach a steady state after approximately 25 days (see §7.3.3 and §8.2.3). However, the GA only runs the model for 10 days and therefore the long term equilibrium is not important in determining the optimal *changeInProfit* value. For short term model predictions, a smaller *changeInProfit* value provides a greater scope for stations to change their price in the model. There is also a very large SD (£22.80) for the parameter indicating a large degree of uncertainty in the optimal value.

The *costToProduce* and *fixedCosts* parameters are closely linked (see §7.2.3) and will therefore be assessed together. The overall cost is a combination of these two values. The GA optimal value for *costToProduce* almost matched the Chapter 8 value. However, the GA values were not tightly clustered (Figure 9.10 (d)) with a SD of 2.27p, again suggesting that a reasonable solution can be obtained from a range of values. The *costToProduce* value is strongly related to the *fixedCosts* value. The GA predicted value (£12.73) is considerably smaller than the value derived for use in Chapter 8 (£80). This was one of the parameters that little information was known about. However, in Figure 9.10(c), the best solutions for the *fixedCosts* are not tightly clustered - they span a £20 range). This suggests that this parameter does not have as great an effect on the overall solution. The *fixedCosts* parameter will affect the profit level the same way at all stations and so it will have less effect on the changes in profit (which control the pricing strategy).

In summary, the values derived on the basis of real data analysis (Chapter 4) and numerical experimentation (Chapter 7) provided a good set of parameter values for the network hybrid model in Chapter 8. The parameters produced by the GA are not constrained to agree with the Chapter 8 parameters, they are merely chosen to minimise the SRMSE of the model. It is reassuring that the parameters obtained by optimising the model using a GA are in close agreement with those obtained independently in Chapter 8 based on features of the real petrol market. This suggests that the model is behaving in a realistic manner and that the assumptions made in deriving the parameter values in Chapter 8 were reasonable.

9.11 Performance on West Yorkshire

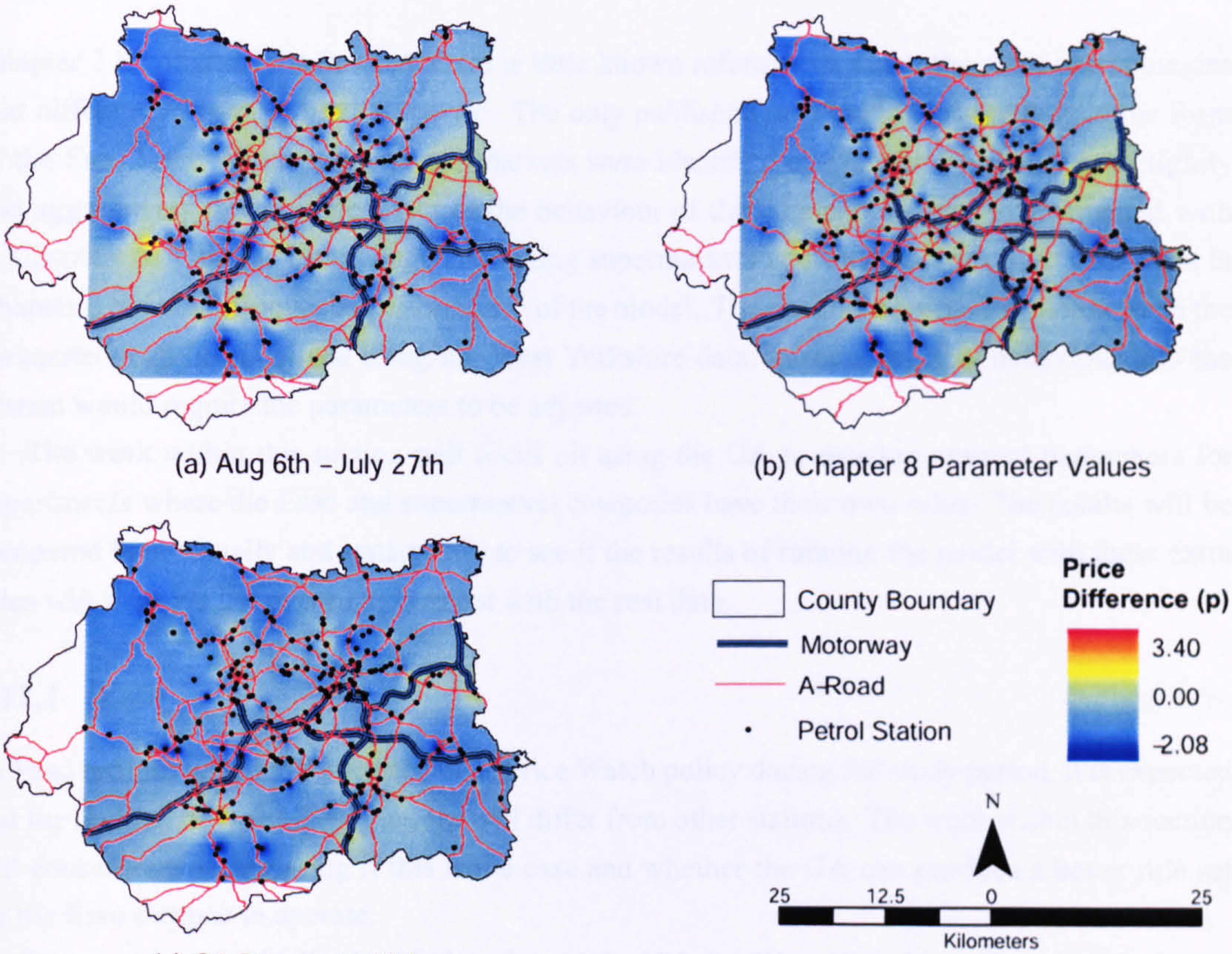
Assessing the SRMSE and parameter values provides an indication of the performance of the model. However, the best test is to see whether the spatial variations within the real data are reproduced. The network hybrid model was initialised with the real data for July 27th and given the optimal parameters produced by the GA (detailed in Table 9.7).

Figure 9.13 presents a comparison of the mean price differences of the real data (between July 27th (day 0) and August 6th), the network hybrid model (day 10 and August 6th) and the GA parameter values (day 10 and August 6th). A perfect match with the real data on August 6th would be indicated by a light blue surface across the entire area.

Figure 9.13(c) shows that the GA parameters are reproducing price variation in the real data more accurately than the constant data or the network hybrid model (Figures 9.13(a) and (b)). For example, in urban areas, both the constant data and network hybrid model display positive price variations. The GA displays fewer price differentials within these areas. However, there are still areas, for example to the east and south-east of the study area, that are being over-predicted, possibly due to edge effects or other unknown factors.

Assessment of the mean price over time (Figure 9.14(a)) shows that the Chapter 8 parameters are on average, performing better than the GA parameters. However, both solutions over-predict the real data. On day 9 (August 5th), the GA is over-predicting by approximately 0.4p (i.e. the model is 0.4p higher than the real data), whilst the Chapter 8 parameters are only 0.1p overpriced. However, on day 10 (August 6th) the GA over-prediction price has fallen to 0.05p. This is due to a marked increase in price between August 5th and August 6th which results in the GA becoming closer to the real data at the end of the period than the Chapter 8 parameters. Since the GA assesses the fitness of the model parameters by comparing with the real data on August 6th it would be expected to perform well on that day. There is little difference in the standard deviations of the GA, real and Chapter 8 prices (denoted by the vertical bars). This indicates that a similar range of prices are being produced in each case.

Figure 9.14(b) shows the SRMSE of the different parameter values over time. The overall model performance for the GA and Chapter 8 parameters are very similar over the first part of the 10 day period and also in close agreement with results assuming the prices remain constant at their values on July 27th (denoted by Constant price on the graph). However, after July 29th the SRMSE of the GA solution decreases and remains lower than the Chapter 8 parameter and the constant price data for the rest of the period. This indicates that the GA parameters are better at representing the real data than either of the other two sets of results.

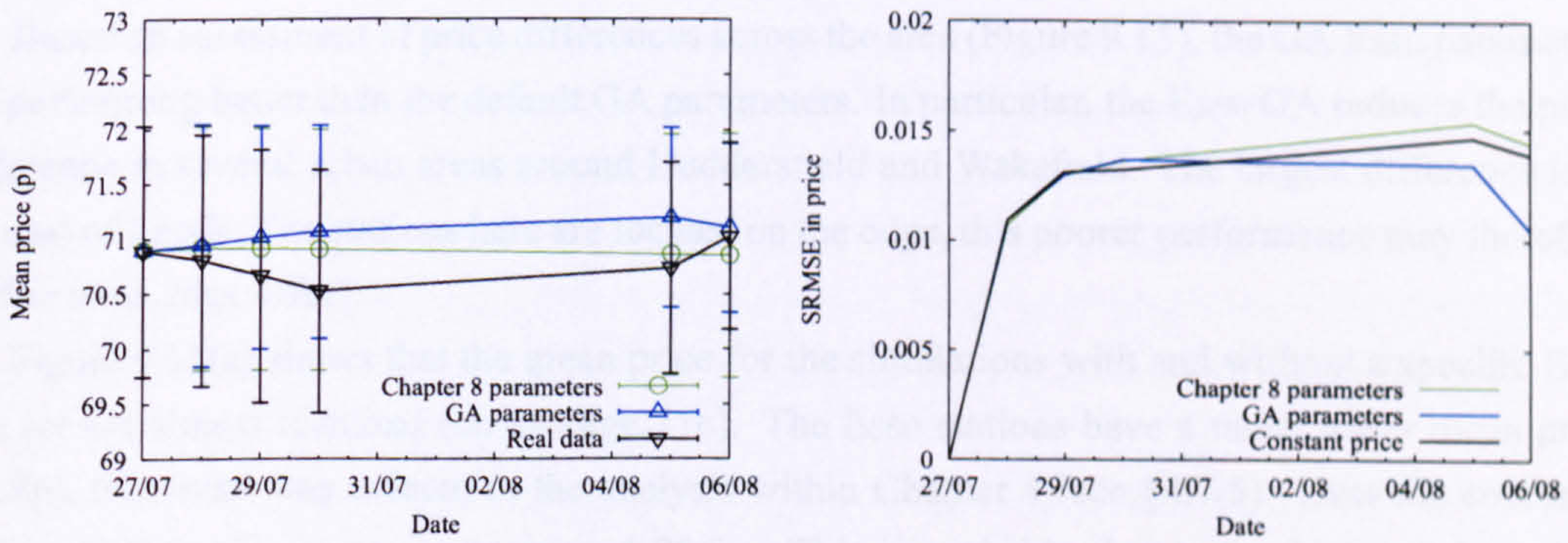


(a) Aug 6th -July 27th

(b) Chapter 8 Parameter Values

(c) GA Parameter Values

Figure 9.13: Price difference between the real data (day 10) and model output using (b) the Chapter 8 parameters and (c) the GA parameters for West Yorkshire. The difference for the real data between the first and last days (July 27th and August 6th) (a) is included for comparison.



(a) Mean Price and SD

(b) SRMSE

Figure 9.14: Comparison of the (a) mean difference and (b) SRMSE over time between the real data, model results using Chapter 8 parameters and model results using the GA parameters. In each case the model was initialised with the real data from July 27th. Standard deviations in the mean price are represented by vertical bars in (a).

9.12 Further Experimentation

Chapter 2 highlighted the fact that there is little known information about the rules and strategies that different companies/brands operate. The only published information available is in the form of the Esso Price Watch policy. Supermarkets were identified in Chapter 4 as being both tightly and aggressively priced. The effect of the behaviour of these categories was experimented with in Chapters 5, 7 and 8. Surprisingly, assigning supermarkets and Esso stations their own rules in Chapter 8 did not improve the performance of the model. The conclusion was drawn that since the parameters had been derived using the West Yorkshire data, introducing new behaviour into the system would require the parameters to be adjusted.

The work within this section will focus on using the GA to develop optimal parameters for experiments where the Esso and supermarket categories have their own rules. The results will be compared both visually and statistically to see if the results of running the model with these extra rules will improve the overall agreement with the real data.

9.12.1 Esso

As Esso are known to have operated their Price Watch policy during the study period, it is expected that the optimal rules for these stations will differ from other stations. The work within this section will concentrate on assessing if this is the case and whether the GA can produce a better rule set for the Esso stations to operate.

The model was initialised with data from July 27th for West Yorkshire. Two rule sets were created, an Esso rule set (for the Esso stations) and a default rule set (for the remainder of the stations). This created an additional set of parameters that the GA had to optimise (with the exception of β and λ which apply to all the stations). The GA was run 10 times using the default GA parameters (see Table 9.8). The best run is used for the analysis within this section. Appendix E.1 shows graphs of the optimal model parameters from the GA against SRMSE for all 10 of the GA runs undertaken in this section.

Based on assessment of price differences across the area (Figure 9.15), the GA Esso parameters are performing better than the default GA parameters. In particular, the Esso GA reduces the price difference in several urban areas around Huddersfield and Wakefield. The largest difference is to the east of Leeds. The stations here are located on the edge, this poorer performance may therefore be due to an edge effect.

Figure 9.16(a) shows that the mean price for the simulations with and without a specific Esso rule set are almost identical (on average 71p). The Esso stations have a much lower mean price (69.9p), this trend was evident in the analysis within Chapter 4 (see §4.7.5). Over the course of the simulation, this increases to around 70.5p. This is probably due to the higher price of the non-Esso stations. Under these rules, there is scope for the Esso stations to increase their prices (and hence profit) whilst still remaining cheaper than the competition. The range of prices are also similar as indicated by the vertical bars. The use of a separate rule set for the Esso stations is not having a large effect on the performance of the model. This is confirmed by the SRMSE. The total SRMSE for the GA derived parameters with and without separate Esso rules are almost identical; the default total is only slightly lower (Figure 9.16(b)).

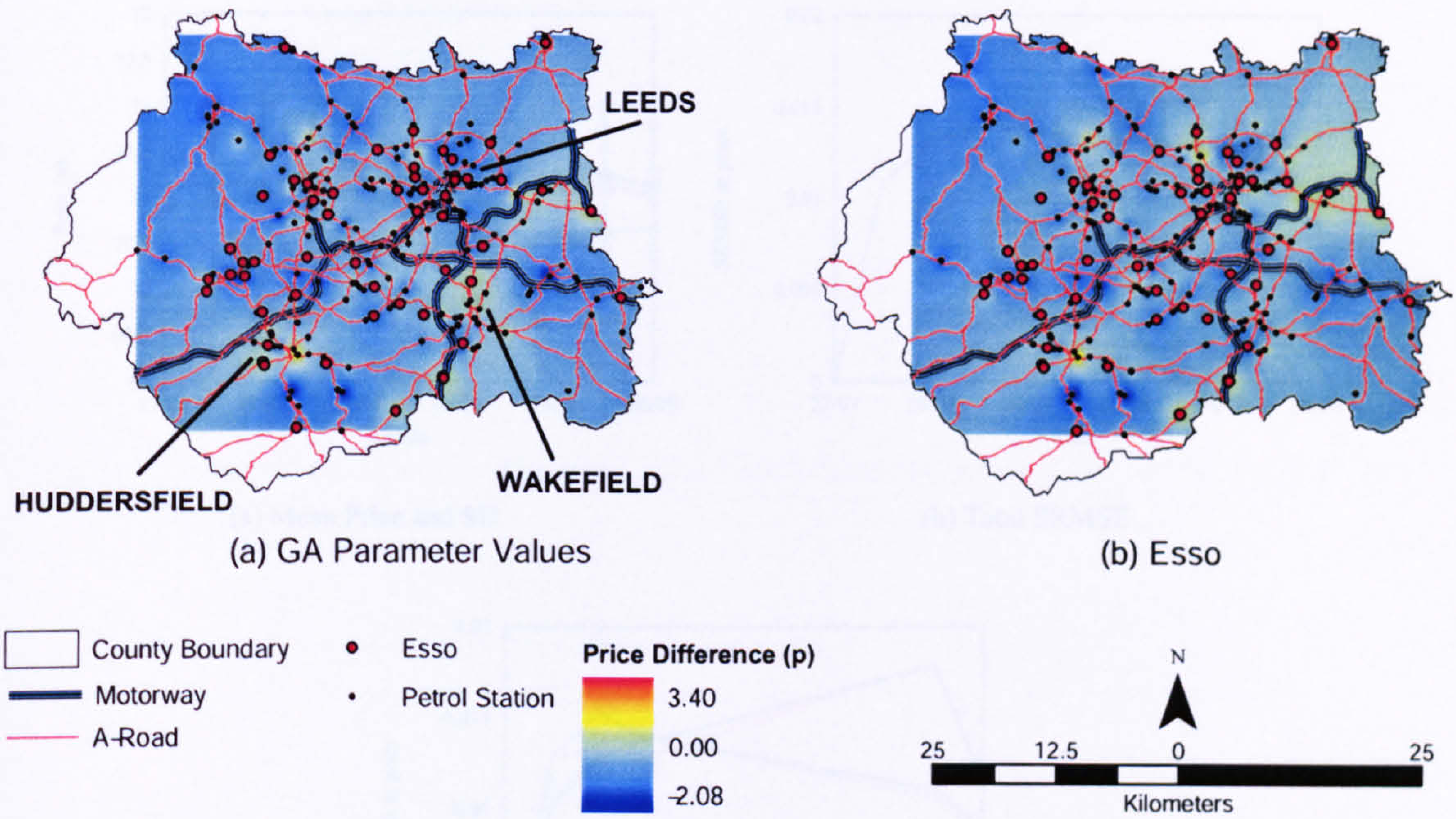
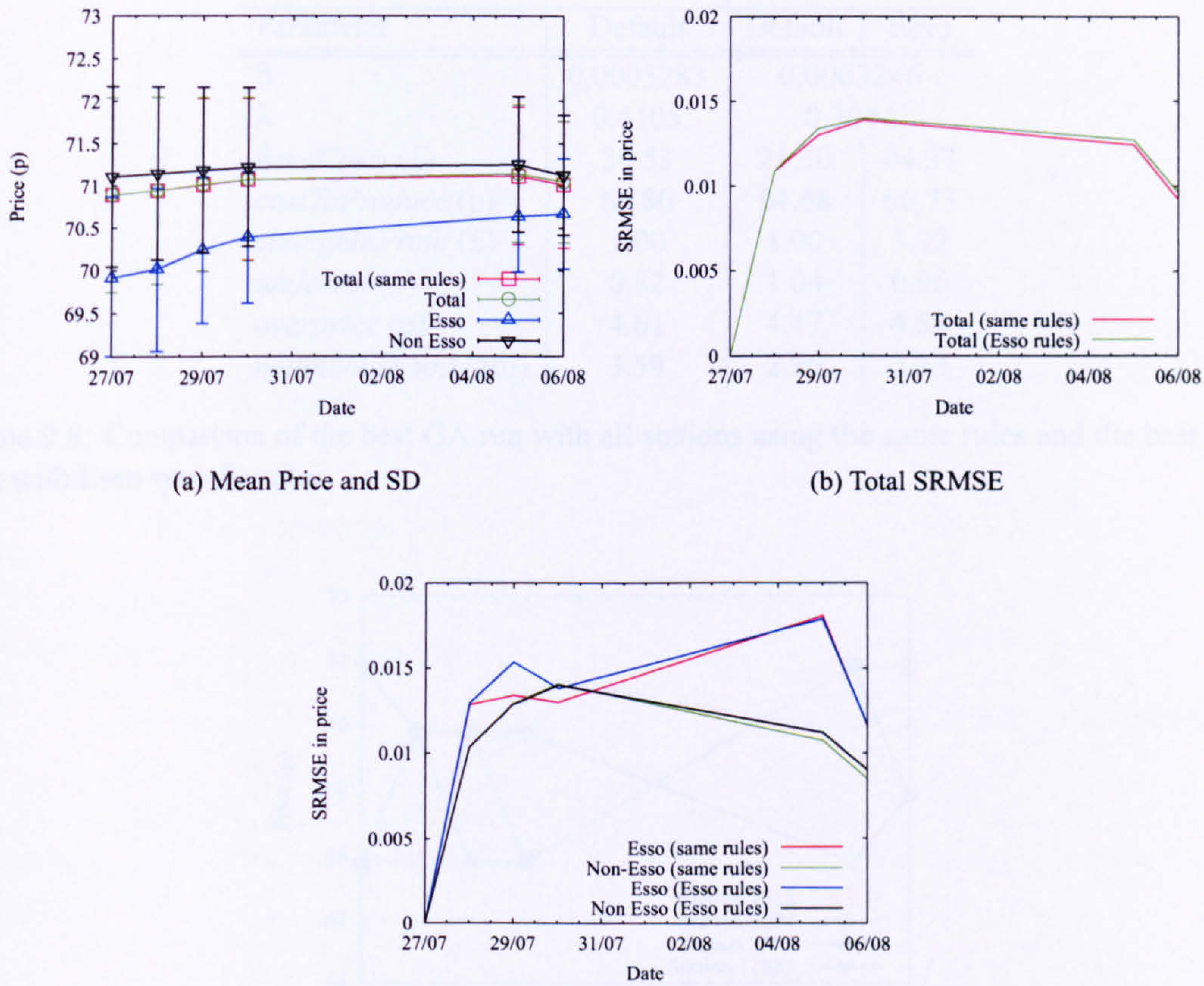


Figure 9.15: Map showing mean price differences between real and model data on August 6th for model runs with (a) the same rules applied to all stations and (b) separate rules for Esso stations.

Figure 9.16(c) shows the separate SRMSE for the Esso and non-Esso stations. The SRMSE of the Esso stations increases over the course of the simulation to 0.0175 on August 5th before dropping to around 0.012 on August 6th. This indicates that the Esso stations are giving a worse than average performance even with their own rule set. Conversely, the non-Esso stations are actually doing better than the default total, these stations are performing better with a rule set of their own.

Table 9.8 presents a comparison of the best GA run with all stations using the same rules and the best GA run with Esso specific rules. On the whole, the parameter values for each of the rule sets are not entirely dissimilar. The largest difference is within the Esso rules, they have a larger *fixedCosts* and *changeInProfit* value than the other stations. This may be because the rules are being fitted to a small data set on one particular day. The *changeInProfit* value is also much higher than the other rules, this means that the stations are less likely to change price. The model doesn't run on absolute levels of profit, only the changes (relative levels), therefore the model may not be particularly sensitive to profit levels.

The Esso Price Watch would operate a 5km distance scale (*neighbourhood*) and 0.0p *overprice* value (to ensure it remains as cheap as its local neighbours). However, the results from the GA suggest that a better set of parameters to use (to maintain profit) would be a *neighbourhood* of 2.1km and an *overprice* of 4.5p. These parameters are most likely a reflection of the location of the Esso stations (normally within urban areas with a number of competitors surrounding them). However, as seen within the assessment of the profitability map in Chapter 8 (see §8.9), the Esso stations are located at sites of high profitability. It can therefore be concluded that the Esso Price Watch is highly effective in reality even if it is not optimal in the model. Very large values for



(a) Mean Price and SD

(b) Total SRMSE

(c) SRMSE by Type

Figure 9.16: Comparison of the (a) mean price difference and (b) total SRMSE over time and (c) SRMSE by station type between optimal results from GA runs with the same rules applied to all stations and from runs with separate rules for the Esso stations. In each case the model was initialised with the real data from July 27th. Standard deviations in the mean price are represented by vertical bars in (a).

overprice are not necessarily indicative of more expensive petrol. If the Esso stations are always competitive this rule may never come into operation in which case its value is irrelevant.

Esso are a much smaller category than the non-Esso's (49 stations compared to 214 stations where a price is recorded on July 27th) and so there tends to be more variability in the data as a change at one stations has a larger effect. This makes it more difficult to obtain optimal variables. In addition, examination of the price variation within the real data (Figure 9.17) shows that on a day to day basis, there is a considerable amount of variation in some of the Esso stations. Such frequent variations are not observed for other station types. These frequent variations, combined with the small number of Esso stations explain the large fluctuations in the SRMSE of the Esso stations seen in Figure 9.16(c). It is uncertain whether these fluctuations are the result of an aggressive price change policy (perhaps related to the Esso Price Watch) or whether they are a result of errors in the available data set.

	Same rules	Esso rules	
Parameter	Default	Default	Esso
β	0.0005283	0.0003246	
λ	0.4105	0.3044	
fixedCosts (£)	20.53	21.30	44.37
costToProduce (p)	65.80	64.68	66.73
changeInProfit (£)	1.00	1.00	3.92
undercut (p)	0.82	1.04	0.96
overprice (p)	4.61	4.47	4.52
neighbourhood (km)	3.59	2.95	2.13

Table 9.8: Comparison of the best GA run with all stations using the same rules and the best GA run with Esso specific rules.

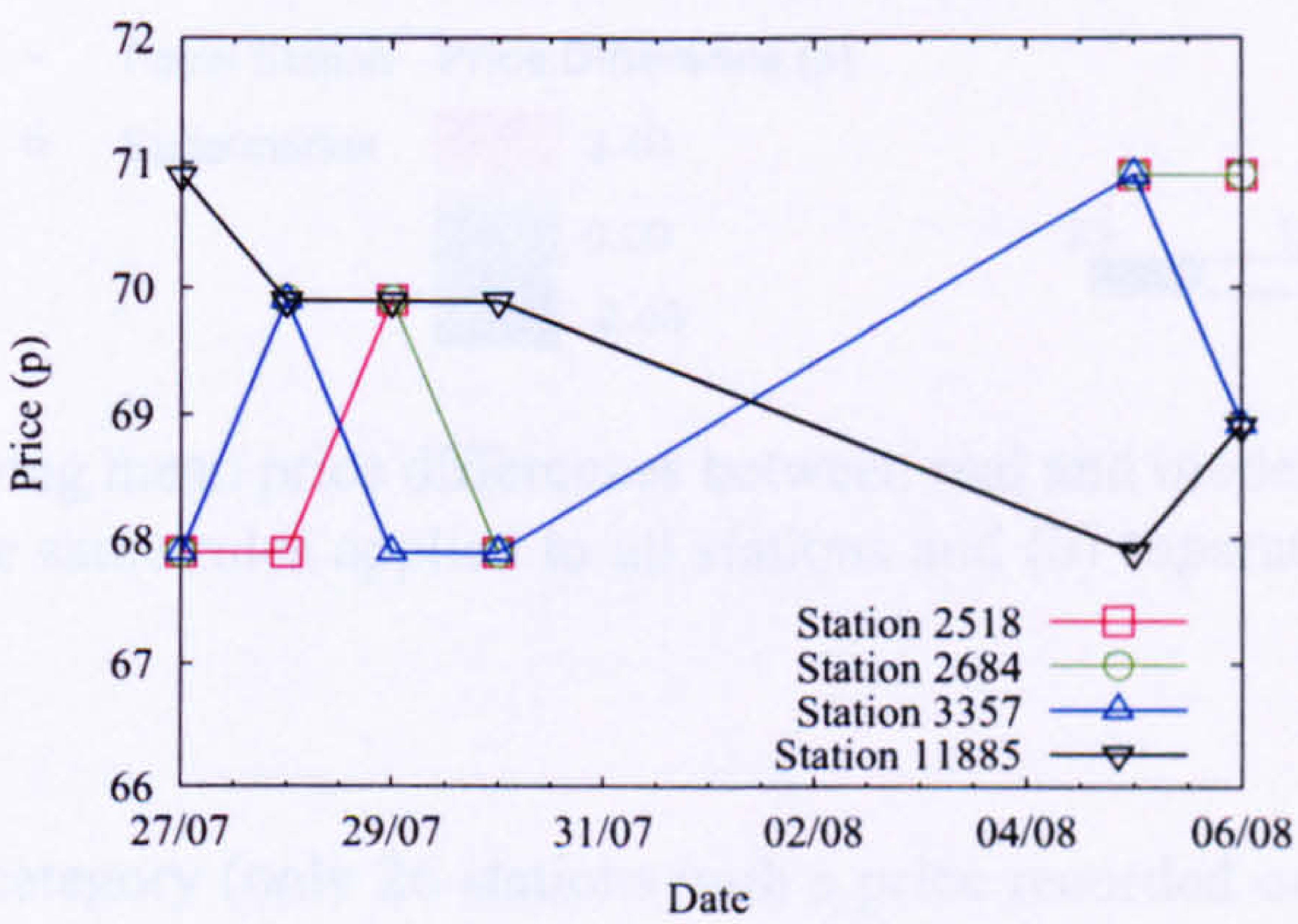


Figure 9.17: Price of unleaded petrol plotted against day for several Esso stations in West Yorkshire during the first half of the data set.

9.12.2 Supermarkets

In Chapter 8, a separate supermarket rule set was experimented with (see §8.6.2). Introducing this new behaviour did not improve the performance of the network hybrid model. As with the Esso experimentation, it was suggested that new parameters would have to be derived to account for this new behaviour.

The procedure used in §9.12.1 is repeated here using supermarkets rather than Esso stations. The model was initialised with data from July 27th for West Yorkshire. The supermarket stations were assigned their own rule set with the rest of the stations assigned default rules. The GA was run 10 times and the best simulation used within this section. Appendix E.2 shows graphs of the optimal model parameters from the GA against the SRMSE for all 10 of the GA runs.

Figure 9.18 demonstrates that adding supermarket specific rules to the GA improves the overall results. This can be seen in particular around the Huddersfield and Wakefield areas. However, an area of high price differences still exists towards the east of the study area.

Figure 9.19(a) shows that, as in the Esso run, the difference in the overall mean price between the two GA runs is very small. This is not totally surprising as the supermarket category is even

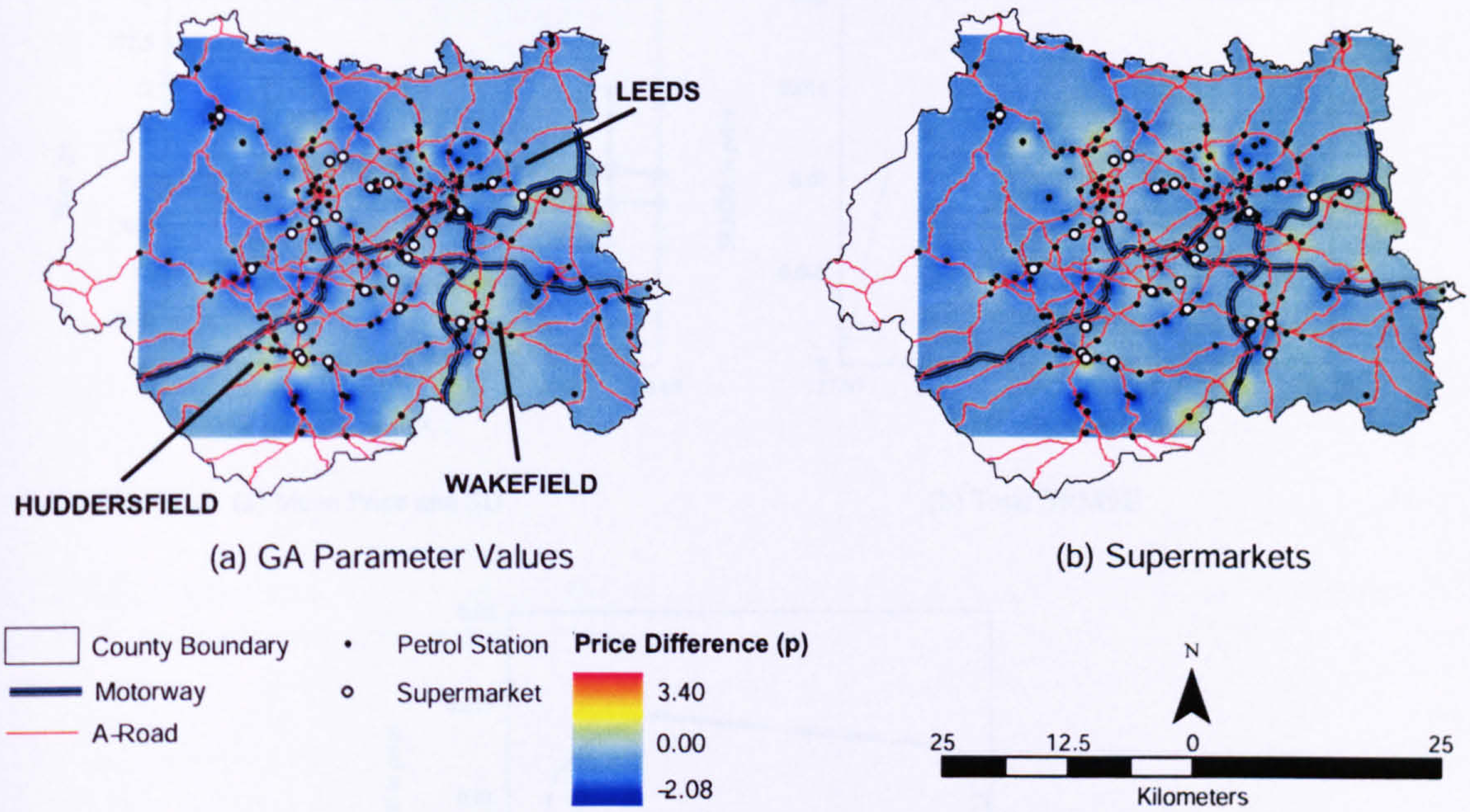
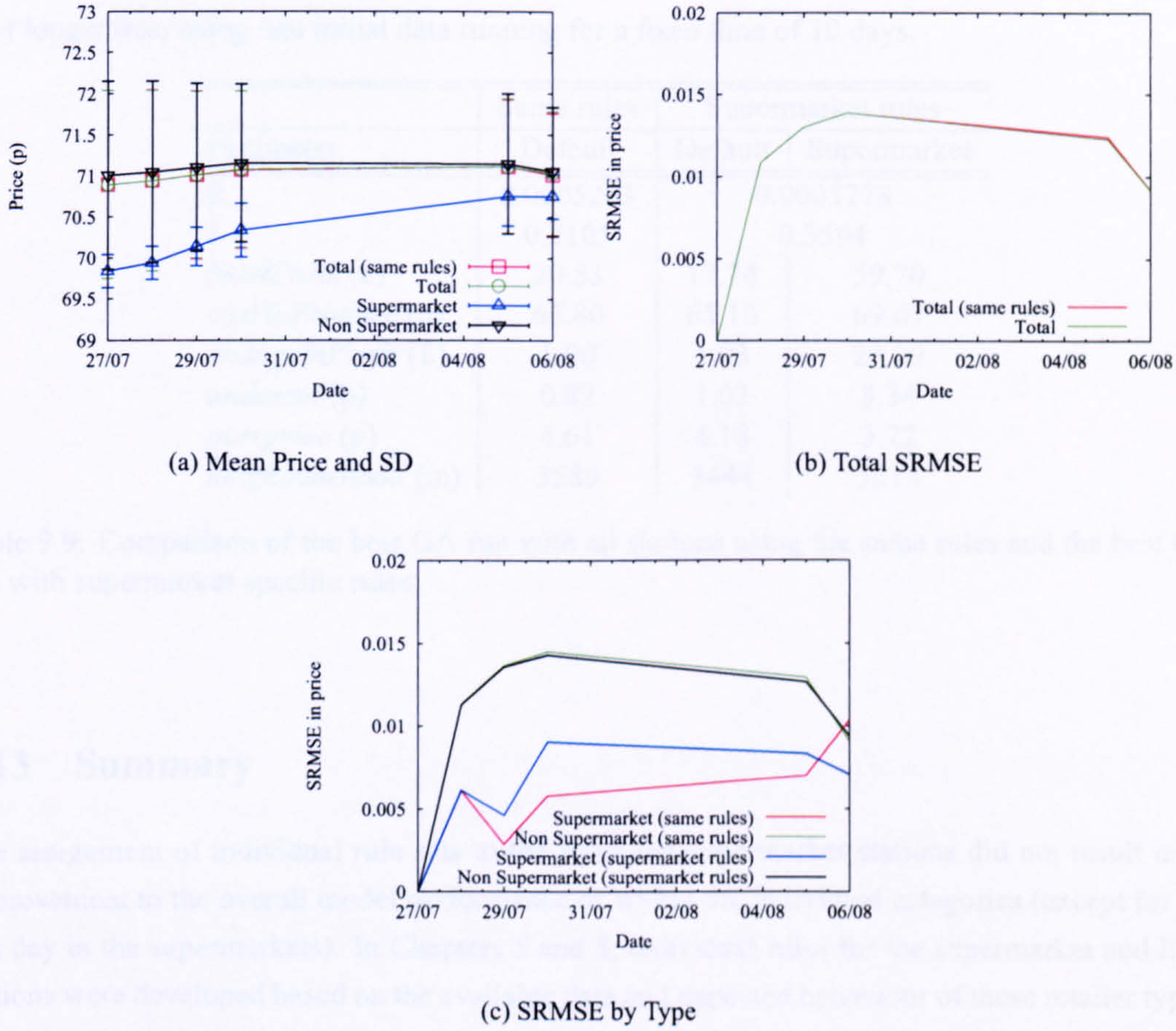


Figure 9.18: Map showing mean price differences between real and model data on August 6th for model runs with (a) the same rules applied to all stations and (b) separate rules for supermarket stations.

smaller than the Esso category (only 26 stations with a price recorded on July 27th). The mean price and SD for the non-supermarket stations is also very similar to the totals. Since there are very few supermarkets, the non-supermarket category and the total category are almost the same. The supermarkets have a much smaller mean price than the non-supermarket stations. This trend was noted in Chapter 4 (see §4.6.2). The difference in mean price decreases over time (also observed in the real data in §4.6.2). The range of prices are also much smaller than the other categories, indicating that supermarkets are operating in a more competitive manner than the other stations. This reflects the findings in Chapter 4 and information from the literature (see §2.3.4).

The SRMSE of the totals are almost identical (Figure 9.19(b)). This shows that assigning the supermarket category a separate set of rules does not significantly improve the overall performance of the model. The supermarket stations only form a small number of the total stations, even if the performance of these stations are improved, they will not have a great effect on the overall result. This is evident within Figure 9.19(c). There is almost no difference in the SRMSE between the rules sets for the non-supermarket stations. Both show a marked improvement in the model results on August 6th. With both models, the supermarkets perform better than the non-supermarkets. For most of the period the results obtained using the same rules for all stations show a better performance than using a specific rule set for the the supermarkets (as indicated by the blue and red lines). The situation reverses on August 6th however. The parameter fitness is solely being assessed on August 6th, the GA is tuning the model to perform well on this day, hence the lower SRMSE on the final day and the excellent visual performance in Figure 9.18.

Table 9.9 presents a comparison of the optimal parameter values obtained from the GA using



(a) Mean Price and SD

(b) Total SRMSE

(c) SRMSE by Type

Figure 9.19: Comparison of the (a) mean price difference and (b) total SRMSE over time and (c) SRMSE by station type between optimal results from GA runs with the same rules applied to all stations and from runs with separate rules for the supermarket stations. In each case the model was initialised with the real data from July 27th. Standard deviations in the mean price are represented by vertical bars in (a).

one set of rules for all stations and using separate rules for non-supermarket and supermarket stations. The parameter values for the GA (same rules) and the non-supermarket rules have similar values. However, the supermarket rules are considerably different for each parameter value. The *undercut* value for the supermarket is much higher than the other rules and the *overprice* is a little lower. This reflects the strongly competitive nature of the supermarkets as would be expected. The other values are counter-intuitive, the *fixedCosts* and *costToProduce* would be expected to be much lower however they are higher than for the non-supermarket stations. This is likely to be for similar reasons to those given for the Esso stations above.

The *changeInProfit* value is very high for the supermarket stations. This means that it is unlikely that the model will change its price much from its initial price regardless of the values of the other parameters. This is one of the problems associated with starting the model on the first day. An alternative approach would be to initialise the stations with the same price and let them evolve to an equilibrium solution, then compare with the data for a given day. Since a simulation make take 100 days or more to reach equilibrium (or indeed never reach equilibrium) this will take a lot longer than using real initial data running for a fixed time of 10 days.

	Same rules	Supermarket rules	
Parameter	Default	Default	Supermarket
β	0.0005283	0.0005778	
λ	0.4105	0.3594	
fixedCosts (£)	20.53	17.74	59.70
costToProduce (p)	65.80	65.10	69.67
changeInProfit (£)	1.00	1.03	23.29
undercut (p)	0.82	1.02	3.34
overprice (p)	4.61	4.18	3.72
neighbourhood (m)	3589	3444	5214

Table 9.9: Comparison of the best GA run with all stations using the same rules and the best GA run with supermarket specific rules.

9.13 Summary

The assignment of individual rule sets to the Esso and supermarket stations did not result in an improvement to the overall model performance or within the individual categories (except for the last day in the supermarkets). In Chapters 5 and 8, individual rules for the supermarket and Esso stations were developed based on the available data and expected behaviour of these retailer types. However, these derived rules are markedly different to those generated by the GA. Unlike the rules developed for Esso stations in §8.6.1 the Esso rules developed by the GA did not reduce the performance of the model.

The GA is ignorant of any specific corporate policies, all it takes into account is the known geographical distribution of these stations and the recorded price at each station. The fact that the Esso and the supermarket rules are different is a reflection of these geographical differences.

The use of different rule set has generated some unexpectedly large values for the *fixedCosts* and *costToProduce* parameters in the Esso and supermarket categories. This may be a result of the model working on changes in profit rather than absolute profit levels. Alteration of the pricing strategies to incorporate an element of both may produce more realistic results for these parameters.

9.14 Further Assessment of GA Performance

The GA has to date been tested using the data for West Yorkshire from July 27th. Using the second part of the data (August 19th - September 1) to test the GA will provide a test of the robustness of the parameter performance under different conditions. In §9.14.2, the ability of the GA parameters to recreate the spatial patterns evident within the real data will be tested on the larger geographical area of the Yorkshire region.

9.14.1 Performance on August 19th

The network hybrid model was initialised using data from August 19th. The model was run for 13 days and compared with the real data from September 1st. The optimal parameters from the GA derived for July 27th (referred to as GA parameters (early)) and the parameters from Chapter 8 (Chapter 8 parameters) were used in addition to the best parameters from a series of GA runs initialised using the August 19th data (GA parameters (late)). Table 9.10 presents each set of optimal parameters.

Parameter	Chapter 8	GA (early)	GA (late)
β	0.00003	0.0005283	0.0006303
λ	0.7	0.4105	0.0840
fixedCosts (£)	50.00	20.53	46.90
costToProduce (p)	65.0	65.80	63.71
changeInProfit (£)	40.00	1.00	36.50
undercut (p)	1.0	0.82	0.22
overprice (p)	1.0	4.61	2.85
neighbourhood (km)	5.00	3.59	8.95

Table 9.10: Comparison of the model parameters from Chapter 8 with the best parameters obtained from the GA using the early and late parts of the data set.

Appendix E.3 shows graphs of the optimal model parameters from the GA against the SRMSE for all 10 of the GA runs.

Figure 9.20 shows that the best performance comes from the GA parameters (late). The GA parameters (late) are slightly over-predicting the prices, but on the whole, the prices do not vary more than 0.2p (Figure 9.20(d)). The exception to this is in the area around Halifax where the model over-predicts the prices by approximately 2p. In Chapter 8 (see §8.8.1), experimentation using the network hybrid model revealed a price war in this area. The parameters from the GA late shows this price war, but not to the same extent as either the Chapter 8 parameters (Figure 9.20(b)) or the GA parameters (early) shown in Figure 9.20(c).

Comparing the performance of the models statistically (Figure 9.21(a)) shows that the mean price for the Chapter 8 parameters is in reasonable agreement with the real data over the study period. However, the price range is much larger than in the real data. The GA parameters (early) start off similar to the real data but from the 23rd August show a downward spiral in price as a result of the price war in both Leeds and Halifax. Running the GA using the August 19th data to initialise it gives a much better performance with the mean price being only slightly less than the real data throughout the simulation. The range of values is also smaller than the real data as well as being smaller than the other simulations.

The price war and the large range of prices in the Chapter 8 and GA parameters (early) results are reflected in the SRMSE values which become progressively larger with time. In contrast, the GA parameter (late) results exhibit a much lower SRMSE and are a definite improvement on the results using a constant price throughout the simulation.

Unfortunately, because of the differences between the early and late data sets, the rules derived for one do not necessarily perform well on the other. This is perhaps not surprising since the

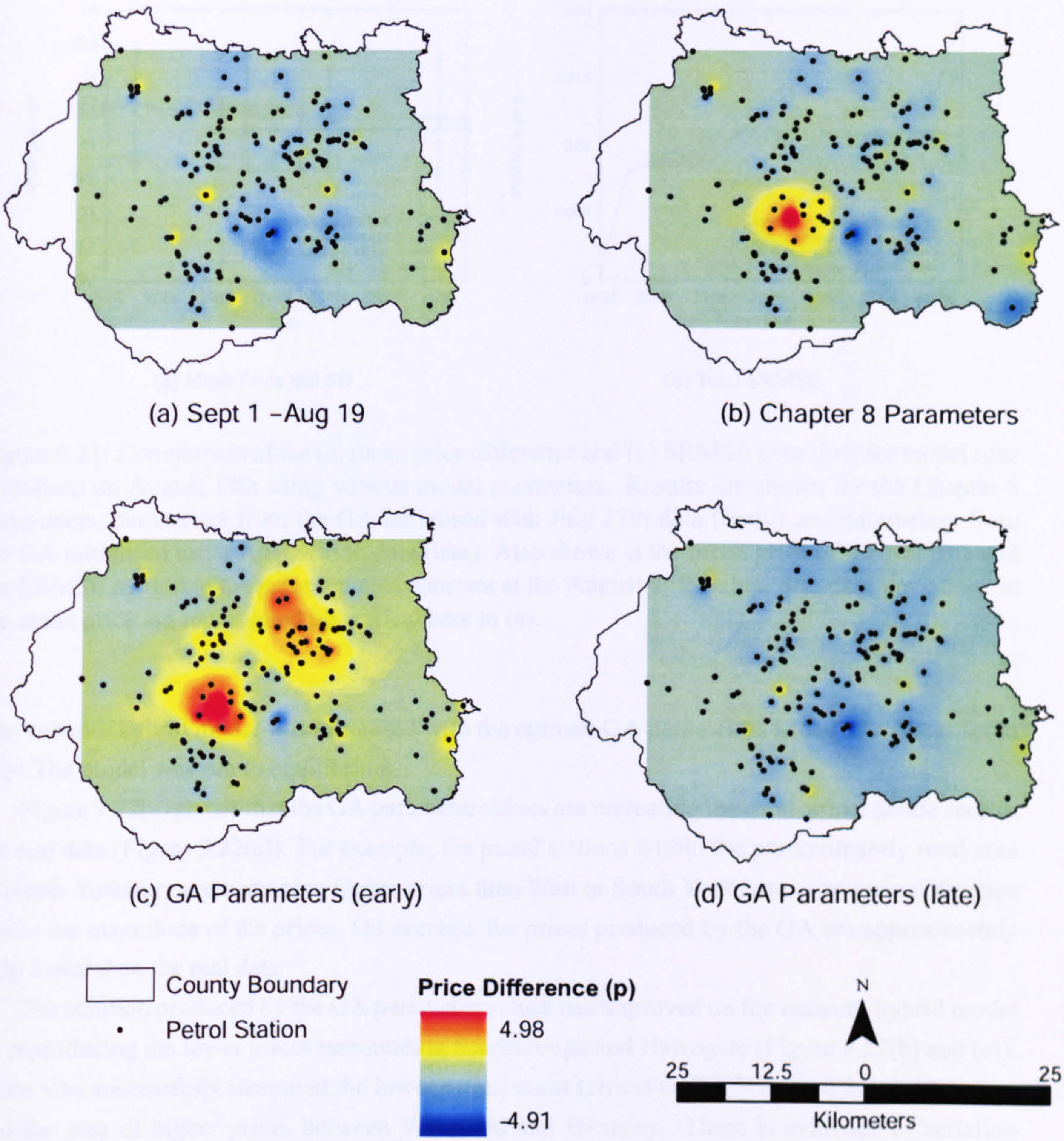
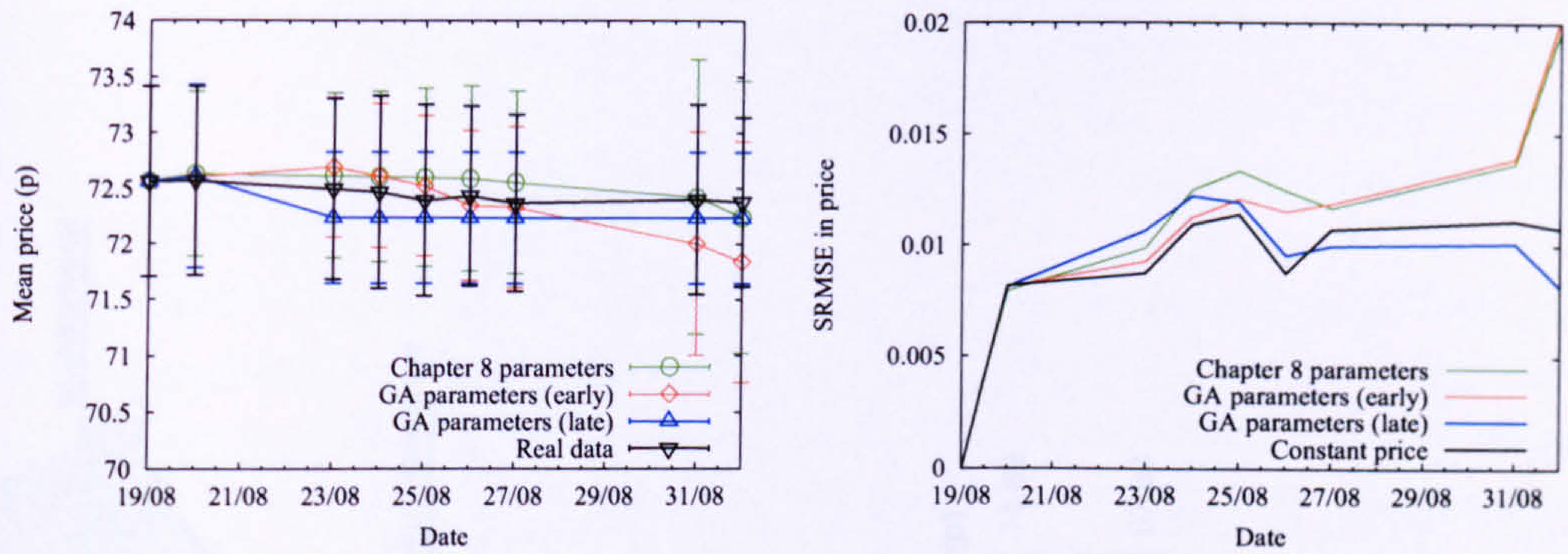


Figure 9.20: Map showing mean price differences between real and model data on September 1 for model runs with (a) the real data (b) the Chapter 8 parameters, (c) the GA rules from the early data and (d) the GA rules from the late data.

analysis in Chapter 4 showed a relatively steep rise in price between the two periods suggesting a change in the costs due to an increase in crude oil prices.

9.14.2 Performance on Yorkshire Region

In Chapter 8, the ability of the network hybrid model to recreate spatial variations was tested by initialising the model using constant price data (see 8.3.2). To test whether the parameter values produced by the GA can generate trends occurring in the real data, this experiment was repeated.



(a) Mean Price and SD

(b) Total SRMSE

Figure 9.21: Comparison of the (a) mean price difference and (b) SRMSE over time for model runs initialised on August 19th using various model parameters. Results are shown for the Chapter 8 parameters, parameters from the GA initialised with July 27th data (early) and parameters from the GA initialised using August 19th data (late). Also shown is the mean price of the real data and the SRMSE assuming the price remains constant at the August 19th value. Standard deviations in the mean price are represented by vertical bars in (a).

The network hybrid model was initialised with the optimal GA parameters and all the prices set to 71p. The model was run to equilibrium.

Figure 9.22(c) shows that the GA parameter values are recreating the rural-urban divide seen in the real data (Figure 9.22(a)). For example, the petrol stations within the predominantly rural area of North Yorkshire are sustaining higher prices than West or South Yorkshire. The main difference lies in the magnitude of the prices. On average, the prices produced by the GA are approximately 1-2p lower than the real data.

The solution produced by the GA parameter values has improved on the network hybrid model by reproducing the lower prices surrounding Scarborough and Harrogate (Figure 9.22(b) and (c)). It has also successfully identified the lower priced areas (cities) within West and South Yorkshire and the area of higher prices between Wakefield and Barnsley. There is evidence of variation within urban areas, although this is not as prominent as within the real data. Based on assessment of Figure 9.22, the GA solution has improved on that produced by the Chapter 8 values.

9.15 Conclusion

The work within this chapter has sought to objectively select optimal parameter values for running the network hybrid model. A Genetic Algorithm (GA) was used due to its ability to optimise multi-dimensional problems with multiple local minima. The chapter provides an overview of this technique and the development of a GA to couple with the hybrid and network hybrid models presented in Chapters 6-8. Experiments varying the parameters controlling the GA were conducted for the hybrid model and data from July 27th leading to the selection of a reasonable set of parameters. Multiple runs of each experiment were undertaken to attempt to find the global minimum.

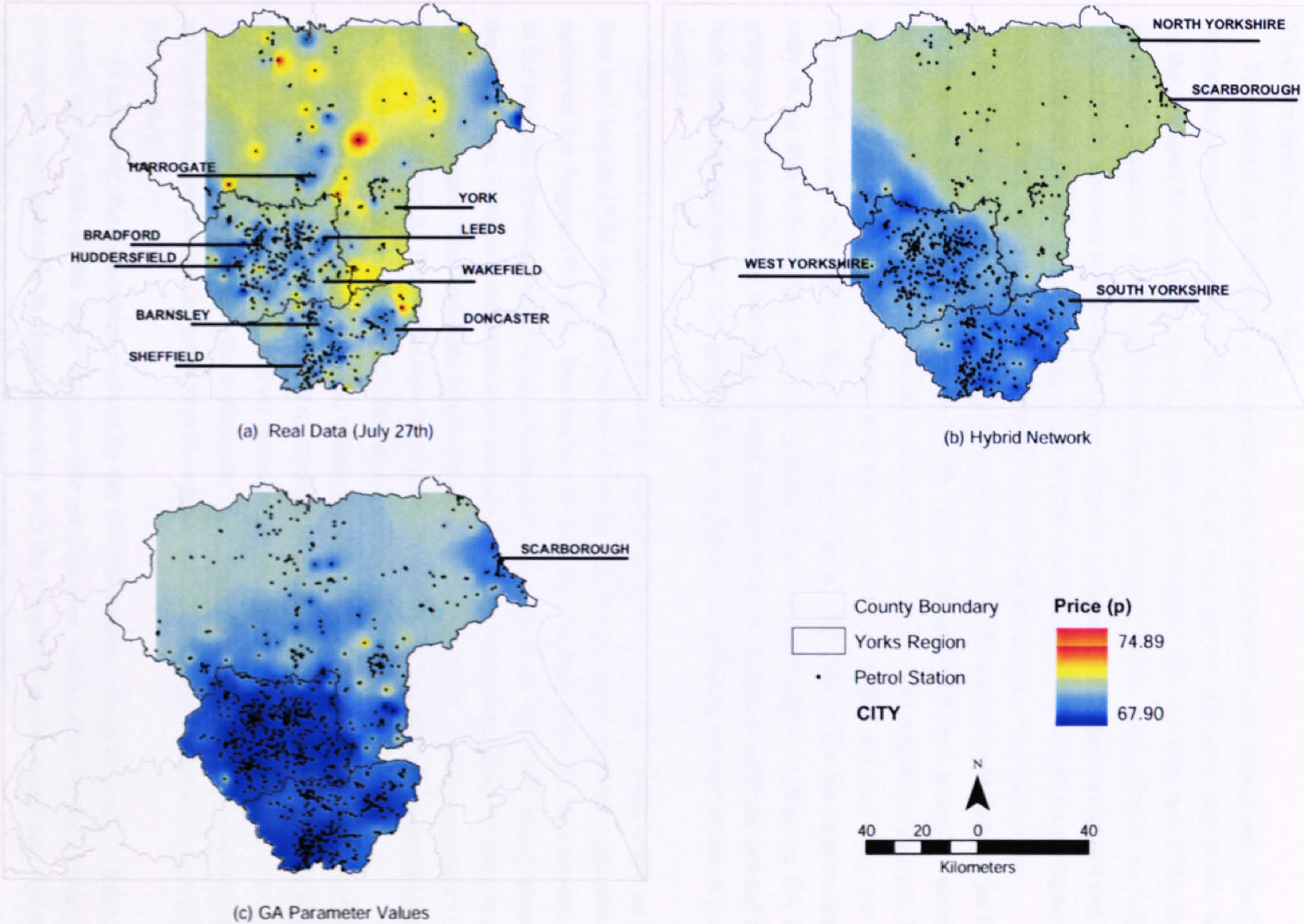


Figure 9.22: Comparison of the model results for Yorkshire using (b) the Chapter 8 parameters and (c) the optimal parameters from the GA using data from West Yorkshire. In each run the stations were initialised with a constant price of 71p. Also shown (a) is the real data for July 27th.

Examination of these runs revealed that, in terms of parameter space the best runs clustered together. This provided an indication of the global minima and some evidence for the robustness of the GA approach. An optimal set of parameters from the GA were obtained for the network hybrid model and comparisons undertaken with the real data and Chapter 8 parameters for West Yorkshire from the July 27th data.

The optimal GA and Chapter 8 parameters were found to be in close agreement. The Chapter 8 parameters overall produced a slightly better result although the difference was small. However, on the day used for comparing the data (August 6th for data initialised with July 27th), the results for the GA parameters were much better than the Chapter 8 values. This is due to the fact that the GA currently assesses model performance solely by comparing the results on the last day (August 6th). Despite the GA not performing quite as well statistically as the Chapter 8 parameters, it did outperform the Chapter 8 model in terms of reproducing the spatial trends in pricing.

Further experimentation concentrated on evolving specific optimal rule sets for the Esso and supermarket categories. The purpose of this was to determine whether assigning these groups individual rules would improve the performance of the model. The results were mixed, the Esso stations did not perform particularly well (due to variability within the real data), but the GA supermarket rules did improve the model performance, particularly for the supermarkets. The only factor that differentiates the Esso stations from the supermarkets within the GA are their geographical location and real prices. Interestingly, the GA suggested different optimal rules for both categories suggesting that geography has a significant influence on the system as previously thought.

The optimal GA parameters derived for West Yorkshire on July 27th were tested on the later data set (August 19th) and on the Yorkshire region. The GA parameters gave a reasonable performance on the August 19th data. Inconsistencies with the real data were due to increases in price in the real data (between July 27th and September 1st) and a small price war that occurred during the simulation. Optimal parameters were evolved using the August 19th data and these produced a better performance. The Yorkshire region was used to test the ability of the parameters to recreate spatial pricing trends. The results were very positive with the GA parameters reproducing more of the pricing variations within the real data than the Chapter 8 parameters.

One issue did come to prominence within the analysis. This involved the method of assessing the model fitness in the GA. Different methods of assessing fitness have their own advantages and disadvantages. Assessing the model on one day did, on occasions, show the model in a better light by over-tuning the parameters for a particular day. Using a combination of techniques (i.e. not just comparing on the last day) may provide a fairer test of model performance. This is a point for further work.

In summary, the parameters evolved by the GA did improve the performance of the network hybrid model while at the same time provide an objective justification for their values. These parameters were generally in close agreement with the Chapter 8 model suggesting that the arguments used there to derive suitable parameter values were justified.

Chapter 10

Conclusions and Recommendations for the Future

10.1 Introduction

The work within this thesis is novel. Currently there are no other published examples of agent-based models being applied to the retail petrol market. More generally, there are no examples of agent models being linked to spatial interaction models, network models and genetic algorithms (GAs) in one application. Using this hybrid agent approach has enabled the modelling of both the temporal and spatial aspects of a complex system whilst charting the effects of an individual's behaviour at different scales of analysis (for example, individual, city, regional or national scales). The application of this methodology has been very successful and has produced a very powerful tool that can not only be applied to the petrol market, but also to other geographical problems.

The work within this chapter concludes the thesis and fulfils the final research objective by summarising the research findings and highlighting the main discoveries made with reference to the research objectives stated in Chapter 1. A critique of the methodology is provided along with points for future research. The chapter ends by drawing attention to those areas where the applications of hybrid agent-based systems may potentially prove beneficial in the future.

10.2 Summary of the Research Findings

As originally stated in Chapter 1, the overall aim of the research was to examine the ways in which hybridised agent-based models can be applied to modelling a dynamic, locally interacting retail market (in this case, the petrol market). To complete this aim, several research objectives were designed. Each of these objectives will now be addressed with a discussion of the main research findings and the primary lessons learnt.

Objective 1: Review and discuss the current state of agent-modelling and the modelling of the petrol market highlighting potential areas for research.

The work within Chapters 2 and 3 provided much of the contextual background for this thesis. The retail petrol market was reviewed for two purposes; to gather information about trends and strategies within the petrol industry and to understand how previous researchers have modelled this system. Several key findings were made which shaped the development of the thesis. Perhaps the most significant finding was the lack of published information available on strategies that companies may operate (the exception to this was the Esso Price Watch). It was concluded that this is due to the commercial sensitivity of the system. This absence of information meant that in subsequent chapters other methods of selecting retailer behaviour had to be employed. The complexity of the petrol price market was also identified through this review. In particular it was noted that pump prices are influenced by several interdependent factors, the most important of which appear to be geographical location and local competition. These factors had to be considered when developing and validating the model. In addition, the review of previous modelling approaches revealed that these were empirically based (normally regression models examining the relationship between petrol price and one other variable), neglecting either spatial or temporal factors as well as being unable to model the impact of individual pricing schemes at any level.

The criticisms of these "traditional" modelling techniques led to the discussion of agent-based models as an alternative methodology in Chapter 3. Through the review of these systems and thus the identification of key characteristics such as flexibility, adaptability and autonomy, these systems were identified as being highly appropriate for application to a complex system such as represented by the petrol price market. The agent paradigm advocates decomposition of the problem into small units, it is therefore possible to provide each agent (or petrol station) individual rule sets (strategies) and examine the effects spatially, temporally and at different scales. It was surmised that this technique could overcome problems associated with traditional approaches. This provided the foundations for the development of the pure agent model in Chapter 5.

Objective 2: Analysis of the real market data to look for evidence of spatial and temporal variabilities and investigation of the suitability of empirical techniques to explain these variations.

One of the key findings in Chapter 2 was the absence of any information concerning strategies that petrol retailers may operate. In the absence of such information, a detailed analysis of the real data was carried out in Chapter 4. This involved examining price variations within different sized neighbourhoods, geographical areas and petrol station categories. This analysis was extremely valuable in providing an in-depth understanding of the system.

From this analysis, conclusions could be formed about some of the behaviour within the market. For example, supermarket stations were identified as being both cheaply and competitively priced, rural stations sustained higher prices than the other stations and over the duration of the data set there was a marked increase in pump prices due to a rise in crude oil prices.

Understanding these trends and patterns within the data enabled appropriate classifications and statistical tests to be employed for comparing prices at the same station over a period of time. These classifications and statistical tests were then used to evaluate the performance of the model throughout the thesis.

Within Chapter 4 empirical models (based on those reviewed in Chapter 2) were built and applied to the real data to ascertain whether, for this data and research aim, these techniques would be sufficient. The results for both were poor, further cementing the statements made in Chapter 2 about the inadequacies of these approaches for modelling a complex system that is both spatially and temporally dynamic.

Objective 3: Use agent technology to build a model to simulate the spatial and temporal variations in price observed in a single commodity retail market.

On the basis of the work presented in Chapters 3 and 4, the pure agent-based model was constructed in Chapter 5. This simple model shared some features with traditional modelling methods highlighted in Chapter 2. The agents (petrol stations) reacted to imposed prices rather than actively predicting them. Initial experimentation was undertaken with individual parameters and separate rule sets (for the Esso and supermarkets stations) assigned. The performance was sensible and the model operated successfully. However, after further analysis of the results it was concluded that the model suffered from an absence of realistic retailer and consumer behaviour: there was no competitiveness in the system and no attempt by retailers to maximise their profits.

These shortfalls led to two further developments within Chapter 6, the spatial interaction model and the network model. Spatial interaction models are a well developed technique for modelling flows of information or people. This technique integrates well with the agent framework and combines the benefits of both traditional and artificial intelligence (AI) techniques. The spatial interaction model provided sales of fuel at the petrol stations based on distance to consumers and price. The agents representing each station could then use this information to calculate their profit and hence determine a suitable pricing strategy. The construction of the network model was a refinement to the model framework to account for the fact that petrol is often not bought on a special trip, but during the course of another journey such as travelling to work. Ideally, the data used would contain detailed information on all journeys (including on which journey petrol is purchased) and journey times. However the data available are based on a 10% journey to work sample, which meant that generalisations had to be made.

Objective 4: Assess model responses to different configurations, initial conditions and rule sets using both idealised and real data.

An important aspect of any model is the suitability of the parameters and their values. Of equal importance is an understanding of the model behaviour and its reaction to different rules and initialisation conditions. This is easier to obtain in an idealised environment where the complications of heterogeneous populations and geography have been removed. In common with many other complex systems, the model responses are unlikely to depend linearly on the input parameters and initial conditions. Methods of validation used must take this into account.

The first part of this validation involved assessing whether the hybrid model[1] performed sensibly under simple idealised cases. This was a necessary prerequisite before application to the more complex, real data. Within Chapter 7, sensible values for the parameters were derived and experiments run that showed the model successfully reproducing rural-urban divides for ideal cities and reacting appropriately in diffusion experiments. Sensitivity testing was also undertaken which

[1] Due to the standardised geography and population, the network part of the model was redundant in Chapter 7.

demonstrated the robustness of the system. There was some evidence of cyclic and chaotic behaviour, however the parameter values chosen for further use were not in the range which produced this behaviour.

After successful testing on the idealised data, the model was then applied to the more complicated real system in Chapter 8. Similar experiments to those performed in Chapter 7 were run. Experiments initialised with the data from July 27th and run for 10 days were in good agreement with observed prices on August 6th. Further validation of the model was obtained by using a larger geographical area and different temporal periods to test the performance of the model parameters. The results showed that the model successfully reproduced patterns found within the real data. As in the idealised simulations, the derived parameter values came from stable parts of the parameter space.

The "rockets and feathers" concept was first introduced within Chapter 2. This effect was simulated in Chapter 8 using rapid changes in the *costToProduce* parameter. As observed in the real petrol market, an increase in costs led to a rapid rise in price with a slow decrease in price when the costs fall. This showed that the model can be used for reproducing temporal patterns in the data due to changes in the parameters as well as for generating spatial patterns.

Most of the validation used in this thesis has involved direct comparisons of real and predicted prices at each petrol station. An alternative approach was also used in Chapter 8. The model was used to predict the profitability of stations in the Leeds areas using the 1999 data set. A comparison with data from 2004 (whether these stations were still operating) was then undertaken. The results showed a high correlation between stations predicted as having a low profitability in 1999 and those that have subsequently closed. This illustrates the potential of the technique for use in long term planning decisions.

Objective 5: Investigation of objective techniques to select optimal values for the parameters in the agent model.

One of the issues that arose within this thesis concerned selecting parameters objectively for the model. In Chapters 5, 7 and 8, parameters for the models were selected by means of a combination of real data analysis and numerical experimentation. The system had already been identified as being non-linear, therefore experimentation with one parameter at a time may not result in an optimal set of parameters. Initial investigations showed the system to be complex with multiple local minima suggesting that there are many different solutions available for the problem. Genetic algorithms (GAs) were chosen due to their usefulness for multidimensional optimisation problems in which the chromosome (set of rules) can encode the values for the different variables being optimised.

The optimal parameters obtained by the GA were quite close to those derived from the real data. This provides support for the arguments used to independently derive the parameters in previous chapters and suggests that the model is behaving in a realistic manner. The robustness of the GA parameters was tested by using different geographical and temporal data. The GA suggested some differences between optimal parameters for the first and second halves of the data set. This could be explained through differences in the underlying conditions such as a rise in the price of crude oil during the period.

10.3 Critique of Methodology

This thesis has examined the application of hybrid agent-based models to the retail petrol market. This has proved to be a successful modelling strategy. The strengths of the agent framework have been successfully combined with embedded market behaviour represented by traditional methods. In contrast to empirical approaches to modelling the retail market, the model developed within this thesis has a sound theoretical grounding based on the idea of competition between petrol stations.

One potential weakness of this model is that the agent approach has not been extended to the consumer. Consumers are modelled in an aggregate manner by the spatial interaction model. This approach can be justified when considering the detail of the available data. To implement a more sophisticated agent approach, data documenting specific purchasing habits would be required as well as access to powerful computing resources. An attempt was made to refine the data by redistributing the population on the basis of journey to work data. This refinement did have the effect of improving the results.

An issue that arises with modelling consumer behaviour is how to deal with consumer knowledge. Chapter 2 made the case that consumers can have perfect knowledge of petrol prices within their area as a result of internet sites such as the AA (accessed 2004). However, in practice, it is expected that consumers will only know the prices of those stations along routes that they frequently take. How is this imperfect knowledge to be modelled? Perhaps this is an opening for fuzzy logic and probabilistic models? This is certainly an area for further development.

The pricing strategy implemented was the same for each petrol station. This was based on the assumption that all petrol stations are individually aiming to make as much profit as possible. This is a reasonable assumption, even if a retailer is operating a company policy of some description, it is unlikely to allow a station to run at a loss. The model captured this in the profitability experiment carried out in §8.9. However, there may be other strategies at work, for example, collusion for market share, rate of profit maximisation as examined by Sheppard *et al.* (1998) or supermarket loyalty schemes which offer discounts on the pump price (cross-subsidising). As already alluded to, inclusion of this behaviour would require more detailed data.

A similar argument to that presented above can be applied to the use of rule sets. With the exception of the Esso and supermarket groups, the same rule set was used by each petrol station. This assumption is plausible as the rule set enables the station to be competitive in order to maximise its profit. If specific information on company strategy was available, assigning stations a combination of rules could potentially improve the model performance.

A combination of standardised root mean square error (SRMSE), mean and standard deviation (SD) was used to assess the fitness of the model solutions in Chapters 5, 7, 8 and 9. These techniques were chosen because they are simple, but offer the advantages of producing an easy to understand measure of the overall model performance. Additionally, Knudsen and Fotheringham (1986) cited SRMSE as the most reliable technique for point to point analysis after a review of several different approaches. However, the downside to using these techniques is that they are unable to indicate whether particular geographical areas or brands are being modelled accurately. This limitation was overcome by the use of detailed geographical classifications and interpolated maps.

In Chapter 9, it was apparent that calculating the SRMSE on the final day of the simulation and comparing with the real data may have drawbacks. The results on the final day of several of the simulations indicated that the performance was good. However, this was in contrast to the results from the preceding days. An alternative approach that could rectify this situation is the calculation of the SRMSE on each day with the average value over the simulation taken as a measure of model fitness.

10.4 Future Research Developments

Having summarised the research achievements and limitations in the research, this section covers some areas for potential future investigation. Further developments of the model for application to the retail petrol market as well as to other geographical problems are discussed.

An area ripe for further development is the construction and implementation of an agent-based consumer model. As highlighted in the previous section, this approach would require both detailed data and vast computational resources. In the absence of such information, the aggregate approach of the spatial interaction model could be improved by the use of regional data on car ownership and purchasing patterns to provide more detailed information on petrol sales at the ward level.

Another area that has the potential to be developed is the method by which pricing strategies are devised. Currently these strategies are fixed for all stations throughout the course of the simulation. The GA could be developed to change strategies "on the fly". This would enable optimal strategies to be continuously used and would produce a dynamic, intelligent pricing function. This could be further improved in the long-term by the use of artificial neural networks (ANNs) to enable the agents to learn which strategies are successful or genetic programming (GP) to evolve entire rule sets.

The experimentation within this thesis has been limited to using two study areas, West Yorkshire and the Yorkshire region. Within these two areas, the costs associated with transporting petrol and maintaining a site could reasonably be expected to be the same. One of the strengths of using the agent methodology is the ability to examine the impact of the model over different scales. Extending the experimentation to using larger regions such as England or Scotland may reveal larger scale variations that have not been fully accounted for. For example, transportation costs would be much higher in the Scottish Highlands than in Leeds. These extensions to the experimentation would require these costs, along with other important factors such as crude oil prices, to be dynamically fed into the model. This would further test the model's ability to cope with different spatial and temporal variations.

As well as using different geographical areas, the experimentation could also be extended to using different temporal periods. This would require data covering a period greater than 3 months. However, it would allow the ability of the model to make long term predictions to be rigourously investigated. In addition, data covering a longer period could be analysed for evidence of different company pricing strategies. The use of evolving rule sets, discussed above, might be useful for modelling prices over periods of months or years rather than just days. Over these long time periods there may also be evidence of homeostasis (the ability of a system to regulate itself and maintain an equilibrium) in the system. This is an interesting avenue of research for future investigation.

The framework developed within this thesis has the potential to be extended to other geographical applications. To do this would require the production of a generic version of the model. Ideally, this would allow easy customisation for a particular application, for example allowing extensions such as the spatial interaction model to be bolted onto the basic agent model. Technologies such as XML parsing could be used to make the input of model structures and data simpler.

10.5 Applications of Hybrid Agent Modelling

The work within this thesis has shown that the hybrid agent model is able to replicate the main trends and patterns observed in the retail petrol market. This success combined with the flexibility of the model suggests the potential for use in various applications. For example, the model could be used for planning decisions such as locating the best site for a new petrol station or deciding whether to close a station because it is no longer commercially viable. It could also be used by companies as a strategy tool for experimenting with different marketing strategies, for example the impact of implementing their own version of the Price Watch. The ability of the model to predict long-term patterns (as seen by the results of the profitability experiment in Chapter 8) as well as short term predictions makes this a powerful tool.

This approach is not limited to the petrol price market. The agent methodology is very appropriate for human geography applications where people and companies are making autonomous decisions based on interactions between the various participants in an area. There are many other geographical systems that possess the characteristics of being both spatially and temporally dynamic and are currently being modelled using empirical approaches. The approaches taken within this thesis could be readily applied to such diverse problems as modelling sub-glacial water pressure, epidemiology (spread of diseases), modelling the housing market or modelling crime.

10.6 Concluding Remarks

The aim of this thesis has been to examine the application of a hybrid agent-based model to the retail petrol market. Agent-based systems provide an ideal framework in which a retail market can be accurately described. Individual agents (petrol stations) can be successfully supplied with detailed knowledge of the market by the attachment of more specialised models, such as spatial interaction models. This type of model can be extremely valuable when modelling trends at a regional level. The outcome of local interactions can be easily seen at a global level and behavioural and quantitative data can be easily combined. This flexibility, coupled with the upsurge in readily available computing power makes agent based modelling a valuable tool for studying market forces and dynamics.

This type of modelling is not just limited to application to the retail market. The range of applications that agent-based models are being applied to is ever increasing. It is to be hoped that, with further development, the ideas and framework developed within this thesis will be used to further our understanding of complex geographical systems.

References

AA (accessed 2004). AA Petrol Busters. URL `http://www.aapetrolbusters.com/`.

Allvine, F. C. and Patterson, J. M. (1972). *Competition Ltd: The Marketing of Gasoline*. Indiana University Press, Bloomington, IN, 1st edn.

Alvarado, M., Cheremetov, L. and Cantu, F. (2004). Autonomous Agents and Computational Intelligence: The Future of AI Applications for the Petroleum Industry. *Expert Systems with Applications*.

Anumba, C. J., Ugwu, O. O., Newnham, L. and Thorpe, A. (2002). Collaborative Design of Structures using Intelligent Agents. *Automaton in Construction*, **11**:89 – 103.

Axelrod, R. and Scott Bennett, D. (1997). Choosing Sides: A Landscape Theory of Aggregation. In Axelrod, R., ed., *The Complexity of Cooperation*, pages 72–94. Princeton University Press.

Bacon, R. W. (1991). Rockets and Feathers: the Asymmetrical Speed of Adjustment of UK Retail Gasoline Prices to Cost Changes. *Energy Econ*, **1**:211 – 218.

Barros, J. and Sobreira, F. (2002). City of Slums: Self-Organisation Across Scales. CASA Working Paper 55.

Batty, M., Desyllas, J. and Duxbury, E. (2003). The Discrete Dynamics of Small-Scale Spatial Events: Agent-based Models of Mobility in Carnivals and Street Parades. *Int. J. Geographical Information Science*, **17**(7):673–697.

Batty, M. and Jiang, B. (1999). Multi-Agent Simulation: New Approaches to Exploring Space-Time Dynamics within GIS. CASA Working Paper 10.

Beasley, D., Bull, R. D. and Martin, R. R. (1993). An Overview of Genetic Algorithms Part 1: Fundamentals. *University Computing*, **15**:58–69.

Beltratti, A., Margarita, S. and Terna, P. (1997). *Neural Networks for Economic and Financial Modelling*. Thomson Computer Press, 1st edn.

Ben-Akiva, M. (1985). *Discrete Choice Analysis: Theory and Application to Predict Travel Demand*. MIT Press, Cambridge MA, 1st edn.

Beneson, I., Omer, I. and Hatna, E. (2001). Entity-Based Modelling of Urban Residential Dynamics: The Case of Yaffo, Tel Aviv. *Environment and Planning B*, **29**:491–512.

Birkin, M., Clarke, G. and Clarke, M. P. (2002). *Retail Geography and Intelligent Network Planning*. Wiley, 1st edn.

Booch, G., ed. (1994). *Object-Orientated Analysis and Design with Applications*. Addison-Wesley, Reading, MA, 1st edn.

Borenstein, S. and Cameron, A. C. (1997). Do Gasoline Prices Respond Asymmetrically to Crude Oil Price Changes? *Quarterly Journal of Economics*, 112(1):305 – 339.

Borenstein, S. and Shepard, A. (1993). Dynamic Pricing in Retail Gasoline Markets. *NBER Working Paper*, 4489.

Brannon, J. I. (2003). The Effects of Resale Price Maintenance Laws on Petrol Prices and Station Attrition: Empirical Evidence from Wisconsin. *Applied Economics*, 35:343 – 349.

Brassel, K.-H. (accessed 2003). VSEit. URL `http://www.vseit./de`.

Braunschweig, B. and Gani, R., eds. (2002). *Software Architectures and Tools for Computer Aided Process Engineering*. Elsevier, 1st edn.

Brekke, K. R. and Straume, O. R. (2004). Bilateral Monopolies and Location Choice. *Regional Science and Urban Economics*.

Brenner, W., Zarnekow, R. and Wittig, H. (1998). *Intelligent Software Agents: Foundations and Applications*. Springer-Verlag, 1st edn.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1998a). Geographically Weighted Regression - Modelling Spatial Non-Stationarity. *Journal of the Royal Statistical Society - Series D - The Statistician*, 47(3):431–443.

Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1998b). Spatial Non-Stationarity and Autoregressive Models. *Environment and Planning A*, 30(6):957–993.

Burse, J. (accessed 2003). QuickSilver. URL `http://sourceforge.net/projects/java4u/`.

Camacho-Cuena, E., Garcia-Gallego, A., Georgantzis, N. and Sabater-Grande, G. (2004). Buyer-Seller Interaction in Experimental Spatial Markets. *Regional Science and Urban Economics*.

Campbell, W. J. and Chisholm, M. (1970). Local Variations in Retail Grocery Prices. *Urban Studies*, 7:76–81.

CASA (accessed 2004). Centre for Applied Spatial Analysis. URL `http://www.casa.ucl.ac.uk/`.

Castanias, R. and Johnson, H. (1993). Gas Wars: Retail Gasoline Price Fluctuations. *The Review of Economics and Statistics*, 75:171 – 174.

Castelfranchi, C. (2001). Again on Agents' Autonomy: A Homage to Alan Turing - Panel Chair's Statement. *Intelligent Agents VII*, LNAI 1986:339 – 342.

Catalist (accessed 2003). Catalist web site. URL `http://www.catalist.com/`.

Cavezzali, A. and Rabino, G. (2003). Multi-Agent Systems and Territory: Concepts, Methods and Applications. In *European Regional Science Association.*

Chan, T. Y., Padmanabhan, V. and Seetharaman, P. B. (2003). A Structural Model of Locational Competition Among Gasoline Retailers: An Empirical Analysis. URL `http://www.kellogg.northwestern.edu/marketing/faculty/marketing_camp_2003/pb_seetharaman-locational_competition.pdf`. Accessed July 2004.

Choi, H. R., Kim, H. S. and Park, B. J. (2004). Multi-Agent Based Intergration Scheduling System Under Supply Chain Management Environment. *Lecture Notes in Computer Science*, **3029**:249–263.

Claycamp, H. (1966). Dynamic Effect of Short Duration Price Differentials on Retail Gasoline Sales. *Journal of Marketing Research*, **3**:175–178.

Clemenz, G. and Gugler, K. (2002). Locational Choice and Price Competition: Empirical Results for the Austrian Retail Gasoline Market. *Working Paper.*

Competition Commission (1990). Petrol Pricing. URL `http://www.competition-commission.org.uk/rep_pub/reports/1990/fulltext/265c4.pdf`.

Crabtree, B. (1998). What Chance for Software Agents? *The Knowledge Engineering Review*, **13**:131–136.

d' Aspremont, C. (2000). Competition Policy and Game Theory: Reflections Based on the Cement Industry Case. In Norman, G. and Thisse, J.-F., eds., *Market Structure and Competition Policy*, pages 9–30. New York: Cambridge University Press.

d'Aspremont, C., Gabszewicz, J. J. and Thisse, J. F. (1979). On Hotelling's 'Stability in Competition'. *Econometrica*, **47**:1045 – 1050.

Davis, L. (1991). *Handbook of Genetic Algorithms.* Van Nostrand Reinhold, New York.

de Frutos, M., Hamoudi, H. and Jarque, X. (2002). Spatial Competition with Concave Transport Costs. *Regional Science and Urban Economics*, **32**:531–540.

De Jong, K. A. (1975). *An Analysis of the Behaviour of a Class of Genetic Adaptive Systems.* Ph.D. thesis, University of Michigan, Ann Arbor, Michigan.

De Palma, A., Ginsburgh, V. and Victor, T. (1987). On Existence of Location Equilibria in the 3-firm Hotelling Problem. *Journal of Industrial Economics.*

Dijkstra, E. (1959). Two Problems in Connexion with Graphs. *Numerische Mathematik*, **1**:269 – 271.

Diplock, G. (1996). *The Application of Evolutionary Computing Techniques to Spatial Interaction Modelling.*. Ph.D. thesis, School of Geography, University of Leeds.

Diplock, G. and Openshaw, S. (1996). Using Simple Genetic Algorithms to Calibrate Spatial Interaction Models. *Geographical Analysis*, **28**:262–279.

Doran, J. (2003). Foundations and Applications of Multi-Agent Systems. *Lecture Notes in Artificial Intelligence*, **2403**:131–151.

Dosi, G., Marengo, L., Bassanini, A. and Valente, M. (1999). Norms as Emergent Properties of Adaptive Learning. *Journal of Evolutionary Economics*, 9(1):5–26.

Dudey, M. (1990). Competition by Choice: The Effect of the Consumer Search on Firm Location Decisions. *American Economic Review*, **80**:1092–1104.

Eaton, B. C. and Lipsey, R. G. (1975). The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition. *Review of Economic Studies*, **42**:27–49.

Ebdon, P., ed. (1985). *Statistics in Geography: A Practical Approach.* Blackwell, 2nd edn.

Eckert, A. (2002). Retail Price Cycles and Response Asymmetry. *Canadian Journal of Economics*, **35**:52 – 77.

Eckert, A. (2003). Retail Price Cycles and the Presence of Small Firms. *International Journal of Industrial Organisation*, **21**:151 – 170.

Economides, N. (1986). Minimal and Maximal Product Differentiation in Hotelling's Duopoly. *Economic Letters*, **21**:67–71.

Economides, N. (1993). Hotelling's " Main Street" With More Than Two Competitors. *Journal of Regional Science*, **33**:303–319.

Epstein, J. and Axtell, R. (1996). *Growing Artificial Societies - Social Science from the Bottom Up.* MIT Press Cambridge, 1st edn.

Evans, T. P. and Kelley, H. (2004). Multi-Scale Analysis of a Household Level Agent-Based Model of Landcover Change. *Journal of Environmental Management*, **72**(1-2):57–72.

Faminow, M. D. and Benson, B. L. (1990). Integration of Spatial Markets. *American Journal of Agricultural Economics*, **72**:49–62.

Ferber, J. (1999). *Multi-Agent Systems: An Introduction.* Addison-Wesley, Harlow.

Fik, T. J. (1988). Spatial Competition and Price Reporting in Retail Food Markets. *Economic Geography*, **64**:29–44.

Fik, T. J. (1991a). Price Competition and Node-Linkage Association. *Papers in Regional Science*, **70**:53–69.

Fik, T. J. (1991b). Price Patterns in Competitively Clustered Markets. *Environment and Planning A*, **23**:1545–1560.

Fik, T. J. and Mulligan, G. F. (1991). Spatial Competition: A Network Approach. *Geographical Analysis*, **23**:79 – 89.

Fischer, M. M. and Gopal, S. (1994). Artificial Neural Networks: A New Approach to Modelling Interregional Telecommunication Flows. *Journal of Regional Science*, **34**:503–527.

Fischer, M. M., Reismann, M. and Hlavackova-Schindler, K. (2003). Neural Network Modelling of Constrained Spatial Interaction Flows: Design, Estimation and Performance Issues. *Journal of Regional Science*, **43**(1):35–61.

Flake, G. W. (2001). *The Computational Beauty of Nature: Computer Explorations of Fractals, Chaos, Complex Systems and Adaption.* MIT Press, Cambridge MA, 4th edn.

Flowerdew, R. (1991). Poisson Regression Models of Migration. In Stillwell, J. and Congden, P., eds., *Migration Models: Macro and Micro Approaches*, pages 92–112. Belhaven Press, London.

Fogarty, T. C. (1989). Varying the Probability of Mutation in the Genetic Algorithm. In *Proceedings of the 3rd International Conference on Genetic Algorithms and their Applications.*, pages 104–109. George Mason University.

Fogel, L. J., Owens, A. J. and Walsh, M. J. (1966). *Artificial Intelligence Through Simulated Evolution.* Wiley, New York, 1st edn.

Foot, D. L. (1981). *Operational Urban Models.* Methuen, London, 1st edn.

Fotheringham, A. S. (1981). Spatial Structure and the Parameters of Spatial Interaction Models. *Annals of the Association of American Geographers*, **71**:425–436.

Fotheringham, A. S. (1983a). A New Set of Spatial Interaction Models; The Theory of Competing Destinations. *Environment and Planning A*, **15**:15–36.

Fotheringham, A. S. (1983b). Some Theoretical Aspects of Destination Choice and their Relevance to Production-Constrained Gravity Models. *Environment and Planning A*, **15**:1121–1132.

Fotheringham, A. S. (1984). Spatial Flows and Spatial Patterns. *Environment and Planning A*, **16**:529–543.

Fotheringham, A. S. (1985). Spatial Competition and Agglomeration in Urban Modelling. *Environment and Planning A*, **17**:213–230.

Fotheringham, A. S. (1986). Modelling Hierarchical Destination Choice. *Environment and Planning A*, **18**:401–418.

Fotheringham, A. S., Brunsdon, C. and Charlton, M. (2002). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Chichester: Wiley, 1st edn.

Fotheringham, A. S., Nakaya, T., Yano, K., Openshaw, S. and Ishikawa, Y. (2001). Hierarchical Destination Choice and Spatial Interaction Modelling: A Simulation Experiment. *Environment and Planning A*, **33**(5):901 – 920.

Franklin, S. and Graesser, A., eds. (1996). *Is it an Agent, or Just a Program? Proceedings Third International Workshop on Agent Theories, Architectures and Languages, Budapest, Hungary.*

Galoetti, M., Lanza, A. and Manera, M. (2003). Rockets and Feathers Revisited: An International Comparison on European Gasoline Markets. *Energy Economics*, **25**:175 – 190.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimisation and Machine Learning.* Addison-Wesley, Reading MA.

Goldberg, D. E. and Deb, K. (1991). A Comparative Analysis of Selection Schemes used in Genetic Algorithms. In Rawlins, G., ed., *Foundations of Genetic Algorithms*, pages 69–93. Morgan Kauffmann, San Mateo, CA.

Goodwin, P. and Wright, G. (1993). Improving Judgemental Time-Series Forecasting - A Review of the Guidance Provided by Research. *International Journal of Forecasting*, pages 147 – 161.

Green, E. J. and Porter, R. H. (1984). Non-Cooperative Collusion under Imperfect Price Information. *Econometrica*, **52**:87 – 100.

Greenhut, M. L., Norman, G. and Hung, C. S. (1987). *The Economics of Imperfect Competition: A Spatial Approach.* Cambridge University Press.

Greenwald, A. and Stone, P. (2001). Autonomous Bidding Agents in the Trading Agent Competition. *IEEE Internet Computing*, **March-April**:52 – 60.

Grefenstette, J. J. (1986). Optimisation of Control Parameters for Genetic Algorithms. In *IEEE Trans. on Systems Man and Cybernetics*, vol. SMC-16(1), pages 122–128.

Guldmann, J. M. (1999). Competing Destinations and Intervening Opportunities Interaction Models of Inter-City Telecommunication Calls. *Papers in Regional Science*, **78**:179 – 194.

Gupta, B. (1992). Sequential Entry and Deterrence with Competitive Spatial Price Discrimination. *Economic Letters*, **38**:487–490.

Guttman, R., Moukas, A. G. and Maes, P. (1998). Agent-Mediated Electronic Commerce. *The Knowledge Engineering Review*, **13**:147–149.

Haff, P. (2001). Waterbots. In Hillebrand, E. and Stender, J., eds., *Landscape Erosion and Evolution Modelling*, pages 239–275. Kluwer Academic / Plenum Publishers, New York.

Haining, R., Plummer, P. and Sheppard, E. (1996). Spatial Price Equilibrium In Interdependent Markets: Prices and Sales Configuration. *Papers in Regional Science*, **75**:41 – 64.

Haining, R. P. (1983a). Anatomy of a Price War. *Nature*, **304**:679–681.

Haining, R. P. (1983b). Modelling Intra-Urban Price Competition: An Example of Gasoline Retailing. *Journal of Regional Science*, **23**:517 – 528.

Haining, R. P. (1985). The Spatial Structure of Competition and Equilibrium Price Dispersion. *Geographical Analysis*, **17**:231–242.

Haining, R. P. (1986). Intra-Urban Retail Price Competition: Corporate and Neighbourhood Aspects of Spatial Price Variation. In Norman, G., ed., *London Papers in Regional Science 16. Spatial Pricing and Differentiated Markets*, pages 144–164. Pion: London.

Hancock, P. J. B. (1994). An Empirical Comparison of Selection Methods in Evolutionary Algorithms. In Fogarty, T., ed., *Evolutionary Computing: AISB Workshop, Leeds, UK, April 1994: Selected Papers*, pages 80–94. Springer-Verlag.

Harp, S. A. and Samad, T. (1991). Genetic Synthesis of Neural Network Architecture. In Davis, L., ed., *Handbook of Genetic Algorithms*, pages 203–221. Van Nostrand Reinhold, New York.

Haverkamp, D. S. and Gauch, S. (1998). Intelligent Information Agents: Reviews and Challenges for Distributed Information Sources. *Journal of the American Society for Information Science*, 49(4):304–311.

He, M., Leung, H.-F. and Jennings, N. R. (2002). A Fuzzy Logic Based Bidding Strategy for Autonomous Agents in Continuous Double Auctions. *IEEE Transactions of Knowledge and Data Engineering.*, **30**:1–19.

Heppenstall, A. J. (2001). *Investigating Relationships in Water Quality using Artificial Intelligence and Statistical Techniques*. Master's thesis, Geography, University of Leeds.

Hewitt, C. (1977). Viewing Control Structures as Patterns of Passing Messages. *Artificial Intelligence*, **8**:323 – 364.

HM Treasury (2000). Prudent for a Purpose: Working for a Stronger and Fairer Britain.

Holland, J., ed. (1975). *Adaption in Natural and Artificial Systems*. MIT Press, Cambridge MA.

Holland, J. H. and Miller, J. H. (1991). Artificial Adaptive Agents in Economic Theory. *American Economic Review*, **81**:365–379.

Hotelling, H. (1929). Stability in Competition. *Economic Journal*, **39**:41 – 57.

Huff, D. L. (1961). A Note on the Limitations of intra-urban gravity models. *Land Economics*, **38**:64–66.

Huff, D. L. (1963). A Probabilistic Analysis of Shopping Centre Trade Areas. *Land Economics*, **39**:81–90.

Huff, D. L. (1970). Defining and Estimating a Trade Area. In Ambrose, P., ed., *Analytical Human Geography*. Longman, London.

Isard, W. (1960). *Method of Regional Analysis*. MIT Press.

Janikow, C. Z. and Michalewicz, Z. (1991). An Experimental Comparison of Binary and Floating Point Representations in Genetic Algorithms. In Belew, R. and Booker, L., eds., *Proceedings of the 4th International Conference on Genetic Algorithms*. Morgan Kaufmann, San Mateo, CA.

Jennings, N., Corera, J. M., Laresgoiti, L., Mamdani, E., Perriollat, F., Skarek, P. and Varga, L. (1995). Using ARCHON to Develop Real-World DAI Applications for Electricity Transportation and Particle Accelerator Control. *IEEE Expert: Special Issue on Real World Applications of DAI systems.*

Jennings, N. R. (2000). On Agent-Based Software Engineering. *Artificial Intelligence*, **117**:277 – 296.

Jennings, N. R., Sycara, K. and Wooldridge, M. (1998). A Roadmap of Agent Research and Development. *Autonomous Agents and Multi-Agent Systems*, **1**:7–38.

Jin, D. and Lee, K. J. (2001). Impacts and Limitations of Intelligent Agents in Electronic Commerce. *Lecture Notes in Computer Science*, **2105**:33–48.

Karrenbrock, J. D. (1991). The Behaviour of Retail Gasoline Prices: Symmetric or not? *Federal Reserve Bank, St Louis Econ Rev*, **73**:19 – 29.

Kephart, J., Hanson, J. and Greenwald, A. (2000). Dynamic Pricing by Software Agents. *Computer Networks*, **32**(6):731 – 752.

Kinny, D. and Georgeff, M. (1996). Modelling and Design of Multi-Agent Systems. In Muller, J., Wooldridge, M. and Jennings, N., eds., *Intelligent Agents III: LNAI Vol 1193*. Springer, Berlin.

Knudsen, D. C. and Fotheringham, A. S. (1986). Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modelling. *International Regional Science Review*, **10**(2):127 – 147.

Kraus, S. (2002). *Strategic Negotiation in Multi-Agent Environments*. MIT Press, 1st edn.

Kreps, D. M. (1990). *Game Theory and Economic Modelling*. Oxford University Press.

Laureano-Cruces, A. L., de la Cruz-Gonzalez, J. M. and Ramirez-Rodriguez, J. (2004). Reactive Agents to Improve a Parallel Genetic Algorithm Solution. *Lecture Notes in Computer Science*, **2972**:698–706.

Lokta, A. (1910). Zur Theorie der Periodischen Reaktionen. *Z. phys. Cheie*, **72:508**.

Los Alamos National Laboratory (accessed 2004). Transims. URL `http://www-transims.tsasa.lanl.gov/`.

Manning, D. N. (1991). Petrol Prices, Oil Price Rises and Oil Price Falls: Evidence for the UK since 1972. *Applied Economics*, **23**:1535 – 1541.

Markopoulos, P. M. and Kephart, J. O. (2002). How Valuable are Shopbots? In *Proceedings of the First International Conference on Autonomous Agent-Based Systems, Bologna, Italy.*

Massive Ltd (accessed 2004). MASSIVE. URL `http://www.massivesoftware.com/`.

May, R. M. (1974). Biological Populations with Non-overlapping Generations: Stable Points, Stable Cycles, and Chaos. *Science*, **186(4164)**:645–647.

May, R. M. (1976). Simple Mathematical Models with Very Complicated Dynamics. *Nature*, **261(5560)**:459–467.

McFarland, O. E. (2003). *The Application of Multi-Agents and Social Network Theory to Petrol Pricing on UK Motorways.* Master's thesis, School of Geography, University of Leeds.

Meyer, B. (1988). *Object-Orientated Software Construction.* Prentice Hall, Englewood Cliffs, NJ.

Michaelewicz, Z. (1992). *Genetic Algorithms + Data Structures = Evolution Programs.* Springer, Berlin.

Mintel Intelligence Group Report (2003). Petrol, Oil and Antifreeze - UK - May 2003.

Mitchell, J. D., Ong, L. L. and Izan, H. Y. (2000). Idiosyncrasies in Australian Petrol Price Behaviour: Evidence of Seasonalities. *Energy Economics*, **28**:243 – 258.

Mulligan, G. (1989). Asymmetrical Price Conjectural Variation in Spatial Competition Models. *Economic Geography*, **65**:19 – 32.

Neven, D. (1987). Endogenous Sequential Entry in a Spatial Model. *International Journal of Industrial Organization*, **5**:419–434.

Ning, X. and Haining, R. (2003). Spatial Pricing in Interdependent Markets: A Case Study of Petrol Retailing in Sheffield. *Environment and Planning A*, **35**:2131 – 2159.

Noel, M. (2002). *Edgeworth Price Cycles in Retail Gas Markets.* Ph.D. thesis, MIT.

Norman, G. (1986). *London Papers in Regional Science 16: Spatial Pricing and Price Differentiated Markets.* Pion: London.

Nwana, H. S. (1996). Software Agents: An Overview. *Knowledge Engineering Review*, **11**(3):205 – 244.

O'Farrell, P. N. and Poole, M. A. (1972). Retail Grocery Price Variation in Northern Ireland. *Regional Studies*, **6**:83–92.

Office for National Statistics (1991). UK census. URL `http://census.ac.uk/casweb/`.

Office for National Statistics (1998). Household Expenditure per Head on Transport at Constant Prices, 1971 - 1998. URL `http://www.statistics.gov.uk/STATBASE/xsdataset.asp?vlnk=1477`.

Openshaw, S. (1976). An Empirical Study of Some Spatial Interaction Models. *Environment and Planning A*, **8**:23 – 41.

Openshaw, S. (1983). Multivariate Analysis of Census Data; The Classification of Areas. In Rhind, D., ed., *A Census User's Handbook*, pages 243–264. Methuen, London.

Openshaw, S. (1988). Building an Automated Modelling System to Explore a Universe of Spatial Interaction Models. *Geographical Analysis*, **20**:31–46.

Openshaw, S. (1993). Modelling Spatial Interaction Using a Neural Net. In Fischer, M. and Nijkamp, P., eds., *GIS, Spatial Modelling and Policy*, pages 147–164. Springer Verlag, Berlin.

Openshaw, S. (1994). Two Exploratory Space-Time-Attribute Pattern Analysers Relevant to GIS. In Fotheringham, A. and Rogerson, P., eds., *Spatial Analysis and GIS*, pages 83–104. Taylor and Francis, London.

Openshaw, S., Kneale, P., Corne, S. and See, L. (1998). The Feasibility of Neural Networks for Flood Forecasting. *MAFF Project OCS967P, School of Geography, University of Leeds.*

Openshaw, S. and Openshaw, C. (1997). *Artificial Intelligence in GIS.* John Wiley and Sons, New York, 1st edn.

O'Sullivan, D. and Haklay, D. (2000). Agent-based Models and Individualism: Is the World Agent-Based? *Environment and Planning A*, **32**(8):1409 – 1425.

O'Sullivan, D. B. (2000). *Graph-based Cellular Automaton Models of Urban Spatial Processes.* Ph.D. thesis, Barlett School of Architecture and Planning, University College London.

Palmer, R. G. (1994). Artificial Economic Life: A Simpler Model of a Stockmarket. *Physica D*, **75**:264–274.

Parker, A. J. (1974). Intra-urban Variations in Retail Grocery Prices. *Economic and Social Review*, **5**:393–403.

Parry, H., Evans, A. J. and Morgan, D. (2004). Aphid Population Dynamics in Agricultural Landscapes: An Agent-based Simulation Model. In *Proceedings of the International Environmental Modelling and Software Society Conference on Complexity and Integrated Resources Management, University of Osnabrück, Germany, 14-17 June.*

Parsons, S. and Wooldridge, M. (2002). Game Theory and Decision Theory in Multi-Agent Systems. *Autonomous Agents and Multi-Agent Systems*, **5**:243–254.

Pave, A. (1994). *Modelisation en biologie et en ecologie.* Lyon, Aleas.

Pham, D. T. and Karaboga, D. (2000). *Intelligent Optimisation Techniques.* Springer-Verlag, London, 1st edn.

Plummer, P. S. (1996). Spatial Competition Amongst Hierarchically Organised Corporations: Prices, Profits and Shipment Patterns. *Environment and Planning A*, **28**:199–222.

Plummer, P. S., Sheppard, E. and Haining, R. P. (1998). Modeling Spatial Price Competition: Marxian versus Neoclassical Approaches. *Annals of the American Association of Geographers*, **88**(4):575 – 594.

Precott, E. and Visscher, M. (1977). Sequential Location Among Firms with Foresight. *Bell Journal of Economics*, **8**:378–393.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 2nd edn.

Rechenberg, I. (1973). *Evolutionsstrategie: Optimierung Techniser Systeme nach Prinzipen der Biologischen Evolution*. Frommmann-Holzboog, Verlag, Stuttgart, 2nd (1993) edn.

Reeves, C. R. and Rowe, J. E. (2003). *Genetic Algorithms: Principles and Perspectives*. Kluwer Academic Publishers, 1st edn.

Reggiani, A., Nijkamp, P. and Sabella, E. (2001). New Advances in Spatial Network Modelling: Towards Evolutionary Algorithms. *European Journal of Operational Research*, **128**(2):385 – 401.

Reggiani, A. and Tritapepe, T. (2002). Neural Networks and Logit Models Applied to Commuter's Mobility in the Metropolitan Area of Milan. In Himanen, V., Nijkamp, P. and Reggiani, P., eds., *Neural Networks in Transport Applications*. Ashgate, Aldershot.

Reilly, B. and Witt, R. (1998). Petrol Price Asymmetries Revisited. *Energy Economics*, **20**:297 – 308.

Roozemond, D. A. (2001). Using Intelligent Agents for Pro-active, Real-Time Urban Intersection Control. *European Journal of operational Research*, **131**:293 – 301.

Roy, J. R. (1999). Areas, Nodes and Networks. *Papers in Regional Science*, **78**:135 – 155.

Roy, J. R. and Thill, J.-C. (2004). Spatial Interaction Modelling. *Papers in Regional Science*, **83**:339 – 361.

Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 1st edn.

Salop, S. (1979). Monopolistic Competition with Outside Gods. *Bell Journal of Economics*, **10**:141–156.

Samuelson, P. A. (1952). Spatial Price Equilibrium and Linear Programming. *American Economic Review*, **42**:283 – 303.

Schaffer, J. D., Caruana, R. A., Eshelman, L. J. and Das, R. (1989). Combinatorial Optimisation by Genetic Algorithms. In Schaffer, J. D., ed., *Proceedings of the 3rd International Conference on Genetic Algorithms*, pages 51–60. Morgan Kaufmann, San Mateo, CA.

Schelhorn, T., O'Sullivan, D., Haklay, M. and Thurstain-Goodwin, M. (1999). STREETS: An Agent-Based Pedestrian Model. *CASA Working Paper 9, University College London, Centre for Advanced Spatial Analysis.*

Schendel, D. E. and Balestra, P. (1969). Rational Behaviour and Gasoline Price Wars. *Applied Economics*, **1**:89 – 101.

Schneider, M. (1960). Panel Discussion on Inter-Area Travel Formula. *Tech. rep.*, Bulletin 253, Highway Research Board, National Academy of Sciences, Washington.

Schwefel, H.-P. (1977). *Numerical Optimisation of Computer Models*. John Wiley and Sons, 1st edn.

See, L. M. (1999). *Geographical Applications of Fuzzy Logic and Fuzzy Hybrid Techniques*. Ph.D. thesis, School of Geography, University of Leeds.

Sen, A. and Smith, T. E. (1995). *Gravity Models of Spatial Interaction Behaviour*. Springer-Verlag, 1st edn.

Sen, S., Sekaran, M. and Hale, J., eds. (1998). *Learning to Coordinate Without Sharing Information*, Proc. 12th Nation. Conf on Artificial Intelligence, Seattle, WA.

Sheppard, E., Haining, R. P. and Plummer, P. (1992). Spatial Pricing in Interdependent Markets. *Journal of Regional Science*, **32**:55 – 75.

Sheppard, E., Haining, R. P. and Plummer, P. (1998). Profit Rate Maximisation in Interdependent Markets. *Journal of Regional Science*, **38**:675 – 76.

Sheremetov, L. B., Contreras, M. and Valencia, C. (2004). Intelligent Multi-Agent Support for the Contingency Management System. *Expert Systems with Applications*.

Sherstuk, Y. (2004). Solutions On Oil Refinery Information Storage Construction on the Basis of MS SQL Server. *Expert Systems with Applications*, **26**(1):73–79.

Shin, D. (1994). Do Petrol Prices Respond Symmetrically to changes in crude oil prices? *OPEC Review*, **18**:137 – 157.

Slade, M. E. (1992). Vancouver's Gasoline-Price Wars: An Empirical Exercise in Uncovering Supergame Strategies. *Review of Economic Studies*, **59**:257–276.

Stone, P. and Veloso, M. (1998). Multi-Agent Systems: A Survey from the Machine Learning Perspective. *IEEE Transactions on Knowledge and Data Engineering*.

Stouffer, S. A. (1940). Intervening Opportunities: A Theory Relating Mobility and Distance. *American Sociological Review*, **5**:845 – 867.

Stouffer, S. A. (1960). Intervening opportunities and Competing Migrants. *Journal of Regional Science*, **2**:1–26.

Swarm development group (accessed 2003). Swarm. URL `http://www.swarm.org/`.

Szyperski, C., ed. (1998). *Component Software*. Addison-Wesley, Reading, MA, 1st edn.

Tabuchi, T. (1994). Two-Stage Two-Dimensional Spatial Competition Between Two Firms. *Regional Science and Urban Economics*, **13**:213–227.

Tabuchi, T. (1999). Pricing Policy in Spatial Competition. *Regional Science and Urban Economics*, **29**:617–631.

Tabuchi, T. and Thisse, J.-F. (1994). Asymmetric Equilibrium in Spatial Competition. *International Journal of Industrial Organisation*, **13**:213–227.

Takayama, T. and Judge, G. (1964). Equilibrium Among Spatially Separated Markets: A Reformulation. *Econometrica*, **32**:510–524.

Tesauro, G. and Bredlin, J. L. (2001). Strategic Sequential Bidding in Auctions using Dynamic Programming. In *Proceedings of the First International Conference on Autonomous Agents, Bologna, Italy.*

Thill, J. C. (1992). Choice Set Formation for Destination Choice Modelling. *Progress in Human Geography*, **16**:361–382.

Thomas, R. W. and Huggett, R. J. (1980). *Modelling in Geography: A Mathematical Approach.* Harper and Row: London.

Tobias, R. and Hofmann, C. (2004). Evaluation of Free Java-Libraries for Social-Scientific Agent Based Simulation. *Journal of Artificial Societies and Social Simulation*, **7**(1).

Torrens, P. M. (accessed 2004). SprawlSim. URL `http://www.sprawlsim.org/`.

Tsvetovat, P. M. and Carley, K. (2002). Emergent Specialisation in a Commodity Market: A Multi-Agent Model. *Computational and Mathematical Organisation Theory*, **8**:221–234.

Turton, I., Openshaw, S. and Diplock, G. (1997). A Genetic Programming Approach to Building New Spatial Models Relevant to GIS. In Kemp, Z., ed., *Innovations in GIS4*, pages 89 – 104. London: Taylor and Francis.

UK Borders (accessed 2002). UK Borders. URL `http://www.edina.ac.uk/ukborders/`.

University of Chicago Social Science Research Computing (accessed 2003). RePast. URL `http://repast.sourceforge.net/`.

Veendorp, E. C. H. and Majee, A. (1995). Differentiation in a Two-Dimensional Market. *Regional Science and Urban Economics*, **25**:75–83.

Walsh, W. E., Das, R., Tesauro, T. and Kephart, J. O. (2002). Analysing Complex Strategic Interactions in Multi-Agent Games. In *Game Theoretic and Decision Theoretic Agents, AAAI-02, Edmonton, Canada.*

Whitley, D. (1989). GENITOR; A Different Genetic Algorithm. *Proceeds of the Rocky Mountain Conference on Artificial Intelligence, Denver.*.

WICID (1991). Web Interface to Census Interaction Data. URL `http://census.ac.uk/cids/`.

Wikipedia (2003). Dijkstra's algorithm. URL `http://en.wikipedia.org/wiki/Dijkstras_Algorithm/`.

Willekins, F. J. (1983). Log-Linear Modelling of Spatial Interaction. *Papers of the Regional Science Association*, **52**:187–205.

Wilson, A. G. (1967). A Statistical Theory of Spatial Distribution Models. *Transportational Research*, **1**:253 – 269.

Wilson, A. G. (1970). *Entropy in Urban and Regional Modelling.* Pion, London.

Wilson, A. G. (1971). A Family of Spatial Interaction Models. *Environment and Planning A*, **3**:1 – 32.

Wilson, A. G. (1974). *Urban and Regional Models in Geography and Planning*. John Wiley & Sons, London.

Wilson, A. G. (1981). *Geography and the Environment*. John Wiley & Sons, New York.

Wooldridge, M. (2002). *An Introduction to Multi-agent Systems*. John Wiley & Sons Ltd, 1st edn.

Wooldridge, M. J. and Jennings, N. R., eds. (1997). *Intelligent Agents: Lecture Notes in Artificial Intelligence.*. CUP, 1st edn.

Wu, C. G., Liang, Y. C. and Lee, H. P. (2003). Study on Optimisation of Agent Initial Positions in Land Combat Simulation. *Prog Nat Sci*, **14**:257–261.

Yano, K. (1993). Integration of Spatial Interaction Models: Towards General Theory of Spatial Interaction. *Geographical Reports of Tokyo Metropolitan University*, **28**:33–78.

Zhou, D. and Vertinsky, I. (2001). Strategic Location Decisions in a Growing Market. *Regional Science and Urban Economics*, **31**:523–533.

Appendix A

Higher Resolution Maps

The following maps are higher resolution versions of the ones presented in §4.5 (Figures 4.5 and 4.6). They show the geographical and petrol station type classifications within West Yorkshire.

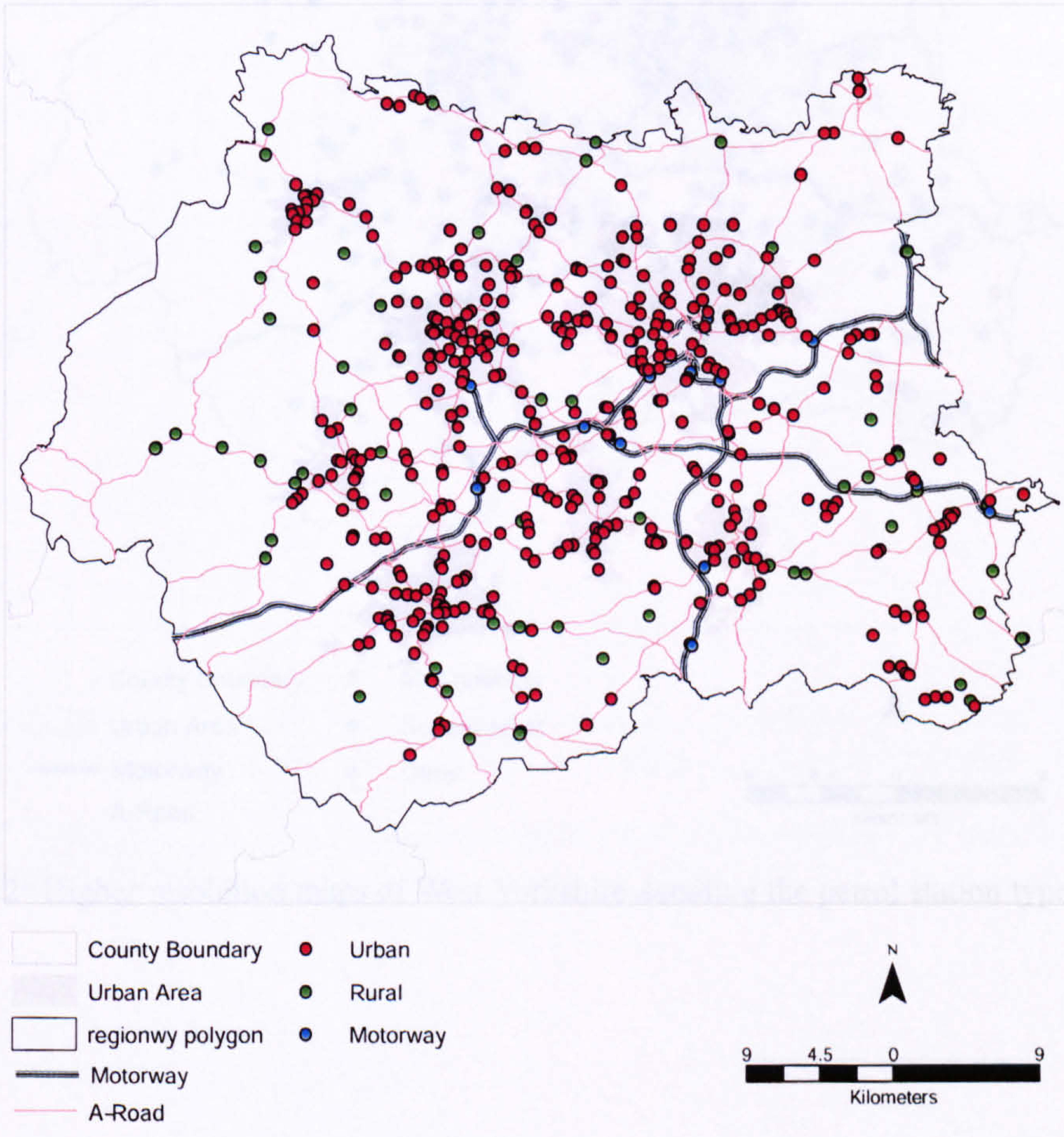


Figure A.1: Higher resolution maps of West Yorkshire detailing the geographical classification.

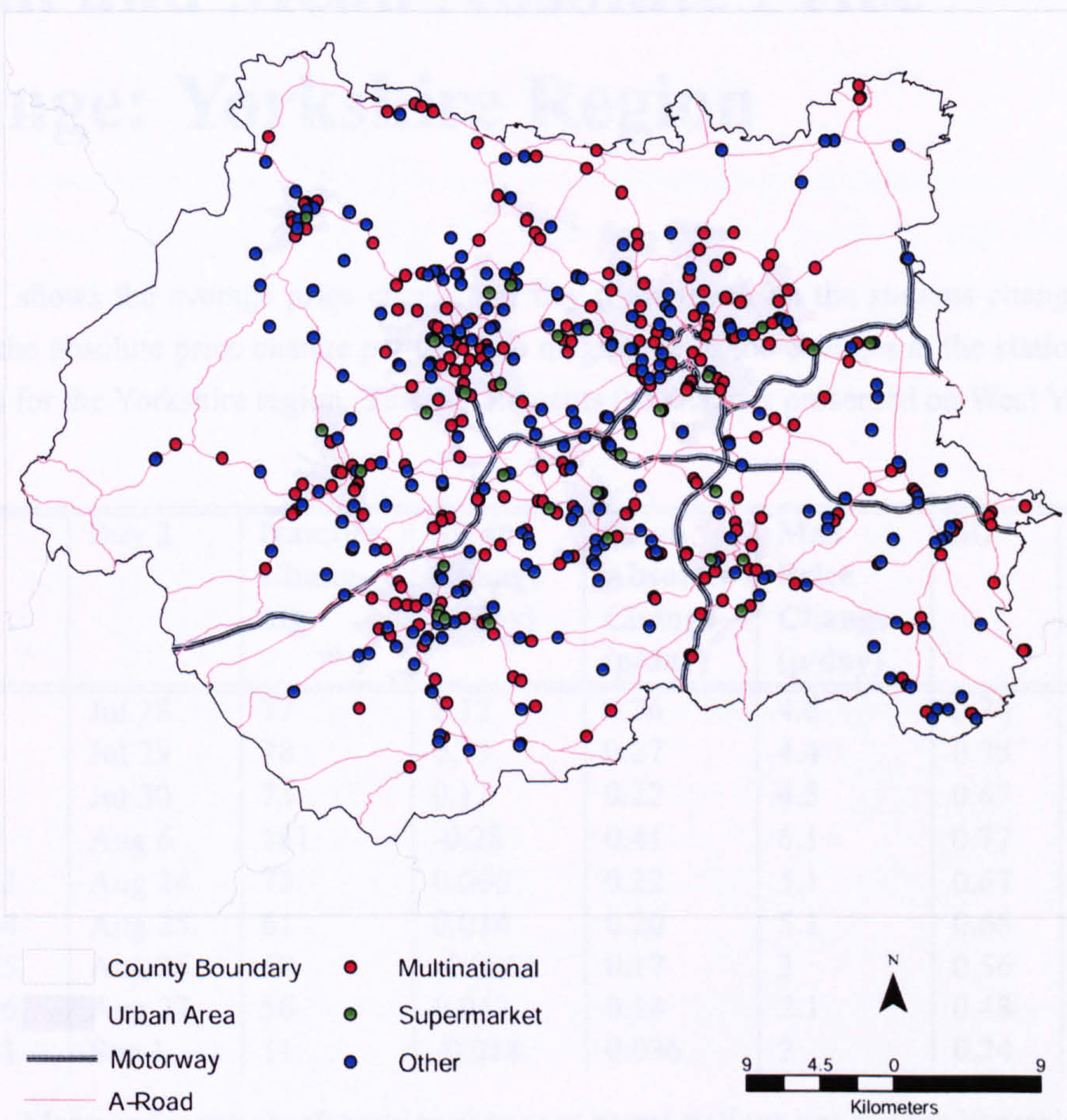


Figure A.2: Higher resolution maps of West Yorkshire detailing the petrol station type classification.

Appendix B

Mean and Mean Absolute Price Change: Yorkshire Region

Table B.1 shows the average price change per day (how much all the stations change on average) and the absolute price change per day (the magnitude of the changes at the stations that are changing) for the Yorkshire region. This supplements the analysis presented on West Yorkshire in §4.7.1.

Day 1	Day 2	Number Changing	Mean Change (p/day)	Mean Absolute Change (p/day)	Max Price Change (p/day)	SD	RMS
Jul 27	Jul 28	77	0.12	0.26	4.6	0.70	0.71
Jul 28	Jul 29	78	0.13	0.27	4.4	0.75	0.76
Jul 29	Jul 30	73	0.13	0.22	4.5	0.67	0.68
Aug 5	Aug 6	161	-0.28	0.41	6.1	0.77	0.82
Aug 23	Aug 24	73	0.066	0.22	5.1	0.67	0.67
Aug 24	Aug 25	61	0.014	0.20	5.1	0.65	0.65
Aug 25	Aug 26	59	-0.0056	0.17	3	0.56	0.56
Aug 26	Aug 27	56	0.013	0.14	2.1	0.48	0.48
Aug 31	Sep 1	11	-0.018	0.036	2	0.24	0.24

Table B.1: Mean and mean absolute price change at petrol stations per day for Yorkshire throughout the duration of the data set.

Appendix D

Esso Price Watch: Pure Agent Model

Figure D.1 shows the comparison of the mean price difference (p) and standard deviation of the Esso Price Watch rules and default rules. The results show that assignment of individual rules has not improved the overall performance of the agent model.

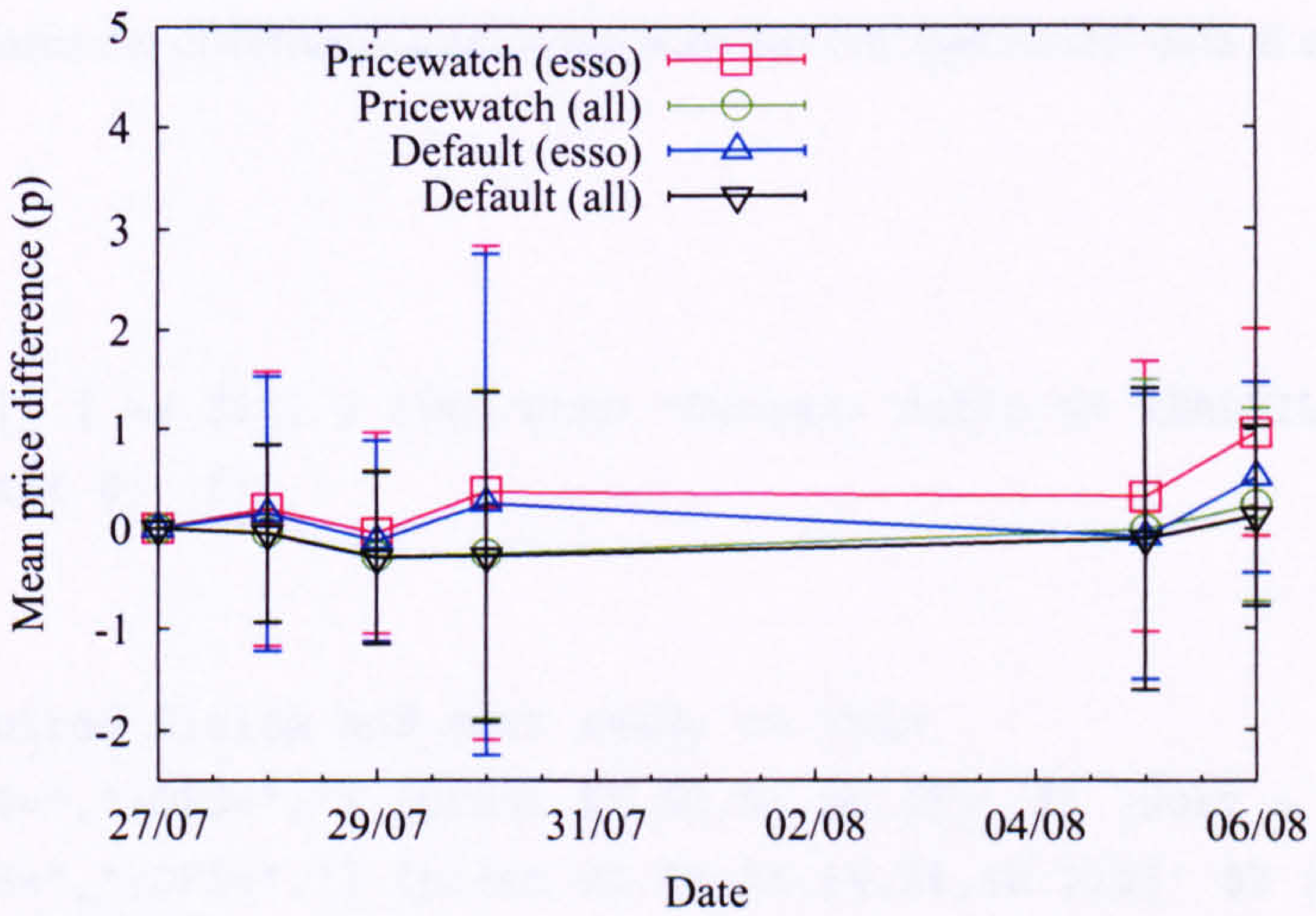


Figure D.1: The mean price difference (in pence) and SD (denoted by vertical bars) were calculated for Esso and non-Esso stations, with and without the Esso rules turned on.

Appendix C

Shell Scripts Used in Data Analysis

These shell scripts are used in the analysis of the real and model data results.

C.1 DIFFS.SH

This script calculates the difference in price between the real and model data at each petrol station on a given day.

```
#!/bin/sh

if [ -z $1 ] || [ -z $2 ] ; then echo "Syntax: diffs.sh REALFILE
MODELFILE"; exit 0;  fi


# Extract required fields and sort ready to join.
awk 'BEGIN {FS=",";OFS=","} {print $1,$2,$3,$4,$5}' $1 |sort > real.csv
awk 'BEGIN {FS=",";OFS=","} {print $1,$2,$3,$4,$6,$8,$10}' $2 |sort >
model.csv


# Join the two files togther - only keep the garages which appear in both
join -t, -1 1 -2 1 model.csv real.csv > merged.csv

# Now work out the differences and print them out.
awk 'BEGIN {FS=",";OFS=","}
{
    if ($4>0 && $8>0) d1=$8-$4; else d1="";
    if ($5>0 && $9>0) d2=$9-$5; else d2="";
    if ($6>0 && $10>0) d3=$10-$6; else d3="";
    if ($7>0 && $11>0) d4=$11-$7; else d4="";
    print $1,$2,$3,d1,d2,d3,d4;
}' merged.csv | sort -n
```

C.2 STATS.AWK

Using the output from the DIFFS.SH, this script calculates the mean, SD, Root Mean Square (RMS) error and the SRMSE of the differences between the real and model data at each petrol station. This can also be used for the prices as well as price differences.

```
#!/usr/bin/awk -f

BEGIN {
    FS=",";
    OFS=",";
    CONVFMT="%.15g";
    for (i=1;i<=4;i++) {
n[i] = 0;
mean[i] = 0;
meansq[i] = 0;
stddev[i] = 0;
    }
    line = 1;
}

{
    for (i=1;i<=4;i++) {
ii = i+3;
price[line,i] = $ii;
if ($ii != "") {
    n[i]++;
    mean[i]+=$ii;
    meansq[i]+=$ii*$ii;
}
    }
    line++;
}

END {
    for (i=1;i<=4;i++) {
if (n[i] > 0) {
    mean[i] = mean[i]/n[i];
    meansq[i] = sqrt(meansq[i]/n[i]);
```

```
}
else {
    mean[i] = "";
    meansq[i] = "";
}
    }
    for (i=1;i<=4;i++) {
if (n[i] > 1) {
    for (l=1;l<line;l++) {
if (price[l,i] != "") {
stddev[i]+=(price[l,i]-mean[i])*(price[l,i]-mean[i]);
}
    }
    stddev[i] = sqrt(stddev[i]/(n[i]-1));
}
else {
    stddev[i] = "";
}
    }
    print
    mean[1],mean[2],mean[3],mean[4],stddev[1],stddev[2],stddev[3],
    stddev[4],meansq[1],meansq[2],meansq[3],meansq[4];
}
```

PAGE

NUMBERING

AS ORIGINAL

Appendix E

Parameter Space for GA Optimal Parameters

E.1 Optimal Parameters for Esso

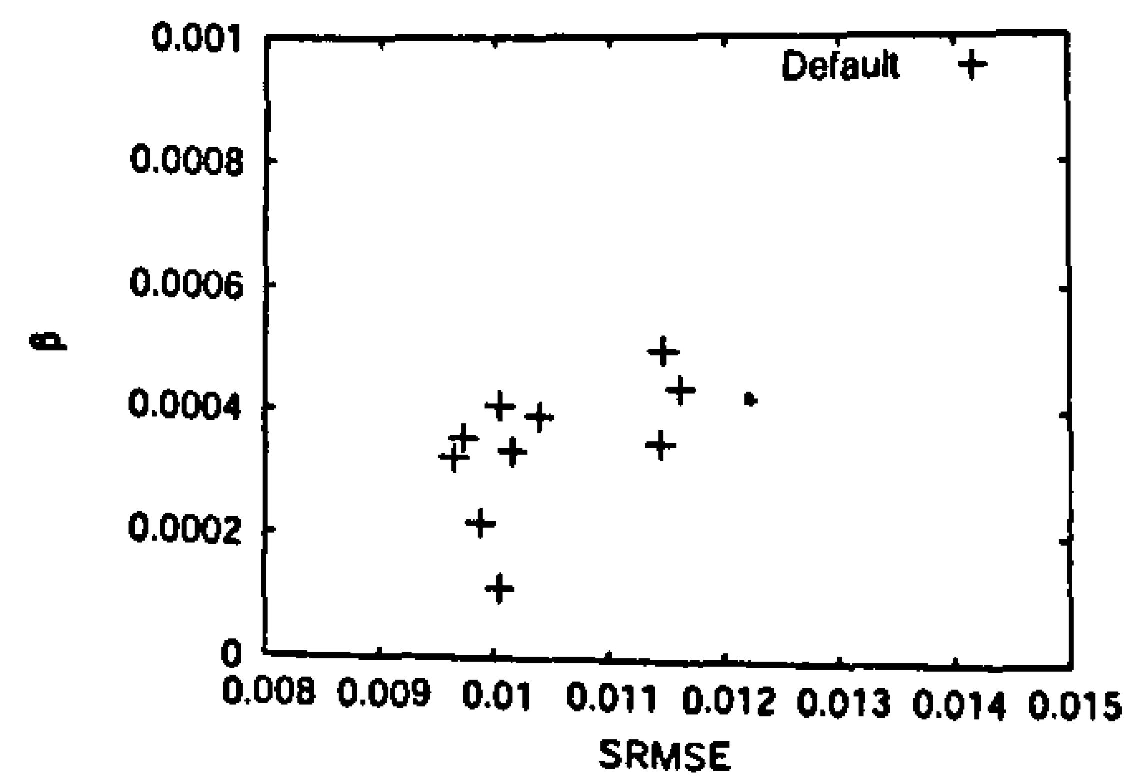
The GA model was run 10 times using the default GA parameters for the Esso rule set. The best of these parameters were taken as the optimal parameters for the model. Figure E.1 shows the variation in model parameters for these simulations.

E.2 Optimal Parameters for Supermarket

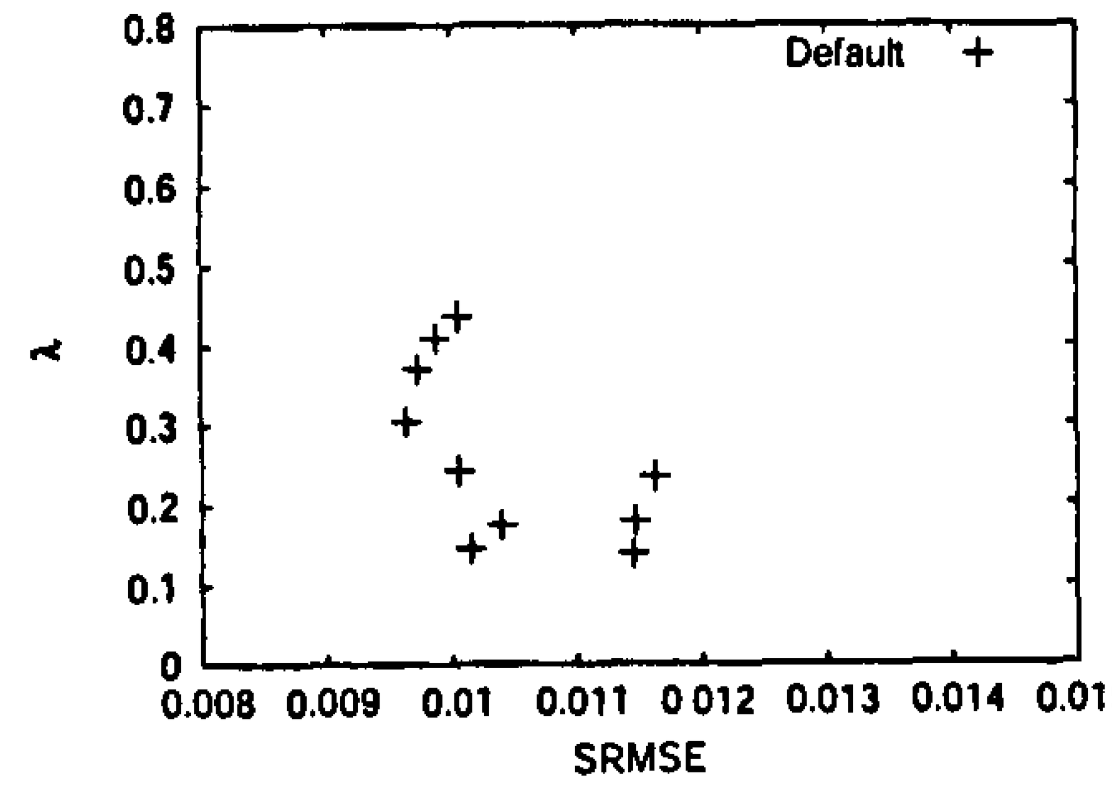
The GA model was run 10 times using the default GA parameters for the supermarkets. The best of these parameters were taken as the optimal parameters for the model. Figure E.2 shows the variation in model parameters for these simulations.

E.3 Optimal Parameters for Late Data

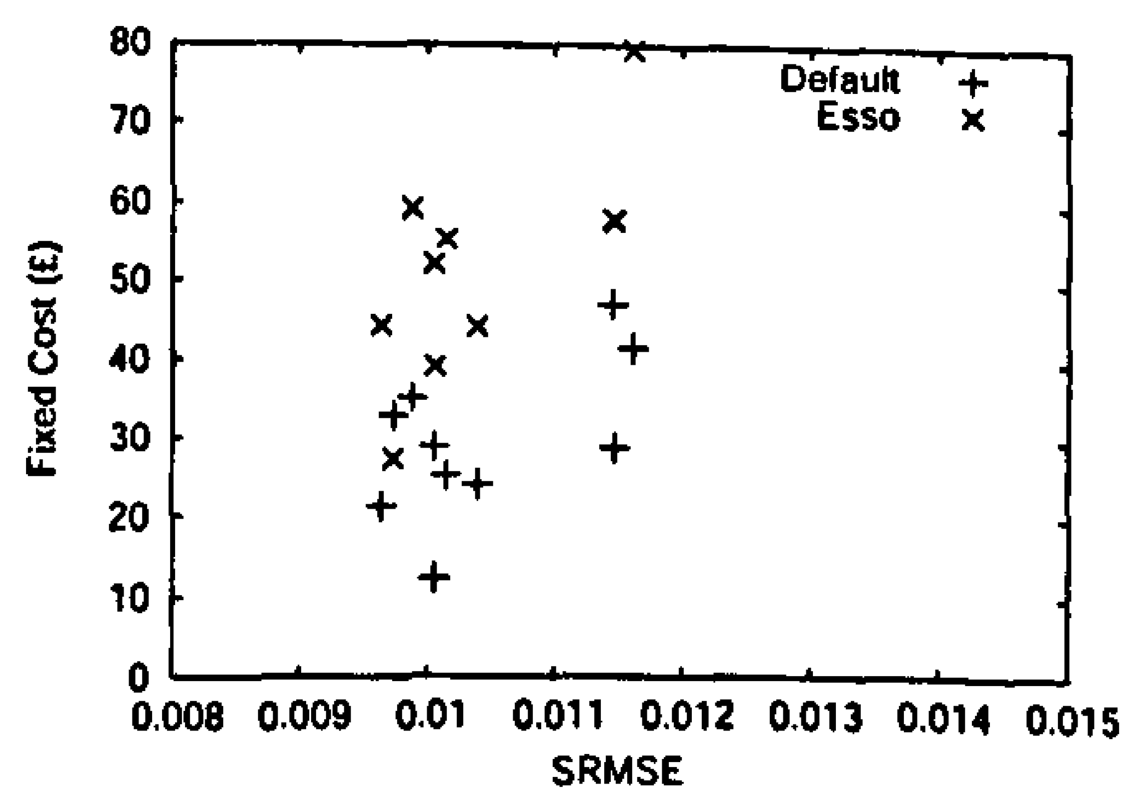
The GA model was run 10 times using the default GA parameters for the August 19th data. The best of these parameters were taken as the optimal parameters for the model. Figure E.3 shows the variation in model parameters for these simulations.



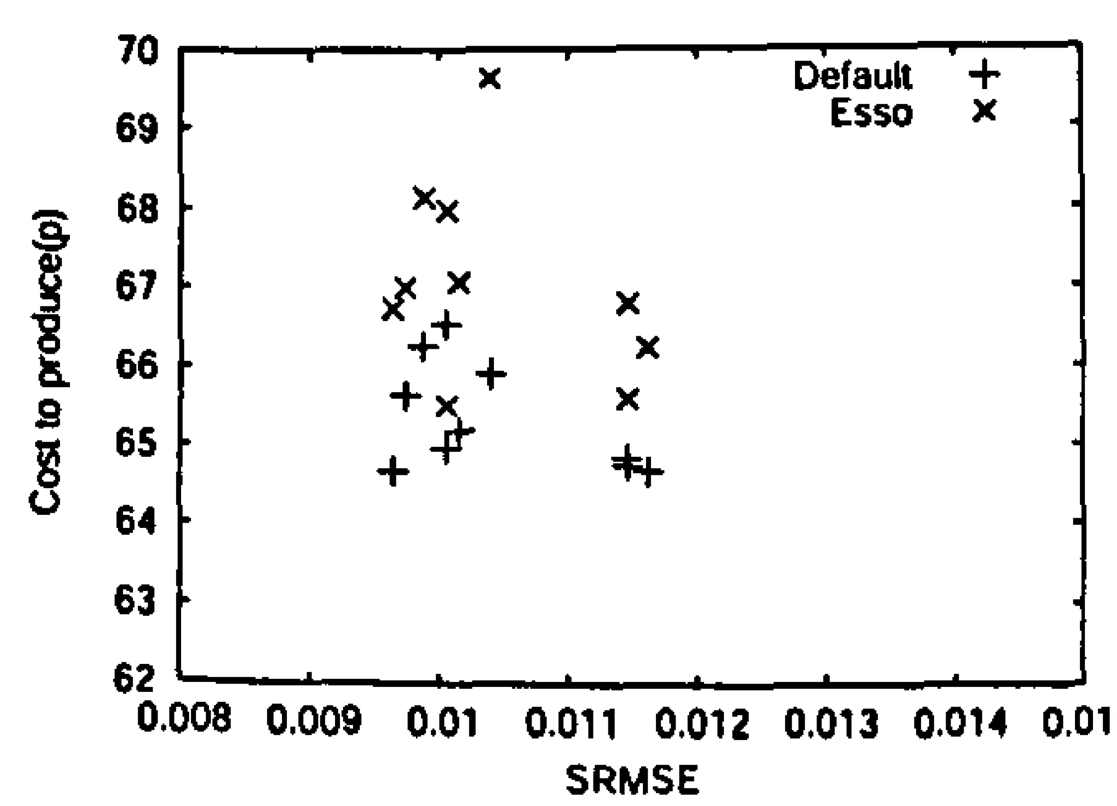
(a) β



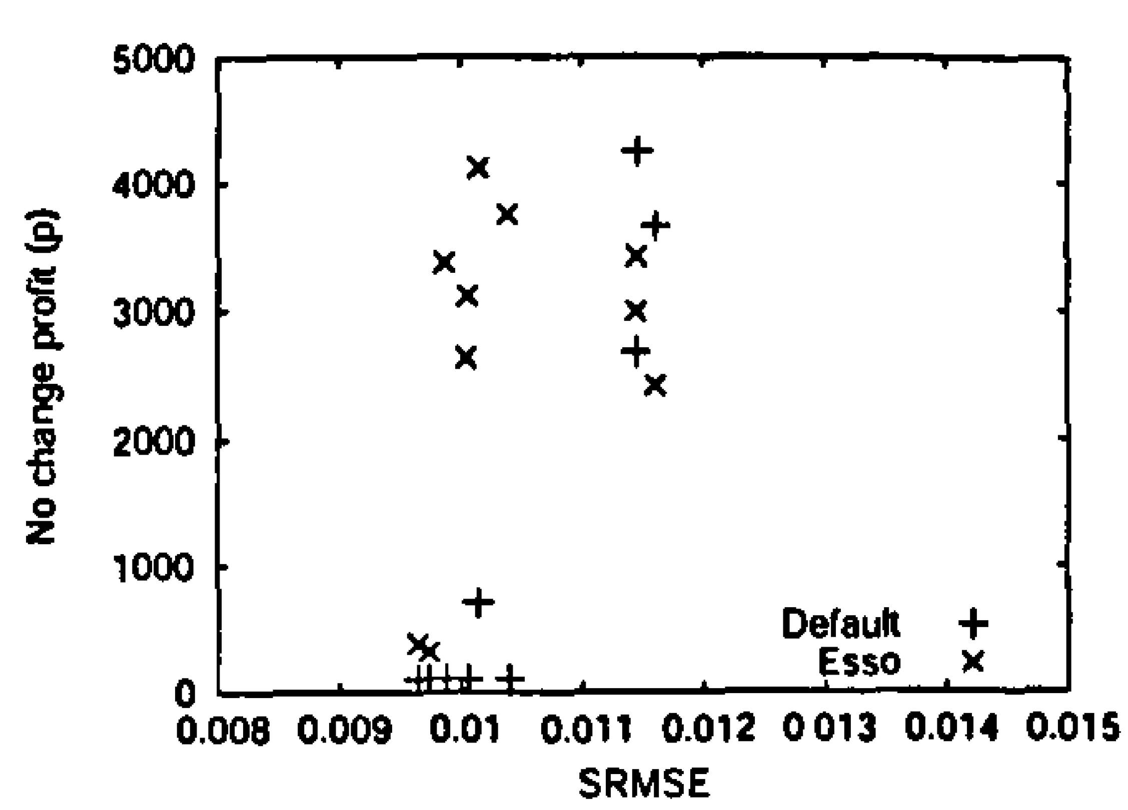
(b) λ



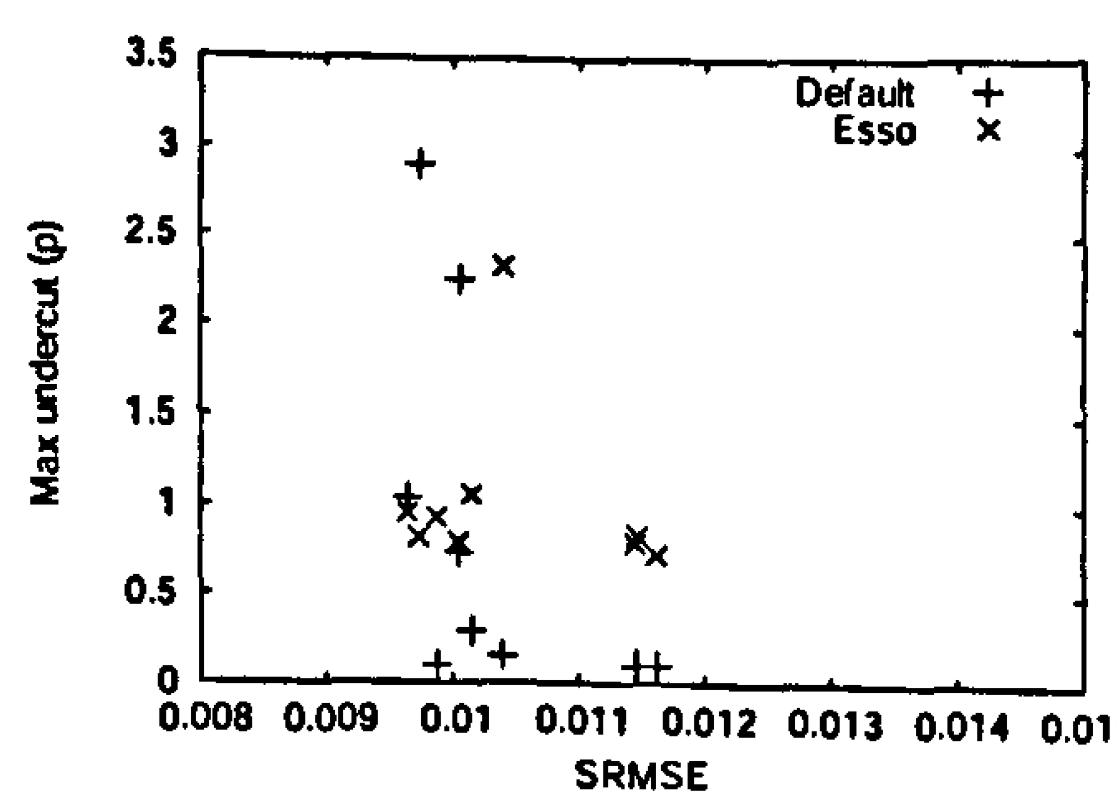
(c) *fixedCosts*



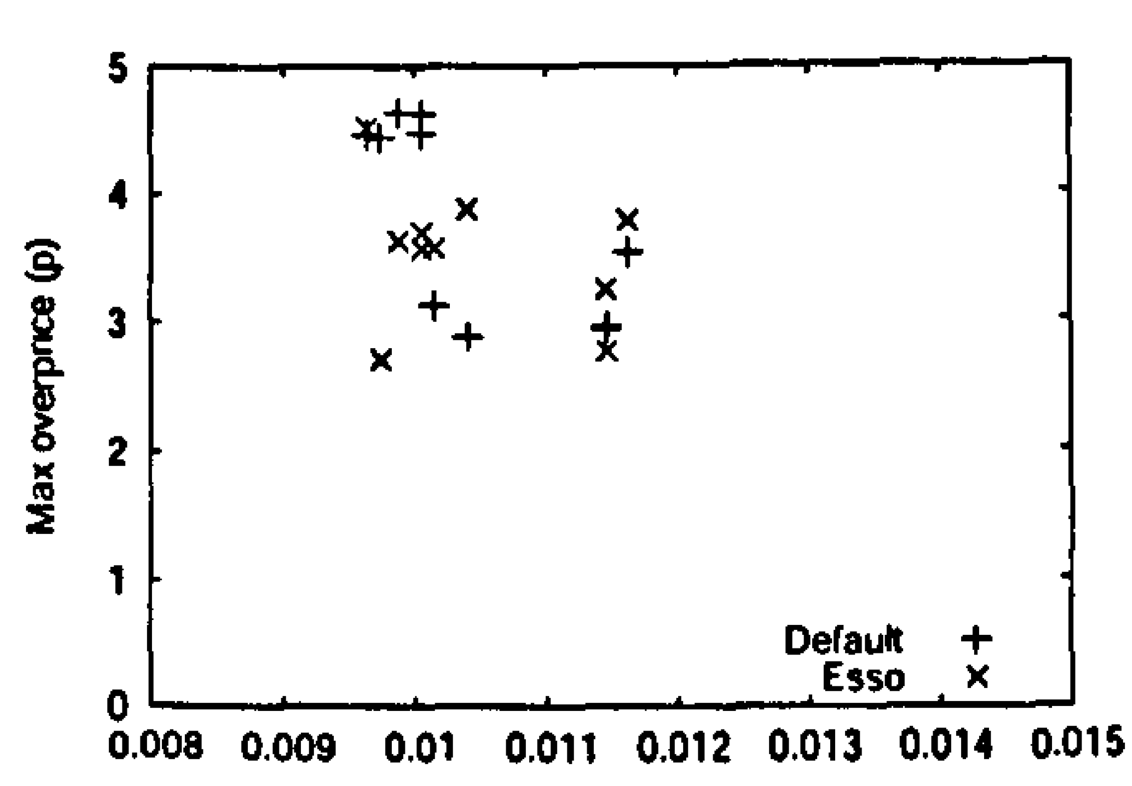
(d) *costToProduce*



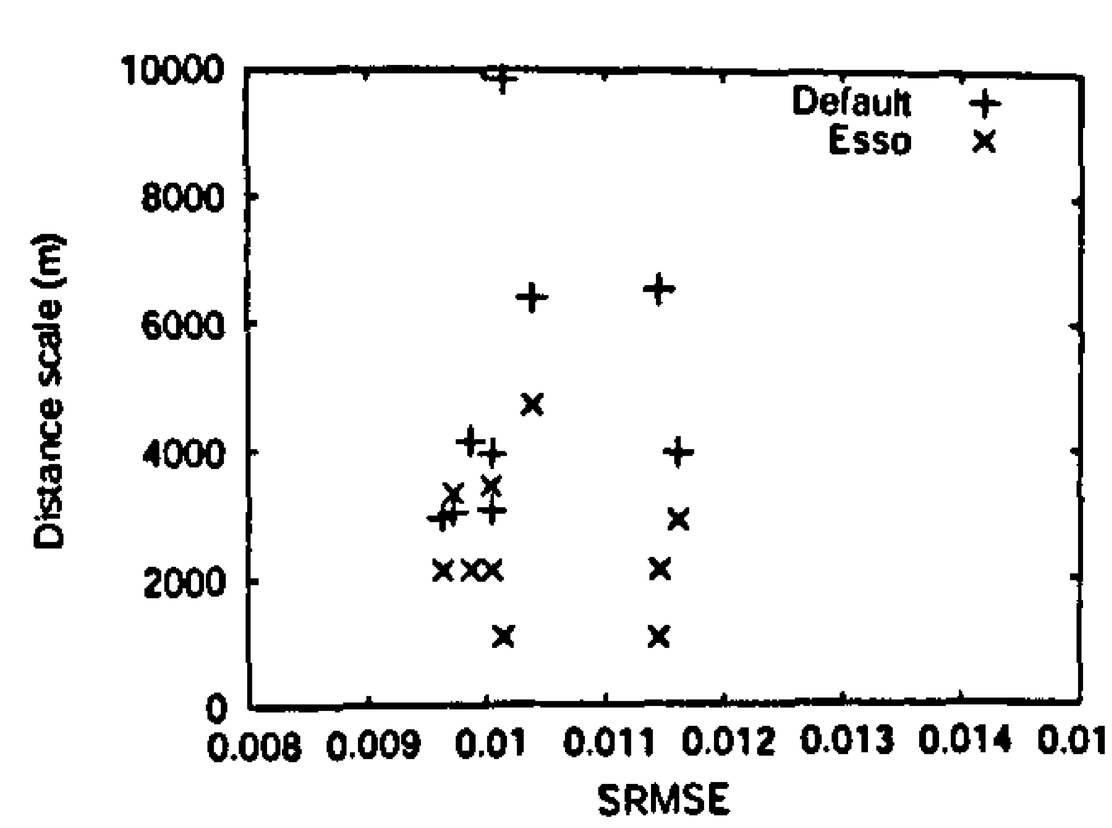
(e) *noChangeProfit*



(f) *undercut*

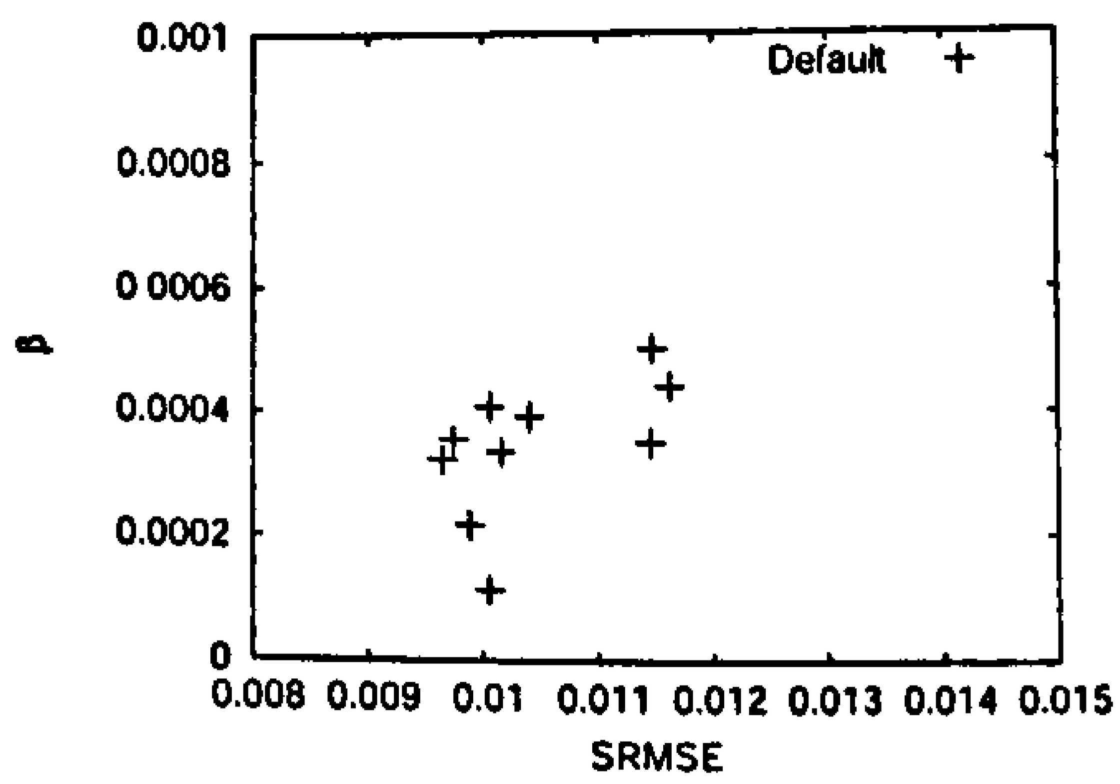


(g) *overprice*

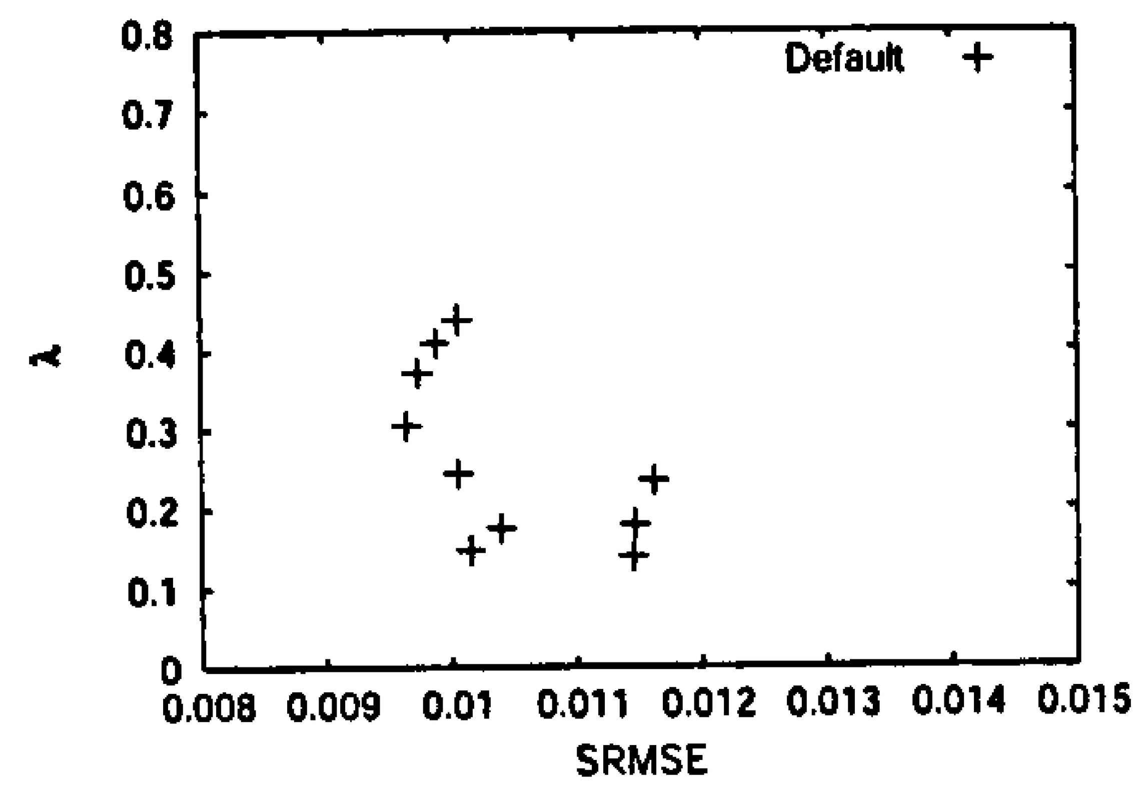


(h) *neighbourhood*

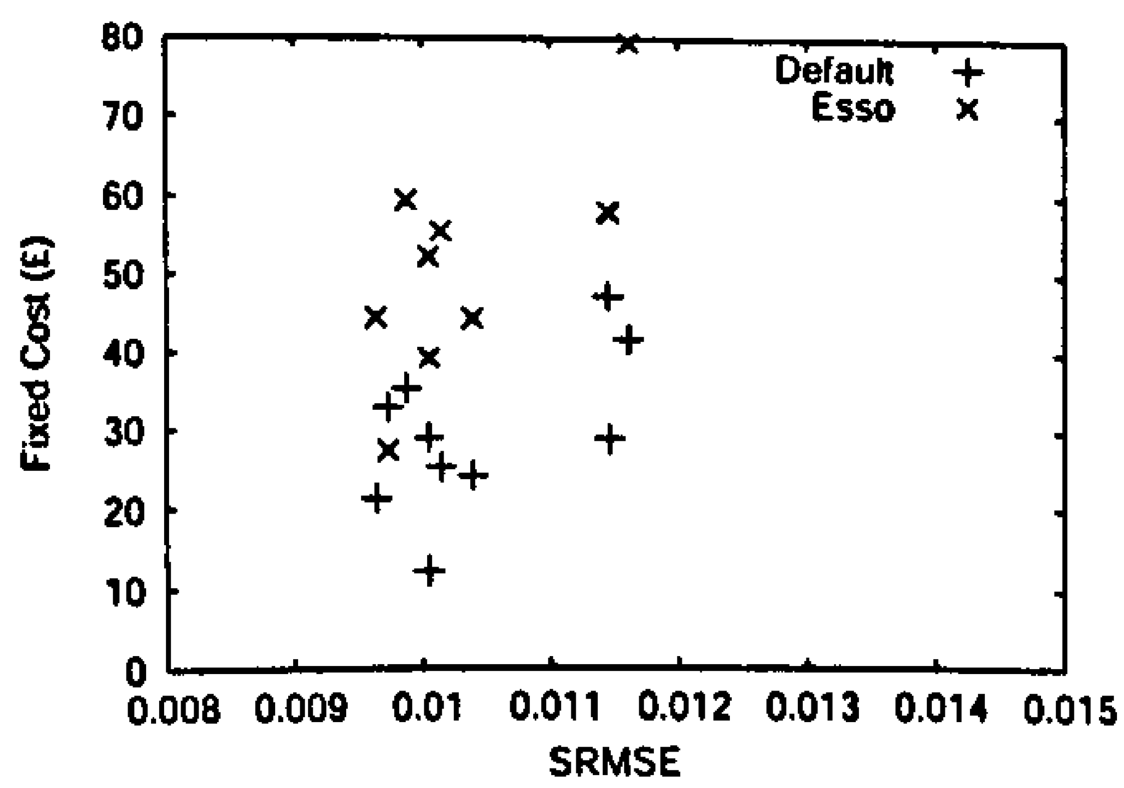
Figure E.1: Variation in model parameters plotted against SRMSE for simulations with different GA control parameters. The experiments (described in §9.12.1) are using the hybrid model to develop rules for both Esso and non-Esso stations.



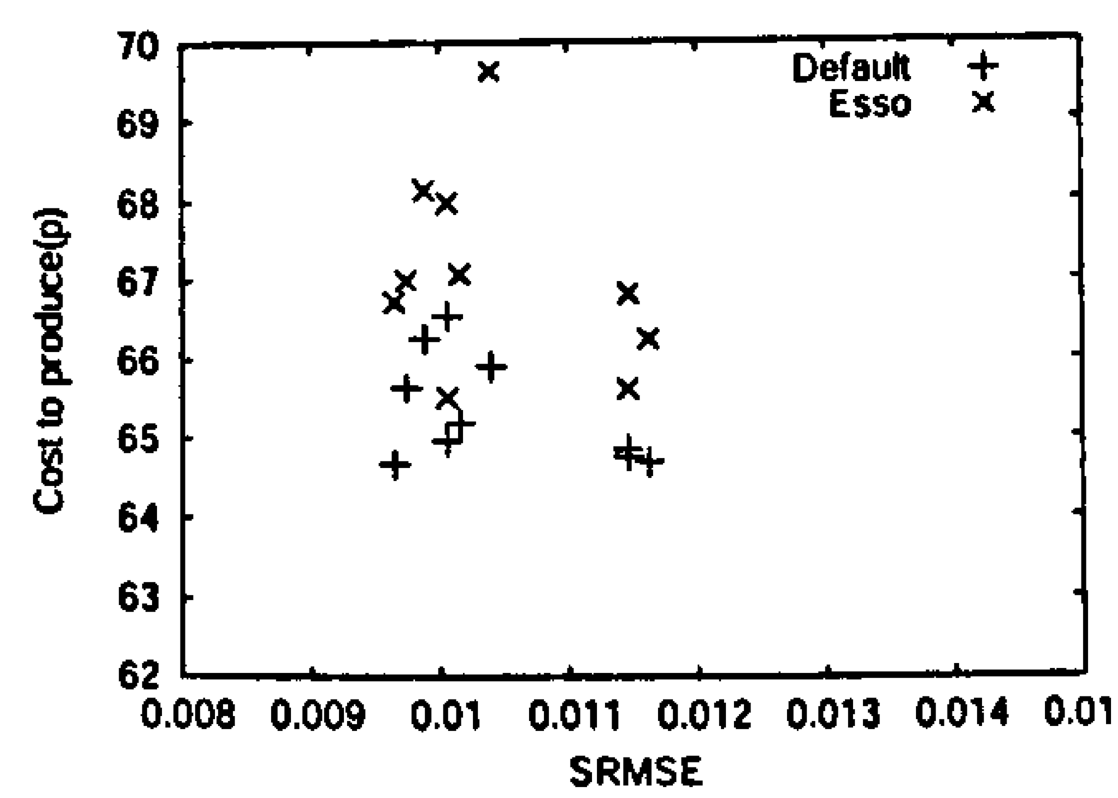
(a) β



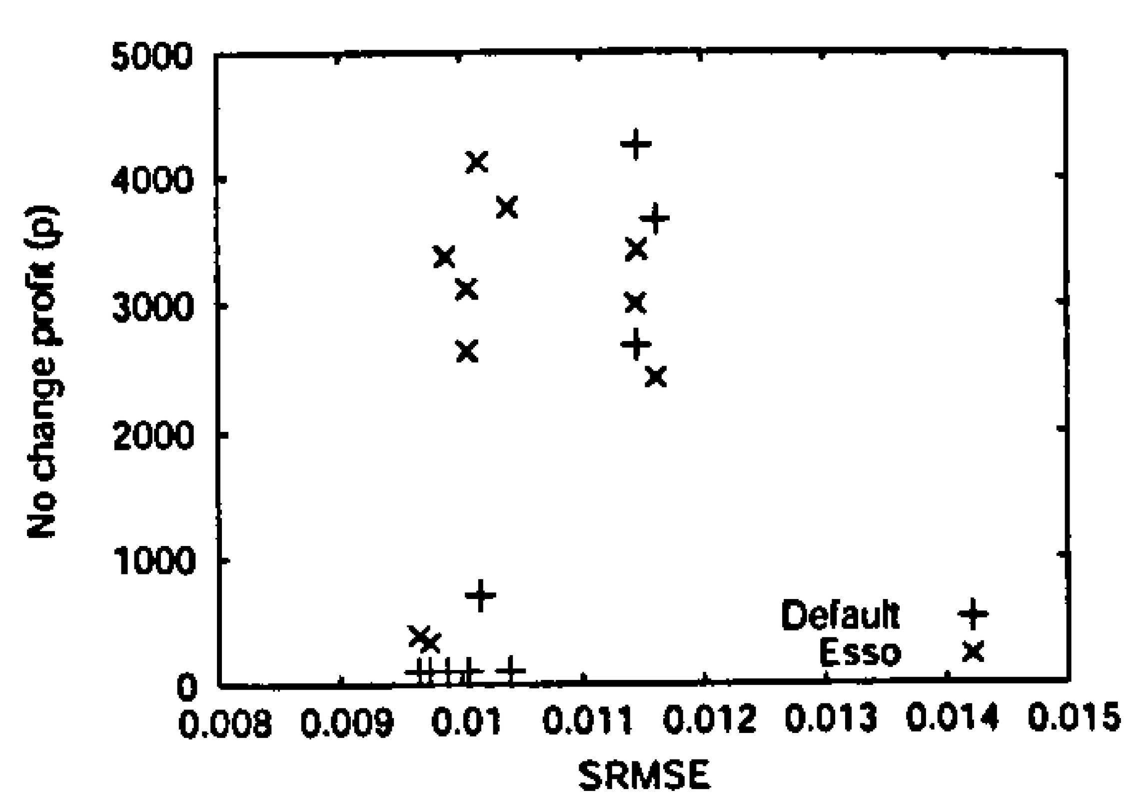
(b) λ



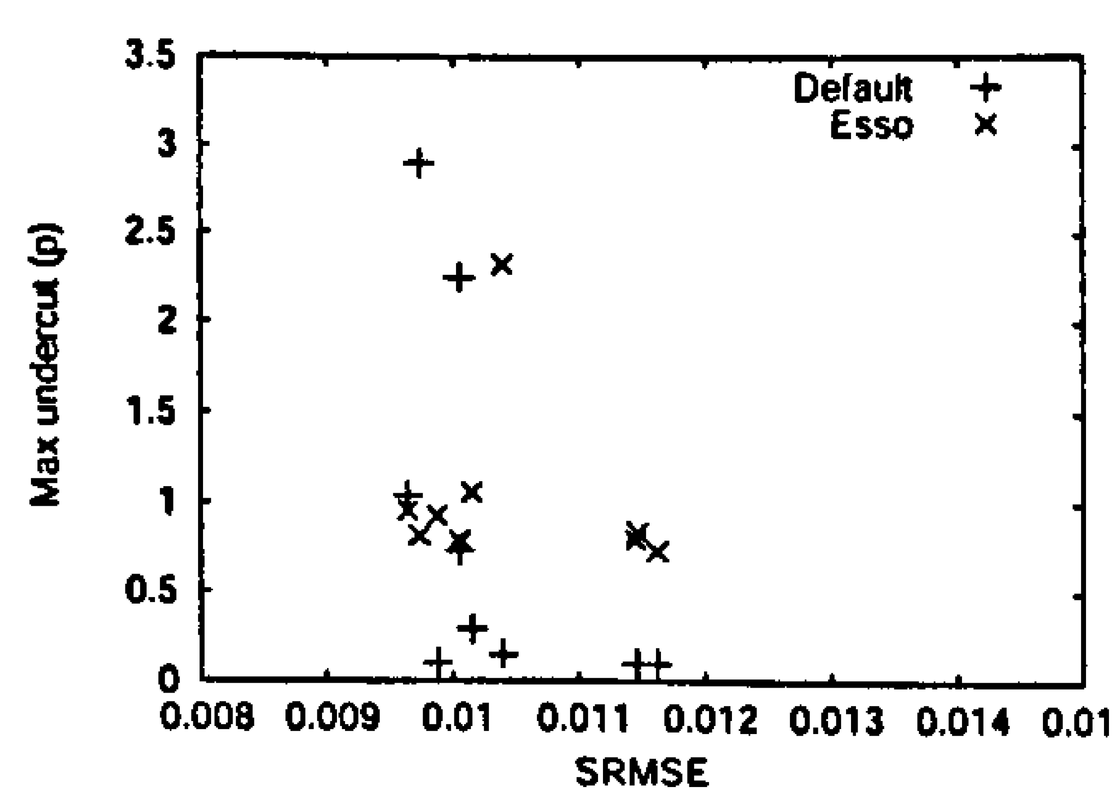
(c) *fixedCosts*



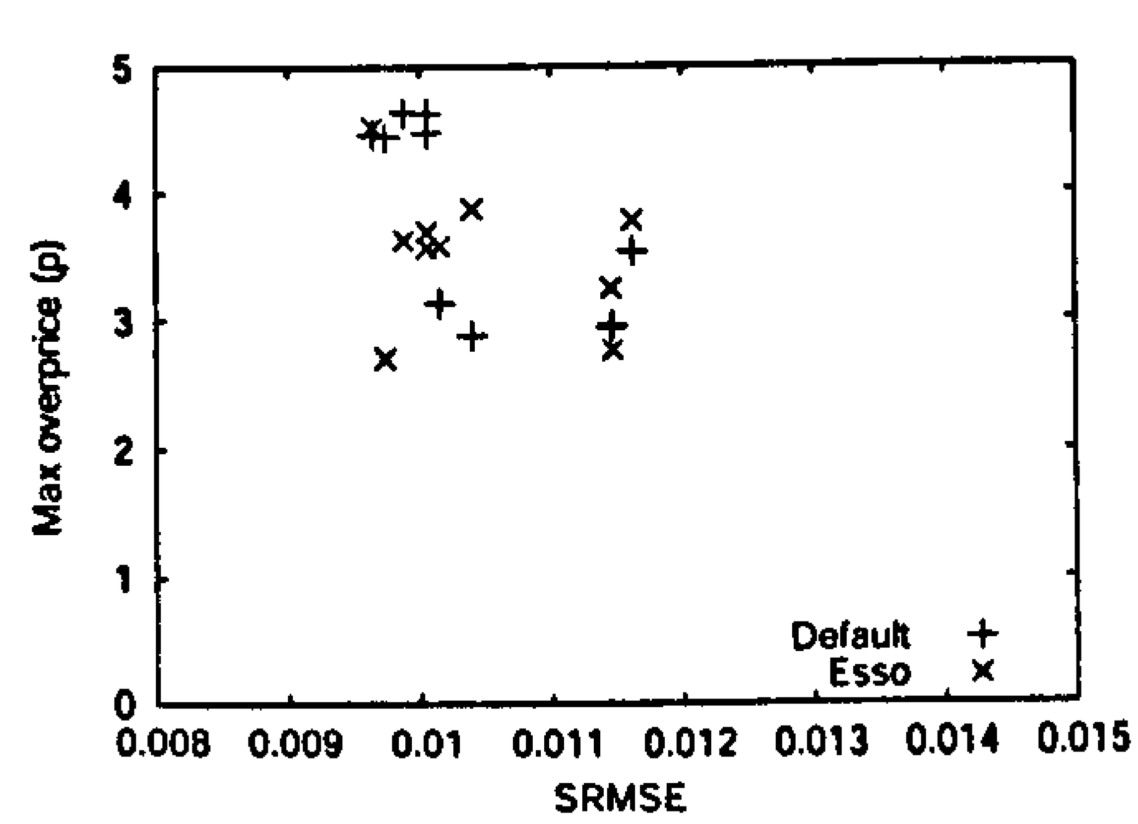
(d) *costToProduce*



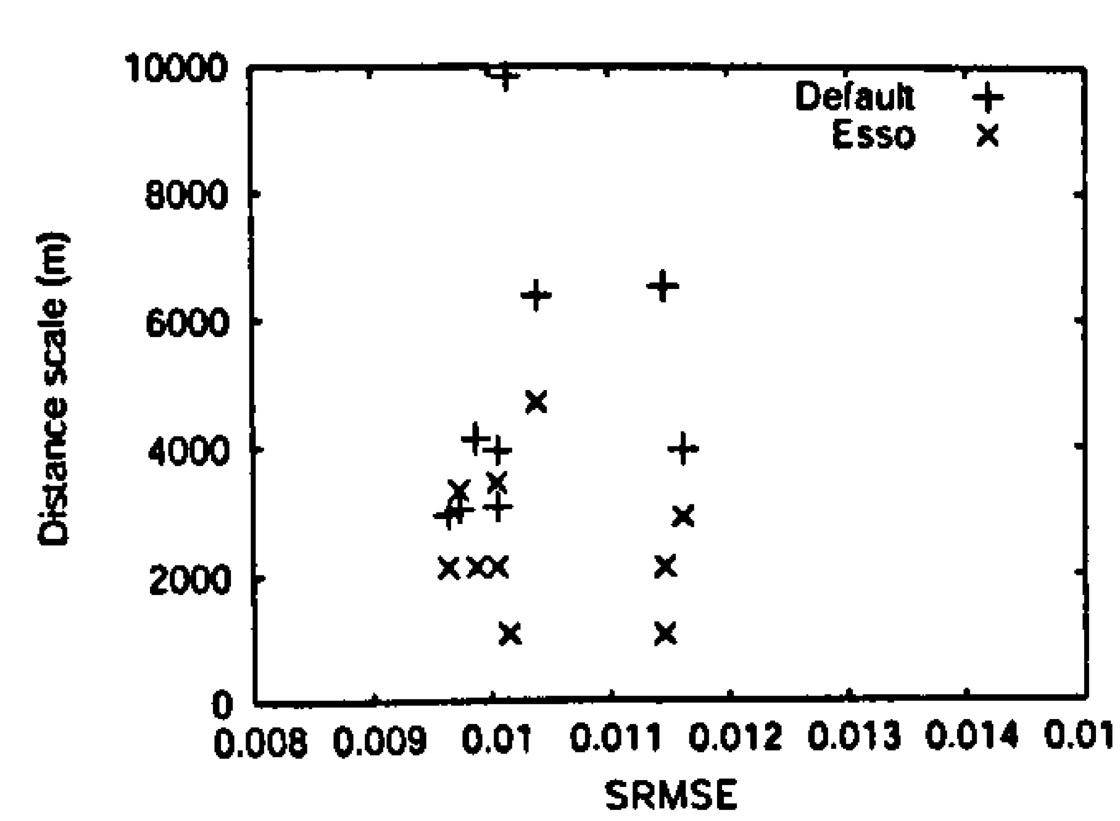
(e) *noChangeProfit*



(f) *undercut*

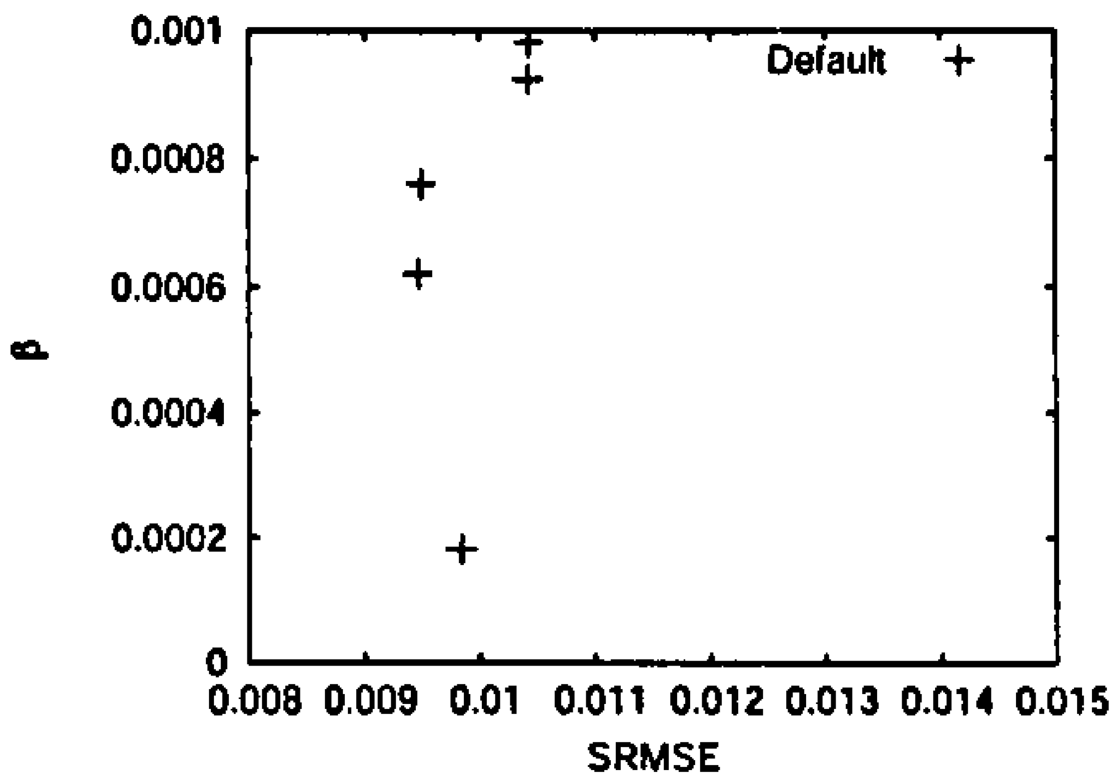


(g) *overprice*

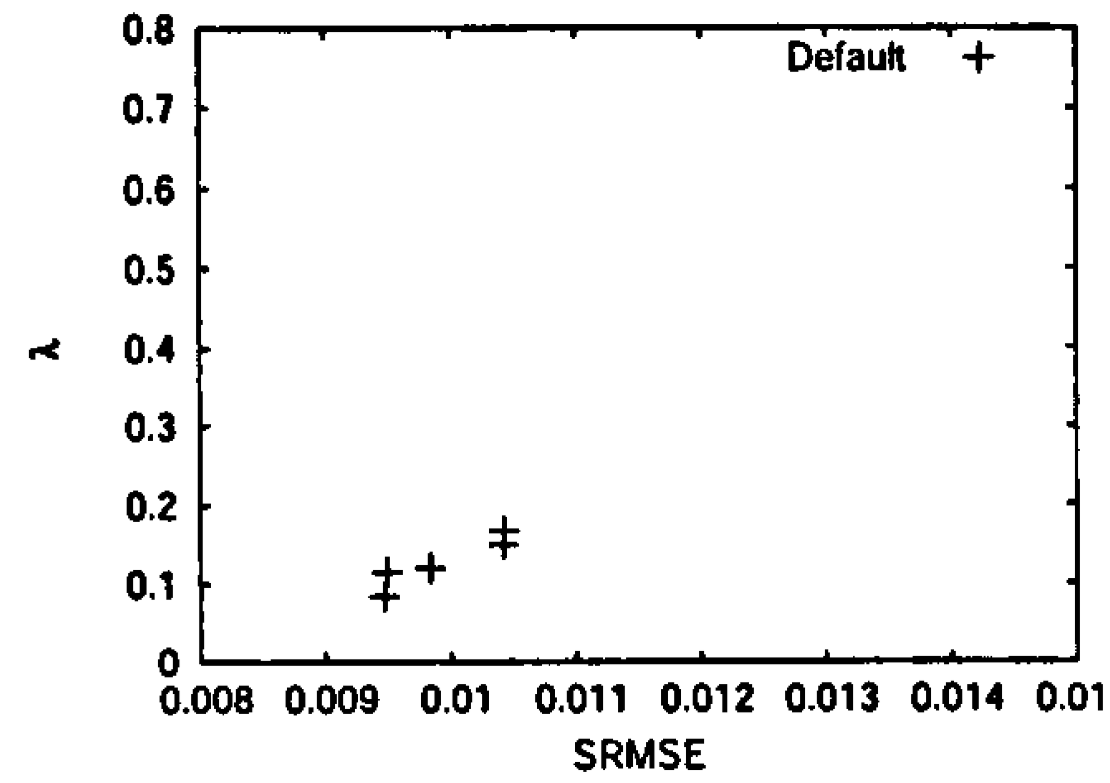


(h) *neighbourhood*

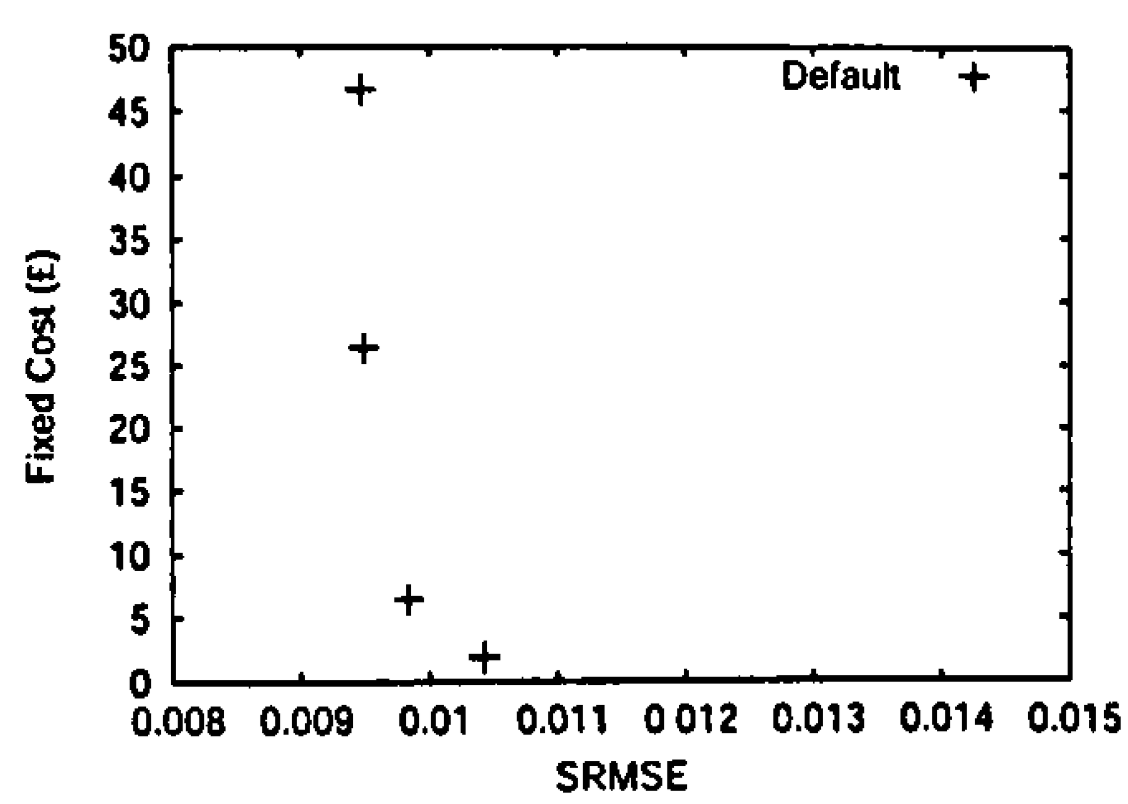
Figure E.2: Variation in model parameters plotted against SRMSE for simulations with different GA control parameters. The experiments (described in §9.12.2) are using the hybrid model to develop rules for both supermarket and non-supermarket stations.



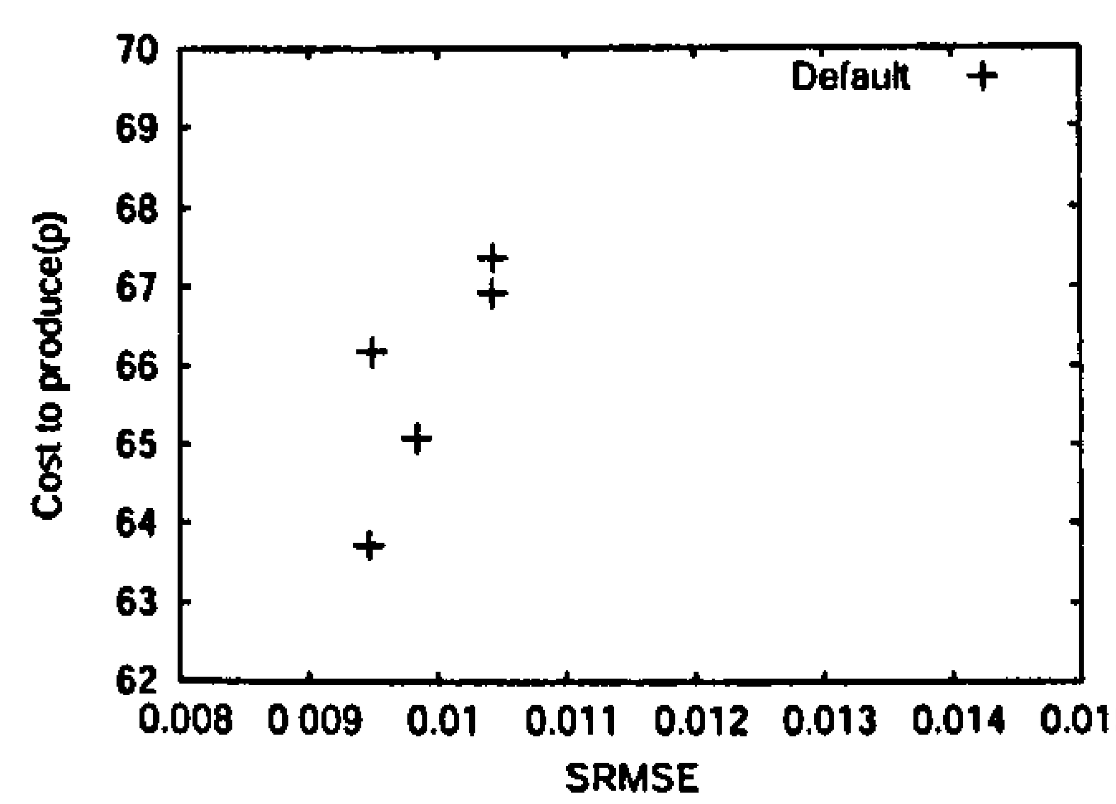
(a) β



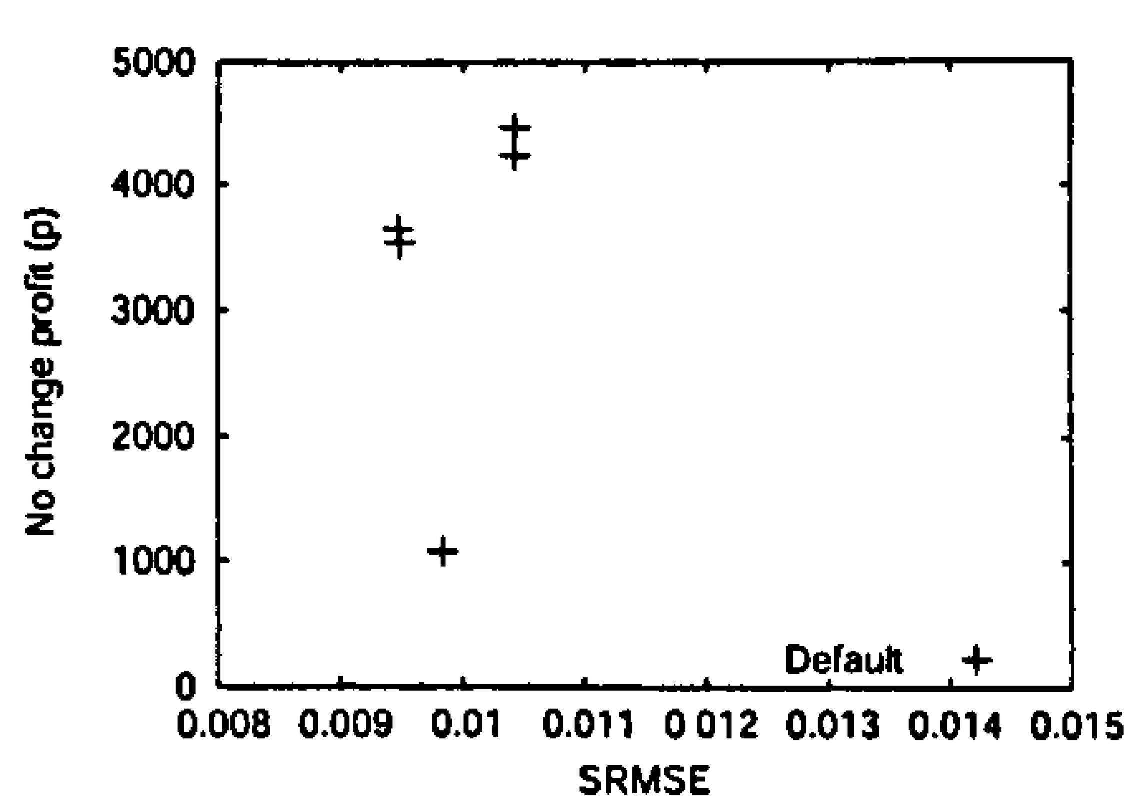
(b) λ



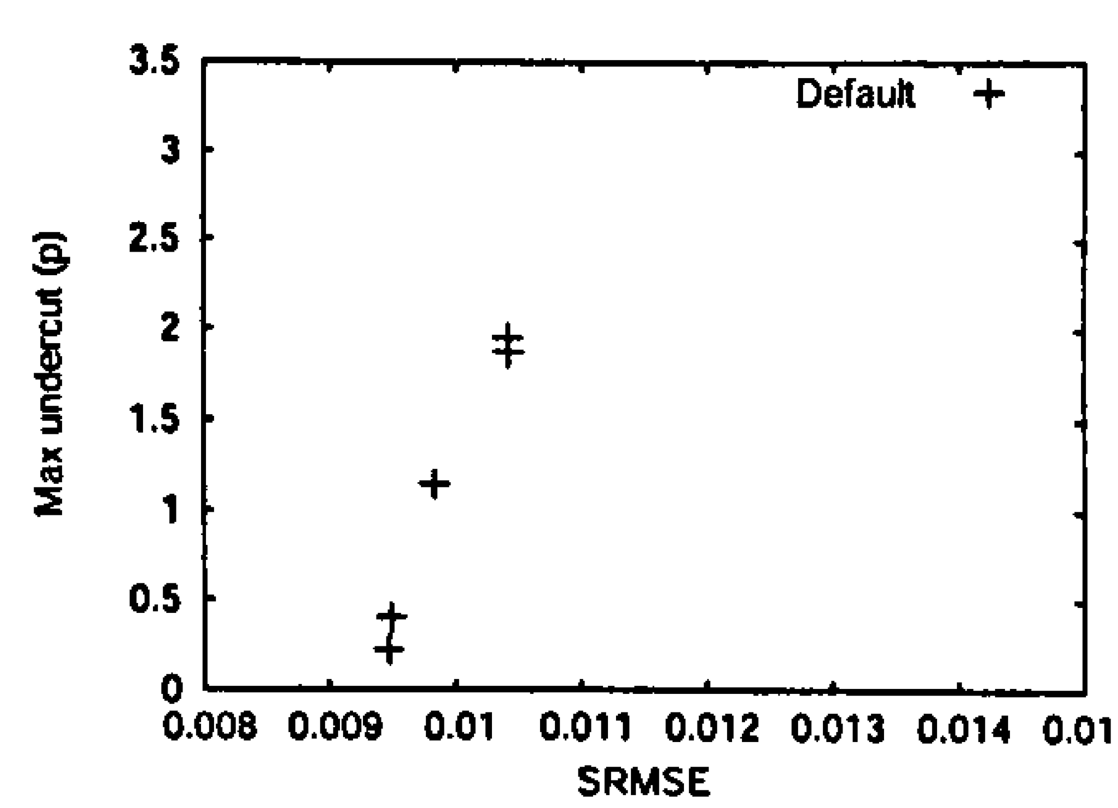
(c) *fixedCosts*



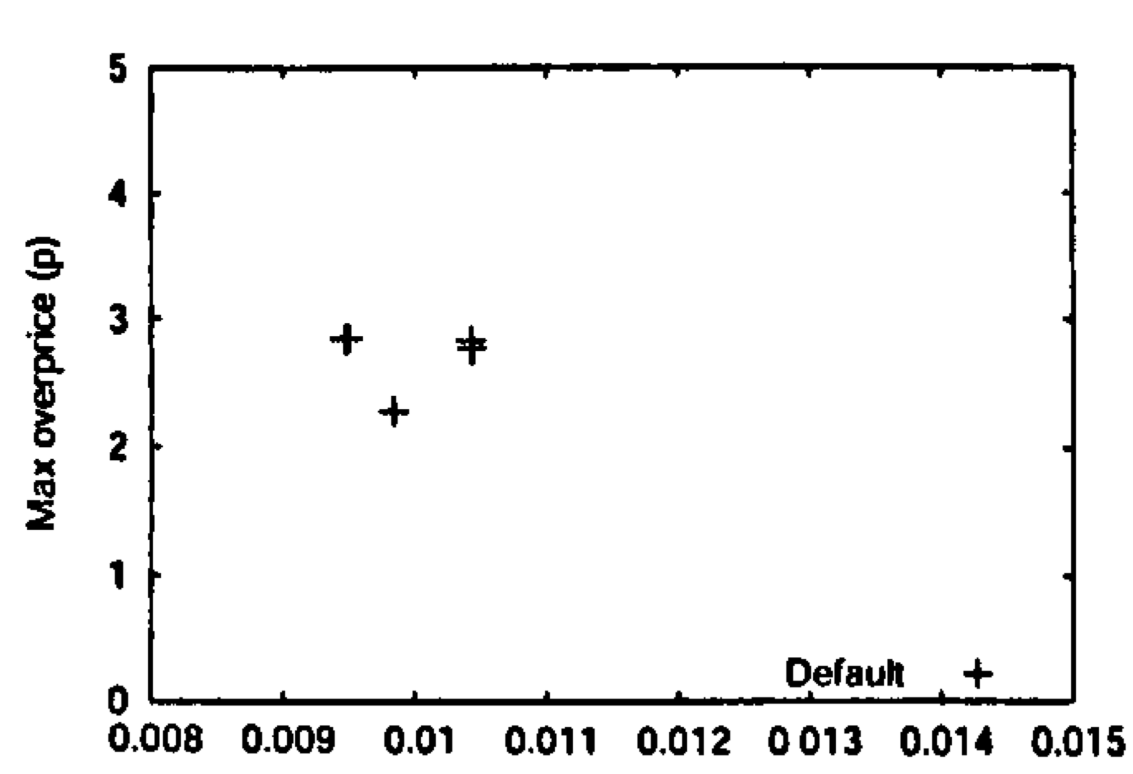
(d) *costToProduce*



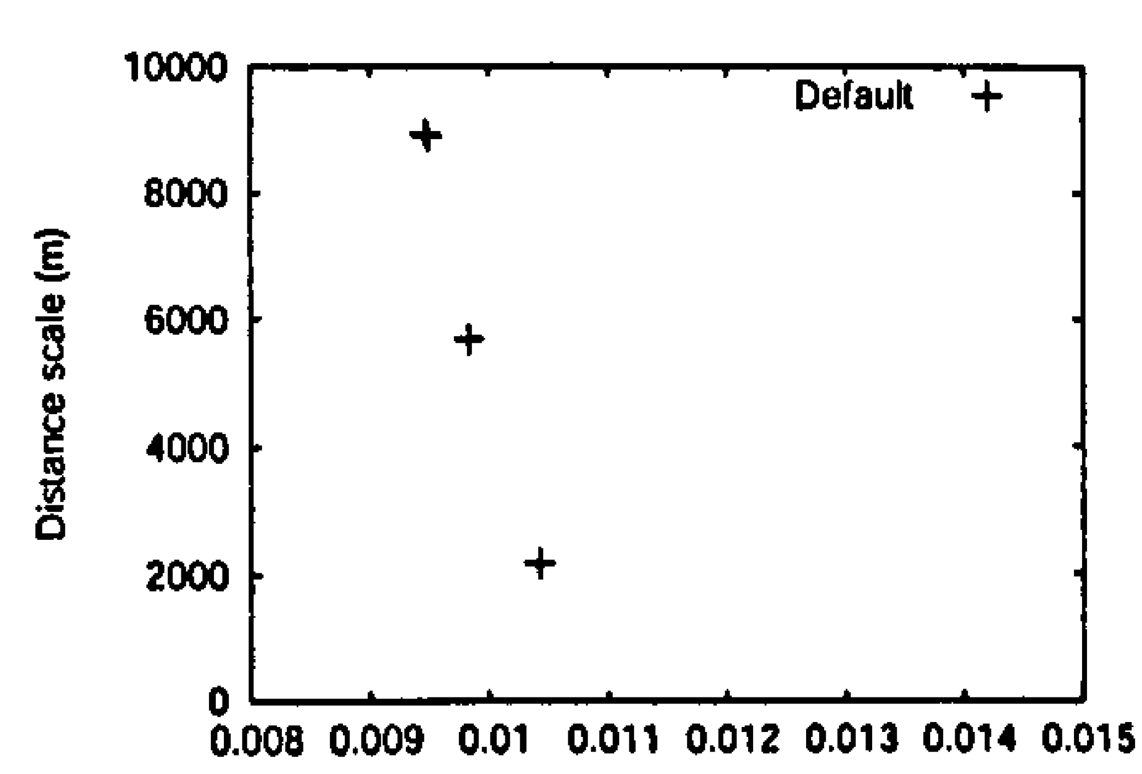
(e) *noChangeProfit*



(f) *undercut*



(g) *overprice*



(h) *neighbourhood*

Figure E.3: Variation in model parameters plotted against SRMSE for simulations with different GA control parameters. The experiments (described in §9.14.1) are using the hybrid model to develop rules using the second half of the data set starting on August 19th.