USING NETWORKS OF JOURNEYS TO IMPROVE A PETROL MARKET MODEL

Alison Heppenstall[1]*, Andrew Evans[1], David O'Sullivan[2], James MacGill[2] and Mark Birkin[1]
[1]University of Leeds. [2]Pennslyvania State University.

A hybrid-spatial interaction model (SIM) was presented in Heppenstall et al. (2005). Despite producing promising results, the model was limited by the assumption that consumers would only buy petrol from the wards that they lived in. Research undertaken by Ning and Haining (2003) showed that the majority of consumers purchased petrol on their way to work or on shopping trips. A network model is presented here that uses journey to work data to redistribute the population around the study area. The results show that this significantly improves the results of the model.

Keywords: Agents, routing, spatial interaction models.

1 INTRODUCTION

The petrol price market is a highly competitive market. 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 11707 sites retailing over 36 billion litres of motor fuel per annum in the UK. This equates to an average of approximately 1,350 litres of fuel consumed by each vehicle per year. Consumers are becoming ever more aware of petrol prices; internet sites in the UK, such as the Automobile Association site, enable consumers to have an almost perfect knowledge of prices within their area. 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. This sensitivity and knowledge has created both a highly competitive and rapidly responding market, with organisations employing various strategies to maximise profits.

Typically, the models developed by researchers to represent relationships within the petrol market are empirically based. They suffer from a number of problems, chiefly: the parameters are all on the same scale (behaviours executed at the *micro* level are not tied to *global* level variables like oil prices); the parameters are often difficult to estimate and lack realism; very little, if any, account of any geographical effects is taken, and, finally, mathematical models by their nature only consider quantitative parameters and therefore miss out on qualitative, behavioural information. Work undertaken by Heppenstall et al.

*School of Geography, University of Leeds, Leeds, LS2 9JT, UK. E-mail: a.heppenstall@geog.leeds.ac.uk. *Acknowledgements:* The work reported within this paper is funded by the ESRC and G-Map.

(2005) and Heppenstall et al. (2004) rectified several of these problems by creating a hybrid Spatial Interaction (SIM)-Agent model that worked at the scale of individual consumers and petrol stations, taking into account both quantitative and global parameters. The results of this hybrid SIM-Agent model were very promising. However, the main criticism was that consumers every day journey's were not taken into account. Instead they were assumed to remain in the ward in which they lived and buy petrol from the nearest station. Within this paper, we present a simple network model that is based on journey to work data. This model distributes the population around the study area thereby providing the consumer a choice of stations to purchase petrol from on their journey.

2 HYBRID SIM-AGENT MODEL

A hybrid Spatial Interaction Model (SIM) - Agent system was developed (see Heppenstall et al., 2005, for further details) in which customers travel to petrol stations depending on the distance to them and their prices. In brief, individual petrol stations were created as agent-objects and supplied with knowledge of their initial starting price, production costs and the prices of those stations within their neighbourhood. Rules were either assigned to individual stations, multiple stations, or simply given identically to all the stations. The rules were based on industry knowledge and implemented after experimentation with differing parameters. The SIM brought consumer behaviour to the agent model by calculating the number of people travelling to multiple locations, based on levels on the attractiveness of price and distance costs.

Interactions between the Agent model and Spatial Interaction Model can be summarised as:

1. The agent model and population are initialised with the real data.
2. The fuel price and station location data is passed to the SIM.
3. The SIM calculates fuel sales for each station and passes the information to them.
4. The stations use this information to choose the rules to implement based on their profitability.
5. Each agent calculates its new price and the simulation returns to step 3 until equilibrium or a set time limit is reached.

The data used for the SIM calculations was primarily drawn from the 1991 UK census. The customer locations were based on ward level data, i.e. number of people living within a ward. Throughout experimentation, these customers remained within their respective wards, buying petrol from the nearest competitive stations.

The consumers remained stationary within their respective wards throughout experimentation travelling buying petrol from the nearest

3 Criticisms/Limitation

The main criticism with this approach is the assumption that consumers buy from the ward that they live in. Research undertaken by Ning and Haining (2003) suggests that this is an unrealistic assumption. Part of Ning and Haining's research involved surveying households about their petrol buying habits. Responses showed that petrol is 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 take this type of behaviour into account, the population needs to be redistributed around the study area. A simple approach is to construct a network model that redistributes the population around the study area based on journeys made rather than just where consumers live.

4 DATA

The data used for redistributing the population was obtained from WICID[1]. It is drawn from a 10% sample and is non-cascading (i.e. only the origin i and destination j of the consumers are known). There are two problems with data of this nature. Firstly, 10% is not representative of the population, but will give an indication of population movements. Secondly, the assumption has to be made that, in the absence of any other information, consumers will take the shortest path between i and j. This is a classical *shortest path problem.*

4.1 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 4.1 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.

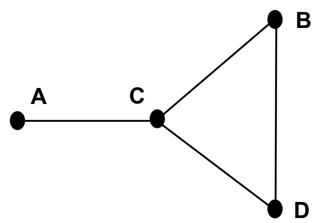

Figure 1: A Typical Graph (After O'Sullivan (2000))

.

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 ($f : E- > \mathbb{R}$)), and given further two elements n, n' of N, find a path P from n to n', so that

$$\sum_{p \in P} f(p) \tag{1}$$

[1]A database of travel to work data derived from the UK census available at http://census.ac.uk/cids/

is minimal among all paths connecting n to n'.

One of the most important algorithms for solving this problem is Dijkstra's algorithm (Dijkstra, 1959).

4.2 DIJKSTRA'S ALGORITHM

Dijkstra's algorithm solves the shortest path problem for the case where the graph is positive weighted and connected. 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. These conditions have to be satisfied as the algorithm assumes a building up of path lengths as it proceeds.

The algorithm 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 organized 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. 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.

4.3 Nodes and Edges

To re-distribute the population around West Yorkshire using Dijkstra's algorithm, nodes and edges have to be established. As the initial data is aggregated into census areas, the nodes will need to be located within these boundaries. The two choices available for use are wards[2] and enumeration districts (ED)[3]. Ideally, the smaller of the two, EDs, would be used. However, use of ward centroids would make the construction of a network model simpler. A compromise that draws the benefits of using the smaller EDs and enables the

[2]Wards have approximately 15,000 people in West Yorkshire. This amount varies across the UK.

[3]These are composed of 200 households (approximately 450 people).

use of wards is a population weighted centroid. Each ward is divided into its constituent EDs. 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 4.3). 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 centroid. This is expected as urban wards are geographically small and densely populated.

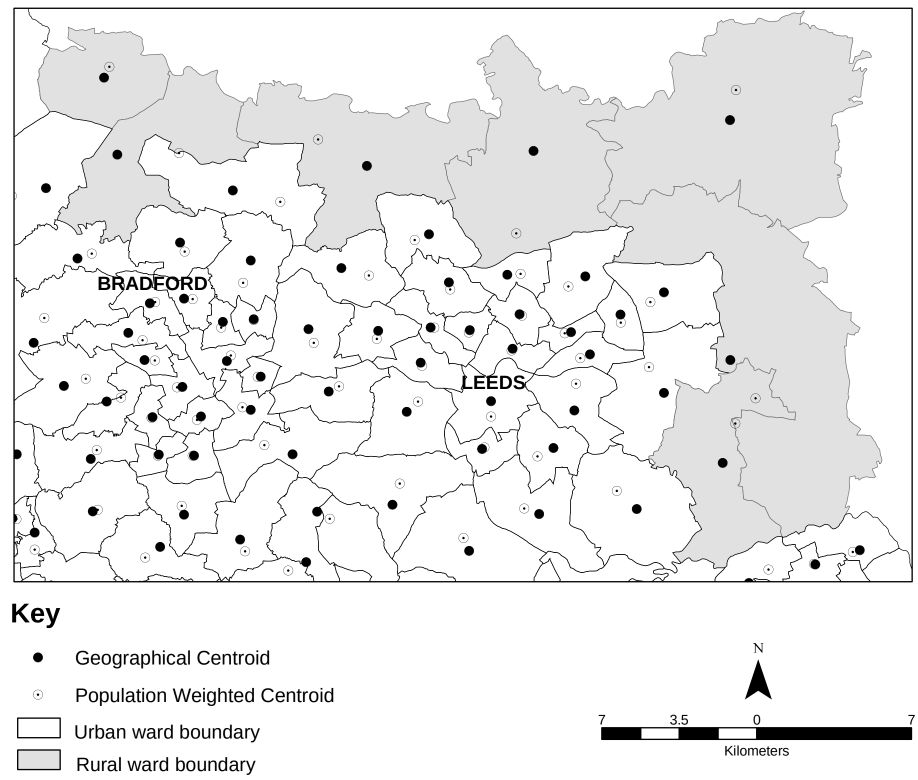

Figure 2: Location Of The Geographical And Population Weighted Centroids Within The Wards Of The Study Area.

The nodes can therefore be represented by these population weighted centroids and connections between neighbouring wards can be represented by weighted edges (Figure 4.3). The weights represent the distance between neighbouring nodes.

5 RUNNING THE NETWORK MODEL

The basic network model is run before any of the interactions between the agent and SIM take place. 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 start ward who take that route equally among each ward the route passes through. This creates a re-distributed population surface that is fed to the SIM in Stage 3 (see the

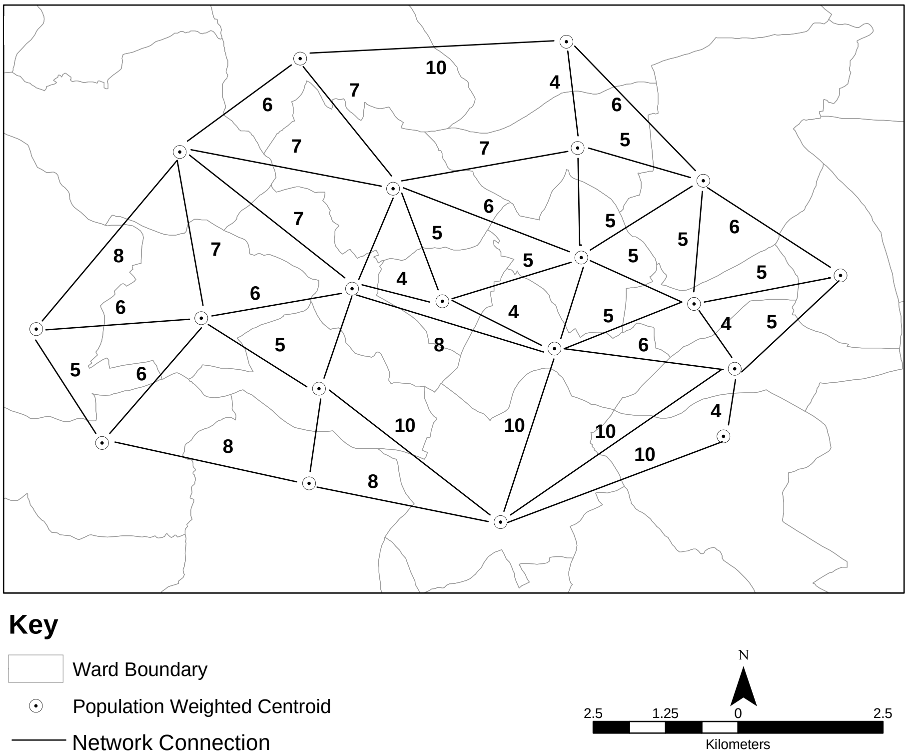

Figure 3: Example Of A Weighted Graph Network.

interactions detailed between the agent and SIM in §2). The network model is not used again during the course of the simulation.

5.1 Re-distribution of Cars

What effect has running this network model had on the performance of the hybrid SIM-Agent model? The distribution of car density before and after the application of the networking was calculated (Figure 5.1). Figure 5.1 (a) clearly shows that before the networking was applied, high densities of cars were found predominately in the centre of the study area. This area covers the West Yorkshire connurbation of Leeds, Bradford, Wakefield, Huddersfield and Halifax. In the suburban areas of Leeds and Bradford, the density of cars is at its highest. After the networking has been applied (Figure 5.1 (b)), all of the cities are characterised by having a higher density of cars.

6 APPLICATION TO REAL DATA

6.1 Case Study 1: West Yorkshire

How will this redistribution of population affect the results of the hybrid-network model? Will it improve on the performance of the hybrid-SIM Agent model? To provide a fair comparison, 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 models were run to equilibrium. The model parameters used are summarised in the Appendix.

Quantifying the degree of these differences between the results of the two models can

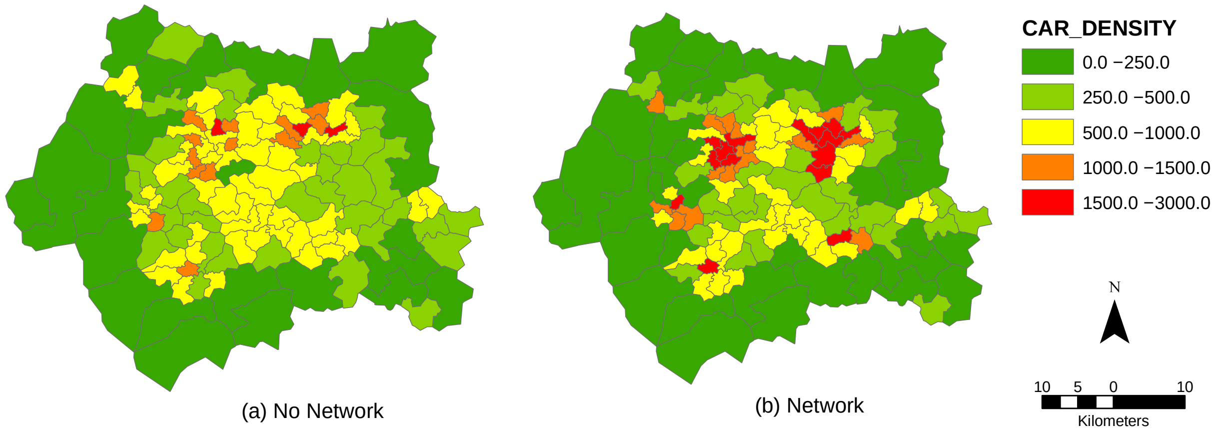

Figure 4: Distribution Of Car Density Before (a) And After (b) Application Of The Networking.

be achieved by using various statistical techniques such as mean and standard deviation (SD) of errors across the region. Figure 6.1 shows that the mean price difference is almost identical for the hybrid SIM-Agent and network hybrid SIM-Agent model. The standard deviation of the hybrid SIM-Agent model is marginally larger than that of the network hybrid SIM-Agent model. However, both model results mirror the pattern of the real data differences almost perfectly.

A better method of assessing the performance of the models is by testing their ability to recreate spatial variations. Figure 6.1 shows the interpolated price surfaces of both models after 10 days. The hybrid SIM-Agent model (Figure 6.1 (b)) is beginning to reproduce some price variations similar to those observed in the real data (Figure 6.1 (a)), for example between high priced rural areas and cheaper urban areas. However, the network hybrid SIM-Agent model (Figure 6.1 (d)) improved on the hybrid SIM-Agent model performance, reproducing a more representative price differentiation within the rural and urban areas. However, neither the hybrid and 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 can be accounted for. These simulations were initialised with each petrol station set to 71p. Over the course of 10 days, it would be almost impossible for a 6p variation in price to be produced with the current rule set and strategies. 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 6.1). This is reflected statistically, with the network hybrid model accurately modelling the prices over a 10-day period.

6.2 Case Study 2: Yorkshire Region

Using a larger geographical region will test the ability of the network hybrid SIM-Agent model to recreate rural-urban patterns on a greater geographical scale. The Yorkshire region (composed of North, West and South Yorkshire) contains a mixed variety of geographical characteristics. West Yorkshire is an urban conurbation surrounded by rural areas whilst North Yorkshire is a much larger, predominately rural county. South York-

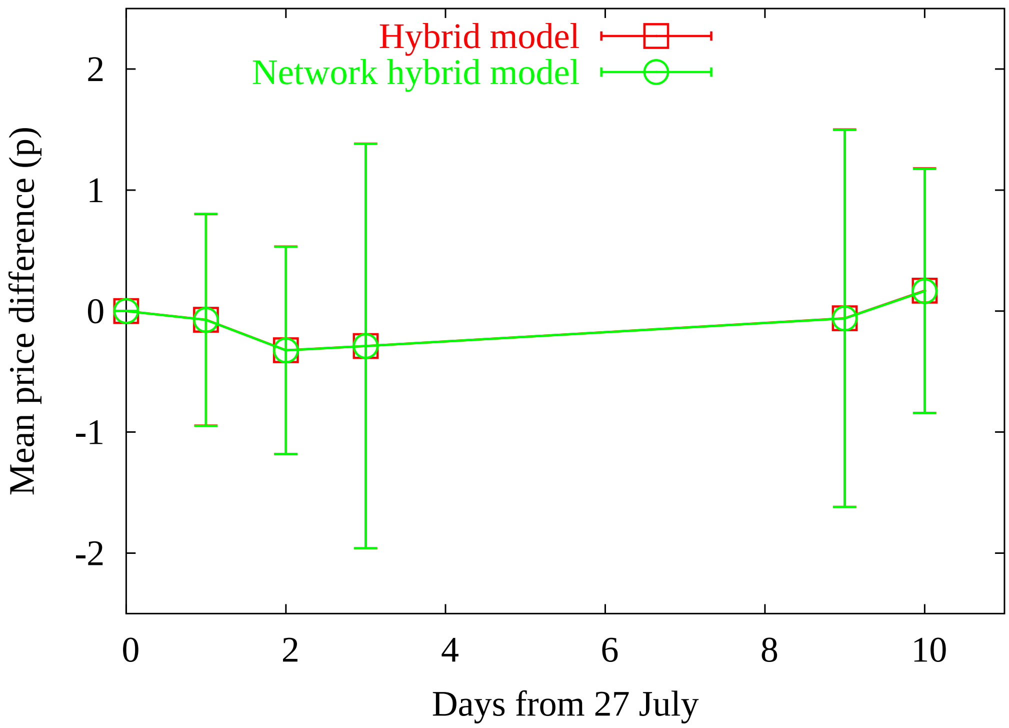

Figure 5: Comparison Of The Mean Price Difference Over Time (SD Is Indicated By The Error Bars) Between The Real (Day 10) And Model Data For The Hybrid SIM-Agent And Network Hybrid SIM-Agent Network Model For West Yorkshire (Model Initiated On July 27th). The Real Data Is Plotted For Comparison.

shire 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 with the constant price data (71p). One of the first observations that can be made from Figure 6.2 (b) is that the network hybrid model has captured the main rural-urban trends within the region. The rural area of North 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 intraurban variations in South Yorkshire have 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.

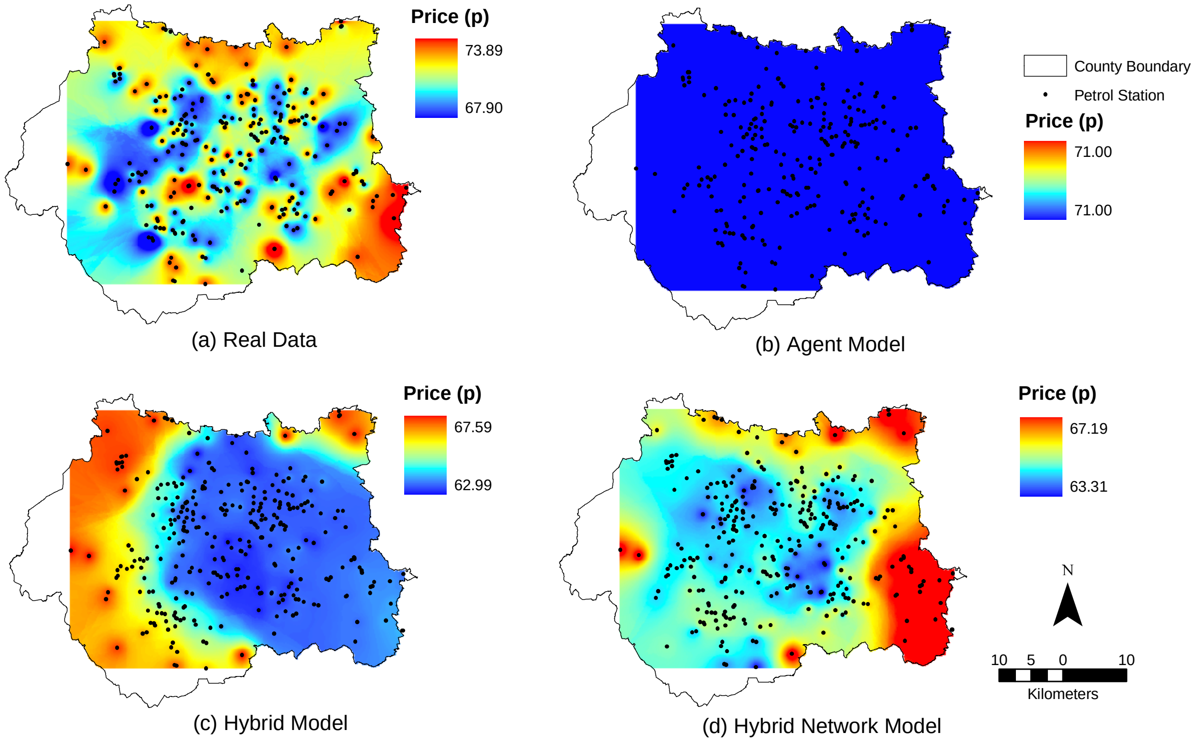

Figure 6: Price Distributions For The Various Models Ten Days After Runs Started With All Stations Initialised At 71 Pence: (b) Hybrid SIM-Agent Model And (c) Network Hybrid SIM-Agent Model. Data From Day Ten Of The Real Data set (a) Is Included For Comparison. Price Surfaces Are Interpolated From Point Station Prices For Ease Of Visualisation.

7 CONCLUSION

The re-distribution of the population using a simple network model has proved to be very effective within this application. The mean price difference of the two models were almost identical. However, the results produced were very good with both models accurately predicting real petrol prices to within 0.2p. The models' ability to recreate spatial patterns existant in the real data were also tested. The results showed that redistributing the population has made a significant difference to the model performance. The spatial patterns created by the network model improved on those produced by the hybrid SIM-Agent model. Areas of high and low pricing were picked out with greater accuracy.

There are improvements that could be made to the network model. For example, incorporating real drive times would ground the network model to the reality of the system. At present, the population is equally distributed throughout each of the wards that the route passes through. Basing this on distance through the ward could be a fairer criteria. This would mean that a greater number of people would be distributed within larger wards. However, the criticism with this approach is that larger wards tend to be rural and do not contain a large number of petrol stations. Ultimately, this is making a large assumption about consumer buying habits. This problem could be rectified by using detailed market research.

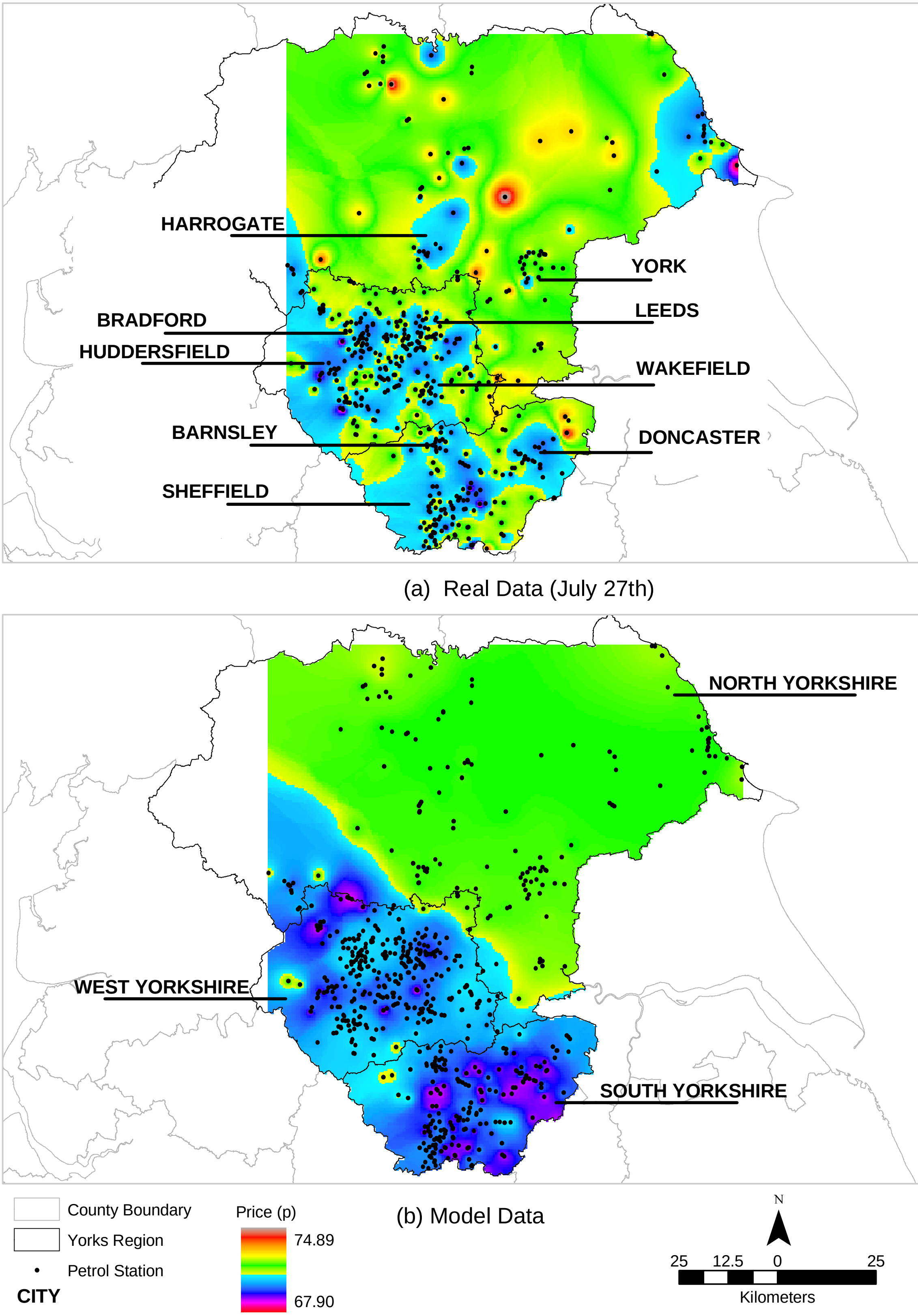

Figure 7: Comparison Of The (b) Network Hybrid Model's Performance On The Yorkshire Region With The Real Data (a) For July 27th.

Within this paper, we outlined a limitation with the hybrid-SIM model presented in Heppenstall et al. (2005). A simple population re-distribution model has been developed and linked to the hybrid-SIM model to rectify this. Results presented show that the inclusion of this model has made a significant to the improvement to the model performance.

8 APPENDIX

Table 8 contains the parameters used within experimentation with both the hybrid-SIM and network model.

Parameter	**Value**
β	0.0003
λ	0.7
Fixed Costs	£80
Cost to Produce	66p
Change in Profit	£40 - £50
Undercutting	1p
Overpricing	5p
Neighbourhood	5km

Table 1: Parameters Used Within Both The Hybrid-SIM Agent Model And Hybrid-SIM Network Model.

References

Dijkstra, E. (1959). Two problems in connexion with graphs. *Numerische Mathematik*, 1:269 – 271.

Heppenstall, A., Evans, A., and Birkin, M. (2005). A hybrid multi-agent/spatial interaction model system for petrol price setting. *Transactions in GIS*.

Heppenstall, A., Evans, A., Birkin, M., MacGill, J., and O'Sullivan, D. (2004). The use of hybrid agent-based systems to model the petrol market. *Proceedings of the Third International Workshop on Agent-based Approaches in Economic and Social Complex System.*, pages 285 – 292.

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.

O'Sullivan, D. (2000). *Graph-based Cellular Automaton Models of Urban Spatial Processes.* PhD thesis, Barlett School of Architecture and Planning, University College London.