SPATIAL NETWORK ANALYSIS AS A LOW COST LAND USE-TRANSPORT MODEL OF CITY WIDE CYCLIST FLOWS

Dr Crispin Cooper
Sustainable Places Research Institute, Cardiff University

1. INTRODUCTION

Despite numerous studies of cycling mode choice (Wardman, Tight, and Page 2007; Parkin, Wardman, and Page 2007; Winters et al. 2013; Ewing et al. 2014) and route choice (Broach, Dill, and Gliebe 2012; Ehrgott et al. 2012) none so far have been turned into a general purpose tool for modelling cycling in the same way as the four step model (Ortúzar and Willumsen 2011) is used to model motor transport. Such a model would have applications in estimating change to mode choice from proposed cycle infrastructure - the key economic justification for investment – as well as highlighting hotspots where new infrastructure would be useful in the first place, assisting with option selection, and illustrating how proposed infrastructure fits in to the wider network. This paper presents such a model which is currently being used to inform production of a city-wide Integrated Network Map (essentially a forward plan for cycle infrastructure) mandated by the Wales Active Travel Act (2013).

The lack of prior models of this type is understandable in light of the fact that motor transport models typically work at transport analysis zone (TAZ) level and thus miss small features that can influence cyclist decision making; indeed a huge proportion of cycling trips are intra- rather than inter-zonal. The answer is not simply to make smaller models, however; instead, equally wide scale models are needed but with a large increase in resolution. Also, an increase in resolution of the *vehicle* model is needed to inform the cycling model. This would entail a greater cost in calibration of the four step model, which is typically too expensive to apply to small scale cycling infrastructure projects in any case.

In the absence of four step models, predictions of cycling have relied on applying an exogenous growth factor to current behaviour (Schwartz et al. 1999), or linking investment in infrastructure to growth in cycling at a course spatial scale (Parkin, Wardman, and Page 2007). The can predict mode choice based on a wide range of infrastructure investment and sociodemographic factors, but the model is not sensitive to the precise location of infrastructure, thus cannot be used to determine the optimum location for it.

This paper presents a high resolution, wide scale model of cyclist behaviour based on spatial network analysis. The fundamental premise of the model is that accessibility itself shapes land use, ultimately creating origins and destinations which cause flows (Chiaradia, Cooper, and Wedderburn 2014).

Spatial Network Analysis (SpNA) has been applied to vehicle, pedestrian (Hillier and Iida 2005; Turner 2007; Cooper 2015) and cycling problems before (Raford, Chiaradia, and Gil 2007; Manum and Nordstrom 2013; Law, Sakr, and Martinez 2014) but without the microeconomic behavioural foundations used here: use of a cycling-specific distance metric both for defining network radius and route choice. Also novel in the current study is the multivariate approach, which can be interpreted as multiple agent models combined through machine learning to fit measured cyclist count data on the network. The multivariate approach is calibrated to match varying individual preferences of cyclists. The latter is acknowledged as a key issue in uptake of cycling as some cyclists are more confident in motor vehicle traffic than others. The model is implemented using the publicly available Spatial Design Network Analysis (sDNA+) software (Cooper, Chiaradia, and Webster 2011) which functions as either a QGIS or ArcGIS plug-in.

2 METHODOLOGY

2.1 Definition of distance

The first step to producing a behaviourally accurate cyclist model using spatial network analysis is to determine an appropriate definition of distance through the network. The metric used is based on a subset of factors identified in the cyclist route choice study of Broach et al (2012), the choice of which is informed by availability of the relevant data. Creating a model sensitive to motor vehicle traffic is considered essential, however, so a submodel is used to predict motor vehicle flows. The definition of distance applicable to cyclists is then determined by a combination of distance, straightness, slope and motor vehicle traffic:

$$cyclist\ distance = Euclidean\ distance\ \times slopefac^{s}\ \times trafficfac^{t} + angular\ distance\ \times \frac{67.2}{90} \times a \tag{1}$$

$$slopefac = \begin{matrix} 1.000 & if\ slope < 2\% \\ 1.371 & if\ 2\% < slope < 4\% \\ 2.203 & if\ 4\% < slope < 6\% \\ 4.239 & if\ slope > 6\% \end{matrix} \tag{2}$$

$$trafficfac = 0.84\ e^{\frac{AADT}{1000}} \tag{3}$$

in which AADT is annual average daily (vehicle) traffic. The structure of this formula and its constants are chosen to match Broach et al, leaving room for calibration by changing *s*, *t* and *a*; with the exponential form of Equation 3 derived by fitting a curve to Broach's fixed distance bands in order to achieve better control over calibration. To match the original study we would set

$$\begin{aligned} a &= 1 \\ s &= 1 \\ t &= 0.05 \end{aligned} \tag{4}$$

Based on previous work we found the following parameter values best fit the Cardiff data in a homogenous model:

$$\begin{aligned} a &= 0.2 \\ s &= 2 \\ t &= 0.04 \end{aligned} \tag{5}$$

All cyclist distances are measured as round trip distances using the same route for the outward and return journey, as a cyclist who goes downhill knows they must later climb back up again, and this will affect their decision to cycle.

2.2 Definition of Betweenness

Having defined distance appropriately, we apply it to the SpNA concept of Link Weighted Betweenness. Betweenness simulates shortest-path trips from everywhere to everywhere, constrained by a certain maximum distance, but we introduce a minimum distance also, so as to simulate each distance band separately and thus reduce multicollinearity in the multivariate models of flow. Distances are defined not in Euclidean terms but in terms of Equation 1.

From a transport simulation perspective, the noteworthy point is that there is no dependence on origins and destinations, or explicit inclusion of land use. Implicit in the SpNA model is an assumption of efficient network use: that the quantity of origins and destinations correlates highly with the quantity of network built to serve them. This assumption has previously been tested in London with very high correlation observed (Chiaradia et al. 2012).

2.3 Distance decay model for flows and mode choice

Having defined distance and betweenness, we compute betweenness within a number of cycling distance bands. The distance bands chosen for cyclists are round trip distances of 3, 5, 8, 11, 15 and 20 'adjusted' km (as per Equation 1). Multi-band Betweenness is then combined using multivariate regression to fit link flow data, such that

$$flow\ on\ link = \beta_1 Bt_1 + \beta_2 Bt_2 + \cdots \tag{6}$$

where Bt_1, Bt_2 is the betweenness in distance bands 1, 2, etc; and the βs are regression coefficients. This is nonparametrically fitting a distance decay curve.

The cycling mode choice model is also based on distance decay. However, instead of computing betweenness, the variables used to predict mode choice are simple counts of network quantity (measured in number of links) within each distance band. Together, these form a multi-dimensional definition of accessibility.

The submodel used to compute vehicle flows to inform the cyclist model is an angular betweenness calculation similar to those used in Cooper (2015) but with a higher (calibrated) maximum trip distance.

2.4 Further models for cyclist flow

The interpretation thus far presented is that the distance decay model uses regression to perform nonparametric calibration of distance decay and economic scaling curves. However, it can also be interpreted as modelling different types of behaviour: trips of varying lengths from everywhere to everywhere. We now include (i) trips of varying lengths to the city centre and known cyclist recreational facilities, (ii) trips made by agents with varying aversion to motor vehicle traffic. The regression model thus chooses an appropriate 'balance' of behaviours that best explains the observed flows. Referring to Equation 1, we compute betweenness with t=0.06 and t=0.08 as well as t=0.04.

2.5 Data sources

Road network data is based on OpenStreetMap (2015), which at time of writing contains more information on traffic free cycle routes than any other publicly available routable network data (including commercial offerings, Lovelace 2015). The network data is prepared according to the instructions in the sDNA user manual, including planarization and use of a high cluster tolerance to correct errors (Cooper 2016).

Flow models are calibrated to measurements of cycle flows for 107 locations on roads (Department for Transport 2014) and 14 locations on traffic free paths (provided by Cardiff Council). Traffic free path counts are derived from electronic counters covering a three month period plus a year-round counter which is used to deduce a scaling factor to estimate average annual daily count. This differs from the on-road counting methodology (Department for Transport 2011) which is likely to under-count cyclists. In the final model therefore, a dummy variable is introduced to account for data source, to estimate the effect of differing count methodology between the data sets.

Mode choice models are calibrated to journey-to-work data (proportion of working population travelling by bicycle) for 1077 census Output Areas covering Cardiff (Office for National Statistics 2011).

2.6 Model fitting

All models are fitted using ridge regression (Tikhonov 1943; Amemiya 1985) and fit reported using generalized cross-validation. Regression is weighted to reflect the same balance of reducing absolute vs relative error as the commonly used GEH statistic. The open source sDNA Learn tool is used to perform the regression based on the glmnet package in R (Friedman, Hastie, and Tibshirani 2009).

3 RESULTS

The vehicle traffic sub-model showed optimum fit to the data with $r^2 = 0.81$ for a 28km (one-way) trip distance.

Table 1 Cross-validated fit for cyclist flow models

Flow Model	GEH-Weighted r^2, cross-validated
Multiple radius, medium traffic aversion only (t=0.04)	0.65
Multiple radius, trips to centre and recreation, medium traffic aversion only (t=0.04)	0.73
As above with mixed traffic aversion (t=0.04, 0.06, 0.08)	0.75
As above with dummy variable accounting for data source	0.78

Table 1 shows results for the cycling flow models. The best model has GEH<5 for 93% of data points, with mean GEH of 1.9. GEH is computed for the peak hour based on estimation that the peak carries 10% of daily flow.

For mode choice, the multivariate distance decay model has cross-validated r^2 with census travel to work data of 0.45. Note that numerous sociodemographic and promotional factors are known to affect mode choice (Parkin, Wardman, and Page 2007) so we should not expect to explain much more of the variance from network design alone.

An example of predicted cyclist flow is shown in Figure 1. The ability of the model to provide detailed spatial information on the links between infrastructure and cycling potential is illustrated in Figure 2, which shows potential increase in mode choice in the hypothetical situation that all routes were traffic-free. This is intended not as a realistic scenario, but an accessibility model which highlights trip endpoints and routes currently worst affected by motor vehicle traffic.

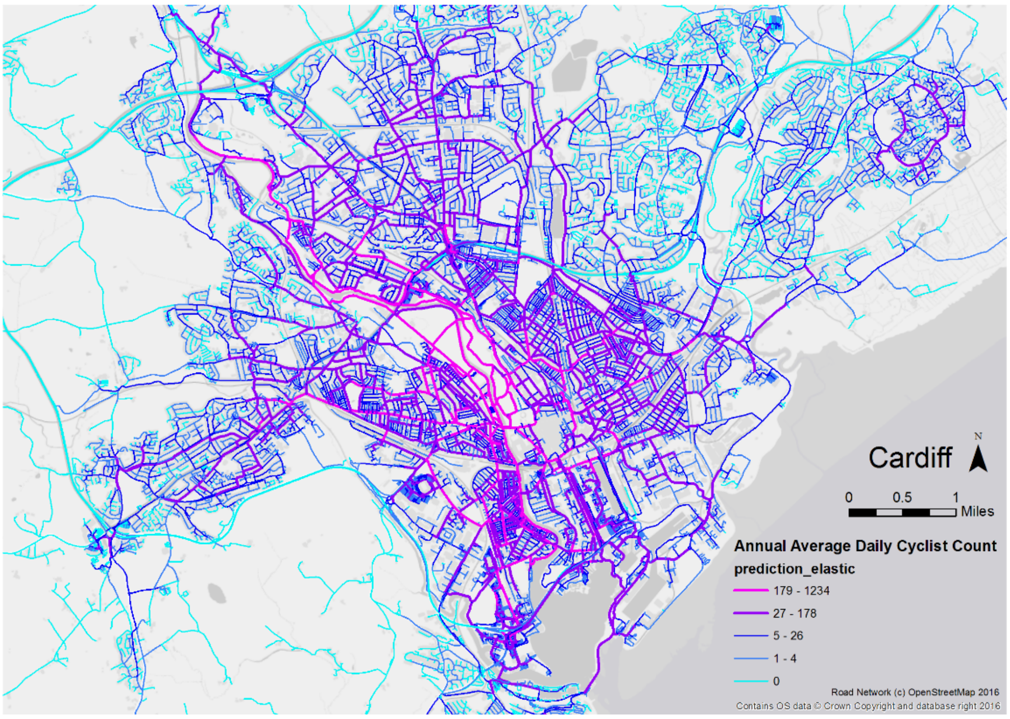

Figure 1 Predicted cyclist flows

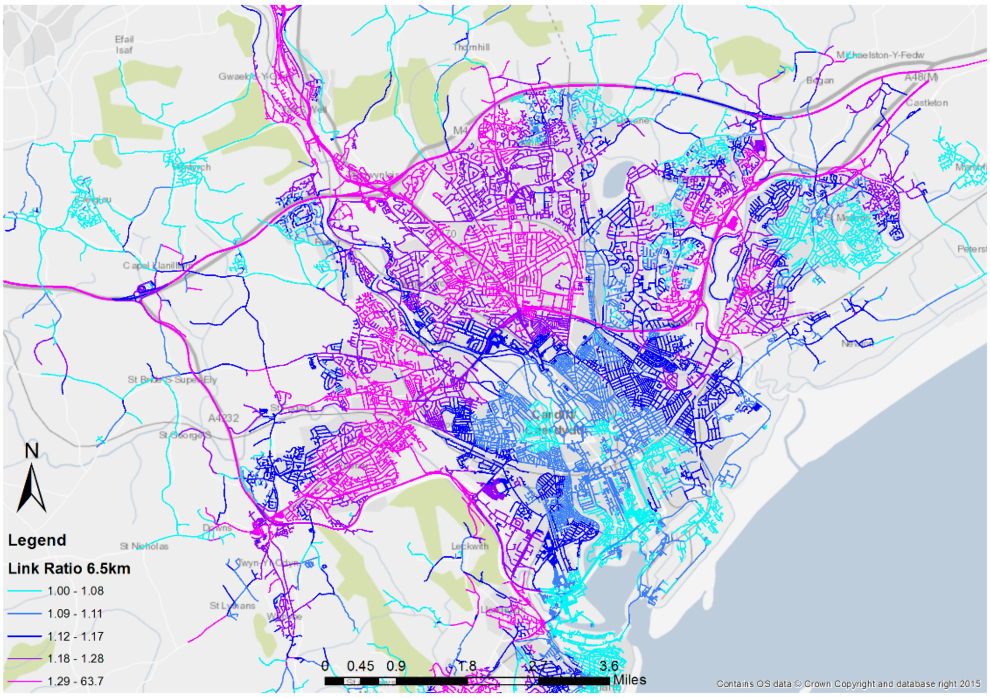

Figure 2 Potential for improvement by introducing traffic free routes

4 CONCLUSIONS

This paper has presented a range of options for modelling cycling which a modeller can choose from depending on their own problem constraints and available time. There are likewise a multitude of ways in which model outputs can be presented, including accessibility/scenario maps, to inform spatially sensitive models of cycling potential.

The main contribution is a spatially-detailed methodology for obtaining a cross-validated fit to existing cycle flow data on a city-wide scale. This is a step forward both for SpNA methodologies which do not normally use cross-validation, and transport methodologies which are not spatially detailed.

While spatial network analysis may be unfamiliar to those with a transport modelling background, the practical modelling process is relatively simple and inexpensive in terms of data collection. The sDNA Integral software is run a number of times to compute different agent behaviours, which are then calibrated against real data with sDNA Learn and extrapolated with sDNA Predict. The software can also interface with existing models through export of skim matrices to model accessibility at high resolution, and import of OD matrices to use in the assignment phase (Cooper 2016).

Finally, we have shown that (at least in the UK) existing targets for the GEH statistic are too easily achieved with the small flow numbers present in cycling models. While we used the scale-free r^2 statistic to evaluate overall model performance, the best choice of metric to describe fit of individual data points remains an open question.

REFERENCES

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge, Mass: Harvard University Press.

Broach, Joseph, Jennifer Dill, and John Gliebe. 2012. “Where Do Cyclists Ride? A Route Choice Model Developed with Revealed Preference GPS Data.” *Transportation Research Part A: Policy and Practice* 46 (10): 1730–40. doi:10.1016/j.tra.2012.07.005.

Chiaradia, Alain, Crispin H. V. Cooper, and Martin Wedderburn. 2014. “Network Geography and Accessibility.” In *Proceedings of 12th Transport Practitioners’ Meeting*. London: PTRC.

Chiaradia, Alain, Bill Hillier, Christian Schwander, and Martin Wedderburn. 2012. “Compositional and Urban Form Effects in Centres in Greater London.” *Urban Design and Planning - Proceedings of the ICE* 165: 21–42.

Cooper, Crispin H. V. 2015. “Spatial Localization of Closeness and Betweenness Measures: A Self-Contradictory but Useful Form of Network Analysis.” *International Journal of Geographical Information Science*, March, 1–17. doi:10.1080/13658816.2015.1018834.

———. 2016. “Spatial Design Network Analysis (sDNA) Version 3.4 Manual.” http://www.cardiff.ac.uk/sdna/software/documentation.

Cooper, Crispin H. V., Alain Chiaradia, and Chris Webster. 2011. “Spatial Design Network Analysis (sDNA).” www.cardiff.ac.uk/sdna.

Department for Transport. 2011. “Road Traffic Estimates Methodology Note.” UK https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/230528/annual-methodology-note.pdf.

———. 2014. “Traffic Count Data.” http://www.dft.gov.uk/traffic-counts/.

Ehrgott, Matthias, Judith Y.T. Wang, Andrea Raith, and Chris van Houtte. 2012. “A Bi-Objective Cyclist Route Choice Model.” *Transportation Research Part A: Policy and Practice* 46 (4): 652–63. doi:10.1016/j.tra.2011.11.015.

Ewing, Reid, Guang Tian, J. P. Goates, Ming Zhang, Michael J. Greenwald, Alex Joyce, John Kircher, and William Greene. 2014. “Varying Influences of the Built Environment on Household Travel in 15 Diverse Regions of the United States.” *Urban Studies*, December, 42098014560991. doi:10.1177/0042098014560991.

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2009. “Glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models.” *R Package Version* 1.

Hillier, Bill, and Shinichi Iida. 2005. “Network and Psychological Effects in Urban Movement.” In *Spatial Information Theory*, edited by Anthony G. Cohn and David M. Mark, 475–90. Lecture Notes in Computer Science 3693. Springer Berlin Heidelberg. http://link.springer.com/chapter/10.1007/11556114_30.

Law, Stephen, Fernanda Lima Sakr, and Max Martinez. 2014. “Measuring the Changes in Aggregate Cycling Patterns between 2003 and 2012 from a Space Syntax Perspective.” *Behavioral Sciences* 4 (3): 278–300. doi:10.3390/bs4030278.

Lovelace, Robin. 2015. “Crowd Sourced vs Centralised Data for Transport Planning: A Case Study of Bicycle Path Data in the UK.” In *GIS Research UK (GISRUK)*. Leeds, http://leeds.gisruk.org/abstracts/GISRUK2015_submission_71.pdf.

Manum, Bendik, and Tobias Nordstrom. 2013. “Integrating Bicycle Network Analysis in Urban Design: Improving Bikeability in Trondheim by Combining Space Syntax and GIS-Methods Using the Place Syntax Tool.” In *Proceedings of the Ninth International Space Syntax Symposium*. Seoul: Sejong University.

Office for National Statistics. 2011. “Method of Travel to Work.” QS701EW. https://www.nomisweb.co.uk/census/2011/qs701ew.

OpenStreetMap contributors. 2015. “Open Street Map.”

Ortúzar, Juan de Dios, and Luis G. Willumsen. 2011. *Modelling Transport*. 4th Edition edition. Chichester, West Sussex, United Kingdom: Wiley-Blackwell.

Parkin, John, Mark Wardman, and Matthew Page. 2007. “Estimation of the Determinants of Bicycle Mode Share for the Journey to Work Using Census Data.” *Transportation* 35 (1): 93–109. doi:10.1007/s11116-007-9137-5.

Raford, N., Alain Chiaradia, and J. Gil. 2007. “Space Syntax: The Role of Urban Form in Cyclist Route Choice in Central London.” In *TRB (Transportation Research Record) 86th Annual Meeting Compendium of Papers CD-ROM*, 07–2738.

Washington, DC: Transportation Research Board. http://escholarship.org/uc/item/8qz8m4fz.

Schwartz, W. L., C. D. Porter, G. C. Payne, J. H. Suhrbier, P. C. Moe, Wilkinson Iii, and W. L. 1999. “Guidebook on Methods to Estimate Non-Motorized Travel: Supporting Documentation.” http://trid.trb.org/view.aspx?id=503335. http://trid.trb.org/view.aspx?id=503335.

Tikhonov, A. N. 1943. “Об Устойчивости Обратных Задач.” *Doklady Akademii Nauk SSSR* 39 (5): 195–98.

Turner, Alasdair. 2007. “From Axial to Road-Centre Lines: A New Representation for Space Syntax and a New Model of Route Choice for Transport Network Analysis.” *Environment and Planning B: Planning and Design* 34 (3): 539–555. http://eprints.ucl.ac.uk/2092.

Wardman, Mark, Miles Tight, and Matthew Page. 2007. “Factors Influencing the Propensity to Cycle to Work.” *Transportation Research Part A: Policy and Practice* 41 (4): 339–50. doi:10.1016/j.tra.2006.09.011.

Winters, Meghan, Michael Brauer, Eleanor M. Setton, and Kay Teschke. 2013. “Mapping Bikeability: A Spatial Tool to Support Sustainable Travel.” *Environment and Planning B: Planning and Design* 40 (5): 865 – 883. doi:10.1068/b38185.