Modelling accessibility via transportation networks based upon previous experience: a Geographically Weighted Regression approach

D. M. Mountain, J. L. Y. Tsui, J. F. Raper

giCentre, Dept Information Science, City University, London, EC1V 0HB, United Kingdom Telephone: +44 (0) 20 7040 4044 Fax: +44 (0) 20 7040 8584 dmm@soi.city.ac.uk, abbj518@city.ac.uk, raper@soi.city.ac.uk, http://www.gicentre.org

1. Introduction

1.1 Motivation

Much research is currently being undertaken in transportation to create more realistic models of transportation networks that account for the spatial heterogeneity in those transportation networks (Mark and Sadek, 2004), and also the different modes of transport that may be used. Simply assuming that users will travel at the maximum speed limit for a particular class of road fails to model local environment factors which may influence speed of travel. For example, the average speed of traffic on motorways is related to the volume of traffic, which tends to be higher around major conurbations. Real time traffic information is useful for relatively short journeys that are already underway, but given the ephemeral nature of this information, it is less useful for longer term journey planning.

This study pioneers the Geographically Weighted Regression (GWR) technique (Fotheringham et al., 2002) for modelling realistic transportation network conditions based upon the long-term previous experience of users of that network. It is hoped that this technique could be an effective approach to making longer term predictions of accessibility via transportation networks, which can account for spatial heterogeneity, temporal variation on a number of scales, and the particular characteristics of specific modes of transport.

1.2 Aims and objectives

The aim of this paper is to investigate the potential of the Geographically Weighted Regression (GWR) technique for modelling *realistic* road network characteristics based upon the previous experience of road users. The scope of this study is limited to cycling data collected in Essex, with the intention of examining how *suitable* GWR is to modeling accessibility via transportation networks.

The specific objectives of this study are:

• to acquire network data that models the road transportation network in the study area;

- to investigate how Geographically Weighted Regression can be applied to model accessibility on transportation networks;
- to consider the factors which influence speed of movement on road networks;
- to use the Geographically Weighted Regression technique to generate a model of accessibility on a transportation network based upon the experience of a particular type of road user and;
- to use the network for routing and comparing the results to existing (commercial) routing services.

1.3 Previous research

Geographically Weighted Regression (GWR) is statistical technique which aims to account for the spatial heterogeneity that may exist in regressed relationships (Yu, 2006). It allows the relationships between variables in a regression equation to vary over space by making local parameters estimates for a regression model. The GWR approach has been employed in research on housing (Yu, 2006), rainfall (Brunsdon et al., 2001), population (Yu and Wu, 2004) and other areas, however has had limited application within the field of transportation research. Zhao and Park (2004) have applied the GWR technique to estimate traffic density, but not accessibility.

2. Method and Results

There are four stages to the methodology – input data, geographically weighted regression, network initialisation and routing analysis – as shown in figure 1. Each stage will be discussed in turn.

Figure 1: Methodology for creating transportation network models, based upon previously observed behaviour



2.1 Input Data

The two main data requirements are mobile trajectory data representing spatial behaviour, and transportation network data for a particular study area. The mobile trajectory data should be a large collection with a good coverage of the study area, should represent a single form of transport, and preferably be collected over a long period of time to prevent particular conditions on any one day influencing results. The mobile trajectory data used in this study represented cycling behaviour for a single individual (a keen cyclist in his 30's), collected over a two-year period. The study area was narrowed down to a region in Essex incorporating Bishop's Stortford, Braintree and Chelmsford where the mobile trajectory data was the road centre lines provided by the Ordnance Survey Integrated Transport Network (ITN) (Ordnance Survey, 2005). The final input dataset is a regular lattice of point locations covering the study area, which are the points at which the GWR parameter estimates will be made (Fotheringham et al., 2002). This produces a regular lattice of local GWR parameter estimates, which can easily be converted into continuous raster surfaces.



Figure 2: Input mobile trajectory data: cycling behaviour in Essex.

2.2 Geographically weighted regression

There were many candidate parameters for the regression equation used to predict the speed of movement in the study region. These included temporal parameters (time of day, day of week) network attributes (junction density, road class) and terrain (gradient). In order to keep the equation simple to aid understanding of the suitability of GWR in this exploratory study, only one attribute was used to model speed: edge length. Edge length this is the length of each road section in the network and is the inverse of junction density. The hypothesis was that speed would be faster on longer, unbroken sections of road (edges) with fewer junctions, than the shorter sections with higher junction density likely to be found in more built up areas.

The resulting regression equation for which parameter estimates needed to be found was:

$$speed = c + (m * edgeLength)$$
(1)

where c is the intercept (default speed) and m the parameter estimate for edge length.

2.2.1 Global regression results

First a global regression model was calculated, estimating the parameters c and m for equation 1 for all of the mobile trajectory data. The global regression equation for speed and network edge length for cycling data in Essex was found to be:

$$speed = 6.8 + (0.0006 * edgeLength)$$
 (2)

Where the units for speed are metres per second (m/s) and edge length is measured in metres (m). This suggests a default global speed (intercept) for this dataset of 6.8 m/s, and a positive relationship between road section length and speed: i.e. speed will be faster on longer sections of road. This fits with the hypothesis that higher junction density will result in slower speeds. This global regression model can be seen in figure 3.





Hence for low junction density (road sections of 100m or less), speed will be close to the default (around 6.8m/s). For unbroken road sections of 1000m, a noticeable increase in speed will be evident (7.4 m/s). Clearly for very long sections of road, this linear regression equation will result in unrealistic speeds for cyclists, however the longest edge length in the study area was 3,500m, which results in a realistic speed estimate of 8.9 m/s using equation 2.

2.2.2 Geographically weighted regression results

Next, local parameter estimates (c and m in equation 1) were calculated for each regression point in the regular lattice, using only the *local* mobile trajectory data. An adaptive kernel (Fotheringham et al., 2002) was applied that used the nearest 2000 data points for calculating local parameter estimates. These point lattices of parameter estimates were then transformed to surfaces using local interpolation.

Figure 4: Variation in intercept (default speed) over space (lighter indicates faster)



Considering the variation in the parameter estimates over the study region, the value of c (default speed) varies between 4.5 and 8.4 m/s. As figure 4 shows, the higher default speeds (shown lighter) tend to be found in rural areas, particularly noticeable between

Great Dunmow and Braintree, and between Chelmsford and Harlow. The (visited) urban centres and Chelmsford, Braintree, Great Dunmow and Bishop's Stortford are all associated with lower default speeds.

Figure 5: Variation in 'edge length' parameter estimate over space (lighter is higher). For higher values, road section length has greater influence. High : 0.00743



The 'edge length' parameter estimate varies between -0.0004 and 0.0073, a predominantly positive relationship, again supporting the hypothesis that longer edge lengths are associated with faster speeds. As shown in figure 5, this parameter estimate also varies over space and tends be larger in urban areas, such as Bishop's Stortford to the NW, and Chelmsford to the SE. (Despite the obvious peak around the urban area of Harlow, this location was not visited by the data collector and is likely to indicate an extrapolation artefact, hence is not considered here). This would suggest that edge length has a greater *influence* in urban areas than in rural areas. This urban-rural distinction fits with an observation of the relationship between junction density and speed for cyclists. Considering the urban case, ringroads - with fewer junctions - tend to be the fastest routes through urban areas and offer cyclists a quicker route than going through alternative

urban areas with a higher junction density and more traffic restrictions. In the rural case however, cyclists are unlikely to have to slow down for the small villages that they pass through, even though junction density increases in these village centres, relative to the surrounding countryside.

2.3 Network initialization

Given the parameter estimate surfaces shown above, a unique speed can be calculated for each edge in the network by entering the length of the edge into equation 1, and using the local parameter estimates for default speed (c) and edge length (m). This produces the speed network shown in figure 6. This network demonstrates a clear trend that urban areas tend to be associated with lower speeds (darker brown edges) and rural areas faster speeds (lighter edges). Knowing the length of each edge and the speed associated with it, a traversal time can be also be calculated for each edge which can be used as an impedance value for routing.





2.4 Routing analysis

The network above was used to calculate the fastest routes between two locations, and this result compared to a commercial routing service. Only one such route is described here. The test route was between Galleywood (Chelmsford) and Hatfield Broad Oak (Hatfield Heath). Figure 7a shows the route suggested from the speed network generated in section 2.3. Figure 7b shows the route suggested by Google maps route planner (Google, 2006). It can be seen that the route suggested by the speed network based upon the previous experience of the Essex cycling mobile trajectories takes a more direct route, heading SE on a minor A-road through the open countryside. It then uses the inner ring road to pass to the South of Chelmsford town centre. The Google route planner heads North initially to join a major A road, and sticks to major roads at the expense of the more direct route. This strategy of seeking out more major roads employed by the commercial service may be preferable for users of automated transport, but is unlikely to provide the fastest routes for cyclists.

Figure 7: Routing between Galleywood (Chelmsford) and Hatfield Broad Oak (Hatfield Heath)

a. Route suggested based upon GWR b. Route suggested by Google maps route speed network

planner (Google, 2006)



4. Discussion and Conclusions

This paper presents an exploratory study to apply the geographically weighted regression technique to calculating the speed of movement on a road transportation network based upon the previous experience of users of that network. The limited scope of the study, constrained to a region of Essex and based upon the experience of a single individual (a cyclist), means that care must be taken when interpreting these results, but nevertheless the study suggests that GWR may provide an effective approach to modelling the spatial heterogeneity associated with speed on transportation networks.

The regression model used is a trivially simplistic one, considering only edge length as a predictor for speed, but nevertheless, meaningful trends emerge in the resulting parameter estimate surfaces and resulting speed network. Future research intends to;

- integrate more prediction parameters (for example: road class; gradient; time of day; day of week)
- use a larger volume of mobile trajectory data, representing more transportation network users;
- model other modes of transportation, such as pedestrians and public and private automated transport.

5. Acknowledgements

The research presented in this paper was conducted within the LOCUS project funded by the EPSRC through the Pinpoint Faraday Partnership. The authors would also like to thank Jo Wood for contributing his GPS data to this research.

6. References

- Brunsdon C, McClatchey J and Unwin DJ (2001) Spatial Variations in the Average Rainfall-Altitude Relationship in Great Britain: An Approach using Geographically Weighted Regression altitude; precipitation; regression modelling; Great Britain. *International Journal of Climatology*, **21**, **4**, 455-466.
- Fotheringham AS, Brunsdon C and Charlton M (2002) Geographically Weighted Regression: the analysis of spatially varying relationships, Wiley, London.
- Google (2006) Google local http://local.google.co.uk/ (visited 10 Dec 2006)
- Mark CD and Sadek AW (2004) Learning Systems for Predicting Experiential Travel Times in the Presence of Incidents: Insights and Lessons Learned. *Transportation Research Record*, 51-58.
- Ordnance Survey (2005) OS MasterMap Integrated Transport Network (ITN) Layer <u>http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/itn/index.html</u> (visited 16 Nov 2005)
- Yu D and Wu C (2004) Understanding Population Segregation from Landsat ETM+ Imagery: A Geographically Weighted Regression Approach. *GIScience and Remote Sensing*, **41**, **3**, 187-206.
- Yu DL (2006) Spatially varying development mechanisms in the Greater Beijing Area: a geographically weighted regression investigation. *Annals of Regional Science*, 40, 1, 173-190.
- Zhao F and Park N (2004) Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. *Transportation Research Record*, 99-107.

Biographies

Dr David Mountain is a research fellow on the LOCUS project and lecturer in Geographic Information within the Department of Information Science at City University, London.

Ms Jaclyn Lai Ying Tsui is studying for an MSc in Information Science within the Department of Information Science at City University, London.

Prof Jonathan Raper is a Professor of Geographic Information Science within the Department of Information Science at City University, London.