



# Using spatial network analysis to model pedal cycle flows, risk and mode choice

Crispin H.V. Cooper

Sustainable Places Research Institute, Cardiff University, 33 Park Place, Cardiff CF10 3BA, United Kingdom

## ARTICLE INFO

### Article history:

Received 19 June 2015

Received in revised form 28 November 2016

Accepted 4 December 2016

Available online 10 December 2016

### Keywords:

Spatial network analysis

Cycling

Modelling

Gis

## ABSTRACT

Spatial network analysis (SpNA) provides a promising alternative to traditional transport models for the modelling of active travel, because walking and cycling behaviour is influenced by features smaller than the scale of zones in a traditional model. There is currently a need for link-level, city wide modelling of cycling, both to ensure the needs of existing cyclists are catered for in planning, and to model the effects of changing infrastructure in shaping cyclist behaviour. Existing SpNA models treat cyclists and car drivers as if they make navigational decisions in a similar way, which in reality is not the case.

This paper presents an SpNA model using hybrid betweenness, which fits cyclist flows in Cardiff, Wales using distance, angular distance, motor vehicle traffic and slope as predictors of route choice. SpNA betweenness is also shown to implicitly capture the effect of urban density on mode choice. As it handles route finding decisions of drivers and cyclists separately, the model presented is also applicable to road safety models examining the interaction between the two classes of road user. The model has low cost of data collection and is reproducible using publicly available network analysis software and open mapping data. Further avenues for modelling the effect of infrastructure on cycling are discussed.

© 2016 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

In a world facing both resource depletion and global warming, and in societies facing obesity as well as transport congestion problems, the use of cycling for transportation seems like a good idea. Forsyth et al. (2009) documents the promotion of cycling as sustainable transport from the 1960s onwards.

There is currently therefore, considerable interest among planners both in catering to the needs of existing cyclists, and in encouraging non-cyclists to switch to this option. There is no simple way to achieve these aims, and in particular the literature suggests that only comprehensive packages of policies can help to achieve the latter (Forsyth and Krizek, 2010, 2011; Handy et al., 2014; McCormack and Shiell, 2011; Pucher et al., 2010). For better or worse, the benefits of active travel schemes are principally evaluated on a financial basis; the primary benefit from this perspective is usually the saving to healthcare providers dealing with a more active population, although a secondary saving arises from increased safety of those who are already active (Canning et al., 2012; sample analyses of this type include Wilson and Cope, 2011). Various authors, possibly inspired by the overlap of cycling with public health issues, have noted the need for developing a more robust quantitative evidence base for cycling in urban planning (Krizek et al., 2009), more reliable and better validated GIS measures, e.g. for street

patterns (Brownson et al., 2009) and better ways to ensure integration of new infrastructure with the “continuous network” (Forsyth and Krizek, 2011).

Traditional transport modelling methods have to date had limited application in the modelling of active travel, due to their focus on large scale transport analysis zones rather than small, link-level features which affect the travel decisions of pedestrians and cyclists (Cervero, 2006). This paper therefore focuses on spatial network analysis (SpNA), which offers an alternative to traditional methods for predicting fine-grained flows on networks, by working at unsimplified network link level. Karou and Hull (2014) demonstrate the use of SpNA to measure accessibility to public transport, and Zhang et al. (2015) show a strong link between various network centrality measures and accidents involving non-motorists. Returning to the prediction of flows, Lowry (2014) uses ordinary (shortest Euclidean network distance) betweenness to interpolate from a small set of measured vehicle flows to the remainder of the road network. In the case of bicycles, spatially localized angular betweenness (reviewed in Cooper, 2015) is used in a series of space syntax studies (Law et al., 2014; Manum and Nordstrom, 2013; Raford et al., 2007), often in combination with other variables, to predict flows of cyclists on the network. Angular betweenness is a measure which identifies the links most commonly used in straightest-path (rather than shortest-path) routes through the network. While this does correlate with flows of cyclists to some extent, angular betweenness is the exact same measure used in the space syntax tradition to predict flows of vehicles and pedestrians (Hillier and Iida, 2005). Combining angular betweenness

E-mail address: [cooperch@cardiff.ac.uk](mailto:cooperch@cardiff.ac.uk).

with other factors in a multivariate regression model could thus be viewed as post-processing (Cervero, 2006) a single underlying direct demand transport model which – apart from the straightest-path aspect – is similar to Lowry's (2014) vehicle model. Prior to post-processing, such a model differentiates between the navigational choices of drivers, cyclists and pedestrians only in terms of the distance that each are willing to travel. Common sense suggests that these three classes of road user are in fact likely to choose routes in a different manner; the observation of Raford et al. (2007) that preferred cyclist routes are often parallel to, but not coincident with routes of high angular betweenness would also suggest this possibility, and discrete choice modelling in fact confirms it (Broach et al., 2012; Wardman et al., 2007). The approach of the current paper, therefore, is to move beyond the SpNA obsession with angular betweenness and define betweenness in terms of factors more likely to influence cyclist route choice (distance, slope, vehicle traffic and angular distance), thus giving the SpNA model a stronger behavioural foundation.

Related to the problem of predicting bicycle flows, is the problem of predicting *trip generation, trip distribution and mode choice*, i.e. whether or not people will travel in the first place, where they will choose to travel and whether they will choose to cycle. For the purpose of this paper we group these three phenomena into a single model of *cycle trip demand*. Any model of cycle flow must contain at least some assumptions on trip demand, if not an explicit model. For the environmental, social and economic policy reasons noted above, the question of whether and to what extent new infrastructure affects the number of people who choose to cycle is an area of active research. Note however, that a model of cycle demand for individual trips is not necessarily the same as a model of cycle demand for the aggregate population. Feedback loops such as the land use-accessibility cycle (Chiaradia et al., 2014 contains a brief review), and residential self-selection (Cervero, 2006) will mean that keen cyclists may deliberately choose to live near a suitable route for cycling to work, or perhaps that an abundance of opportunities for cycling will lead to a population more willing to consider cycling as a mode of transport. Thus, in addition to models of mode choice for individual routes (e.g. Wardman et al., 2007) there exist numerous models of mode choice at district level, which aside from being necessary, are advantageous to explore due to the better availability of data at this level. Some of these models have been unable to discern any effect of infrastructure on cycling (Goodman et al., 2014; Pooley et al., 2011) while others have (Ewing et al., 2014; Winters et al., 2013) including Parkin et al. (2007) which is endorsed by Department for Transport (2014a). Parkin et al. (2007) find a strong correlation between the proportion of off-road cycle path at local authority level, and decision to cycle to work at ward level; however this model does not account for the location of off-road path, and so from the perspective of urban design is not sensitive to the location options a planner might consider. There is also little history of any such district level models being used to predict cycle flows at network link level; traditionally, cycle mode choice and hence demand is considered exogenous to existing transport models such as TRIPS and QUOVARDIS (Schwartz et al., 1999) which themselves tend to operate on a simplified network suitable for motor vehicles rather than cyclists. In the absence of suitable models, current approaches to planning cycle networks involve bypassing the non-existent demand model and instead focussing on *potential*, defined as existing motorized trips that could be undertaken by bike (Lovelace et al., 2015).

The objective of the current study is to create an SpNA methodology for calibrating a more behaviourally realistic, link level model of cyclist flow in cities, and to test whether that model fits available flow data. The ultimate aim is to help planners with decisions on cycling infrastructure, by providing a tool (i) to extrapolate from a small number of measured flows to likely cycling patterns over the entire network; (ii) to predict the effect on flows of new infrastructure schemes, (iii) to understand how proposed schemes will (or won't) integrate with the wider network and (iv) to help understand interactions between flows of cyclists and cars.

The model attempts to form a bridge between SpNA and transport modelling traditions. To this end it is based on discrete choice literature which studies the effect of distance, turns, slope and vehicle traffic on cyclist route choice, combined with the SpNA calculation of spatially localized betweenness reviewed in Cooper (2015). It will be shown that this is equivalent to a mode choice model which uses urban density to predict the decision to cycle. The approach can perhaps be classed as 'extreme' spatial network analysis in that it considers network accessibility itself to be the primary driving cause of land use and hence trip generation, resulting in a model which is low cost with respect to data collection requirements: at a minimum the model can function using only the network itself as input, although in the current case we calibrate using measured flows. As it handles route finding decisions of drivers and cyclists separately, the model presented is applicable to road safety models examining the interaction between the two classes of road user; this application is briefly demonstrated and also serves as validation of the flow model. We conclude by discussing further avenues for modelling the effect of infrastructure on cycling. The model is reproducible using publicly available, general purpose network analysis software sDNA+, which can function either as a GIS or CAD plugin (Cooper et al., 2011).

## 2. Methods

### 2.1. Data

Our spatial network is based on Open Street Map (OSM). OSM is an open access, crowd sourced mapping product with global scope; while coverage is currently patchy, both coverage and data quality are only likely to increase for the foreseeable future, and for coverage of cycle paths OSM is currently the best on offer in the UK (Lovelace, 2015). The OSM network is however, still liable to contain spatial errors which are detected and repaired as described in Cooper (2016, chapter 2). Slope data is taken from OS Terrain 50.

Fig. 1 shows a map of all cycle sampling locations and counts used in the study. Two sources of cycle flow data were used to calibrate the model:

1. Department for Transport (2014b) (DfT) pedal cycle flows, which are derived from a mixture of vehicle gates and manual sampling at 107 locations in Cardiff. The DfT report annual average daily traffic (AADT). However, only locations on roads carrying vehicle traffic are recorded, so this data gives no indication of the use of traffic free paths. The DfT deem any weekday from March–October to be a 'neutral day' on which a representative sample of traffic can be taken; AADT is estimated by applying expansion factors based on type of road, day of year and type of vehicle. Thus while this methodology takes account of national weather variations it discards regional ones, which may have a major effect on pedal cycle usage. Additionally, it may underestimate recreational pedal cycle traffic on weekends. Finally, in some cases roads are not sampled at all, but flows are estimated by applying a growth factor to count data from a previous year (Department for Transport, 2011).
2. Cardiff Council's data, collected from 14 electronic cycle counters on traffic free paths over a 3 month period. Cardiff council report only daily flows for the recorded months (September–November 2014). As cycle traffic is heavily seasonal, a seasonal correction was derived from a 15th electronic cycle counter which records year round flows, to estimate AADT.

As the two sources of data were collected using different methodologies, we cannot be sure whether or not the AADT flows derived from each are truly comparable (a more detailed discussion of issues with cycle count data can be found in Gordon, 2014). However, as both data types are essential to understanding flows of pedal cycles, they must be combined nonetheless. To account for errors generated by the

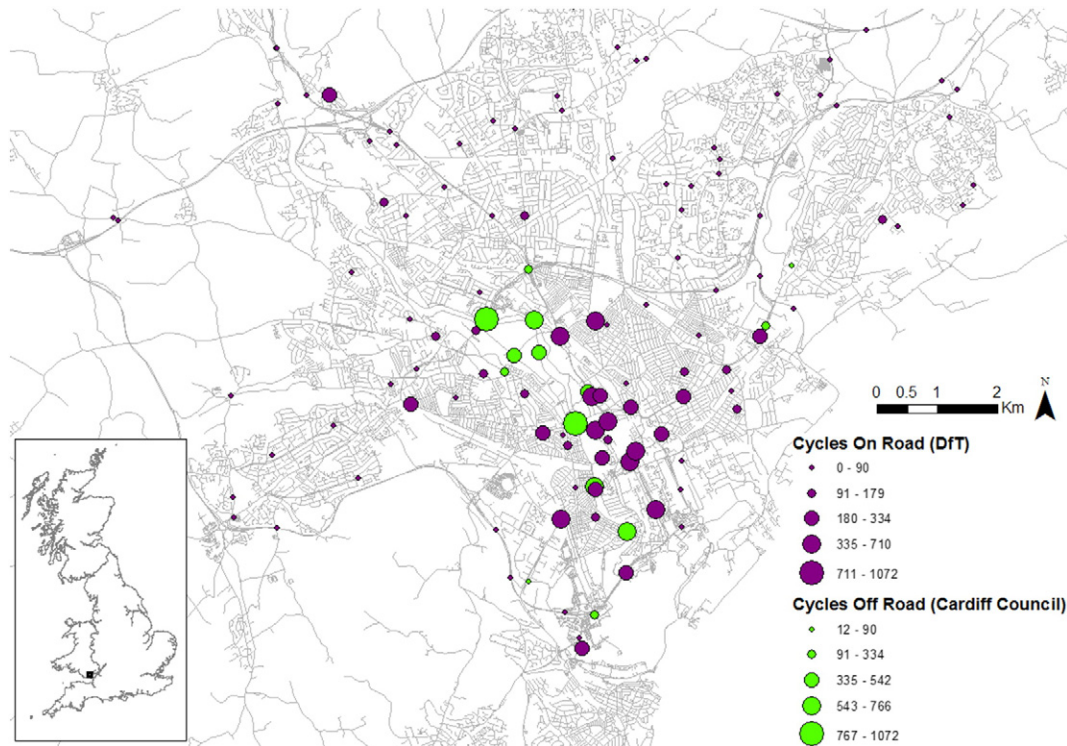


Fig. 1. Measured cycle flows in Cardiff. DfT and Council figures both estimate AADT but are not necessarily comparable.

mismatch of methods, the final flow model includes a dummy variable to account for the data source.

The vehicle flow sub-model is calibrated using vehicle counts from the same 107 locations sampled by the DfT (2014b). Mode choice data is derived from the UK Census (Office for National Statistics, 2011) aggregated to output area level, at which there are 1076 zones within the city of Cardiff.

Road traffic incident data is sourced from DfT (2013) and records 767 incidents involving at least one motor vehicle and at least one cyclist, between 2005 and 2012. The data includes spatial co-ordinates, vehicle types and in some cases road names. The spatial accuracy is limited, with locations taken from grid references recorded by officers attending the scene of incident; in some cases due to human error these do not fall on the named road, and in other cases they do not fall on any road in the spatial model. For the purpose of this study, we consider locations to be accurate to the nearest 30 m.

## 2.2. Trip generation, distribution and mode choice models

The SpNA models of vehicle, pedestrian and cyclist flow cited in the introduction have all used some form of spatially localized betweenness as a predictor of actual flow. In rough terms, betweenness is the output of a flow model which simulates indiscriminate trips from everywhere to everywhere, subject to a maximum trip distance or *radius*, and some criteria by which each trip is routed. Thus it replaces all four stages of the traditional transport model (de Ortúzar and Willumsen, 2011): trip generation, trip distribution, mode choice and route choice. In framing betweenness as a transport model, the first three of these are subsumed into a single, simplified model of transport demand (“indiscriminate trips from everywhere to everywhere”). The route choice model remains distinct, and is used to determine the path of each individual trip, although the betweenness model outputs link-level flows rather than routes. That the literature shows such models fitting the data at all, let alone well, could be considered surprising from a transport perspective as the indiscriminate simulation of trips takes into

account neither land use nor trip purpose. It is therefore prudent to consider why good model fit might arise from this process in order to understand the limits of SpNA models. This exercise also helps to shine some light on the links between SpNA and traditional transport modelling, from which perspective SpNA contains some implicit assumptions which are not usually voiced.

Briefly reviewing the mode choice literature, we note the importance of urban density as a common theme among all models which find a link between the built environment and decision to cycle. Winters et al. (2013) considers the effect of bike paths on mode choice, though this is tested only in a univariate model and it is unclear to what extent urban density confounds this relationship. Ewing et al. (2014) show a weak relationship between cycle mode choice and several variables (intersection density and connectivity, population and jobs) all of which strongly correlate with urban density. Parkin et al. (2007) describe a combined built environment and demographic model which explains a large proportion of variance in cycle use for the journey to work ( $r^2 = 0.82$ ); in this case the two largest built environment coefficients relate to the proportion of off-road cycle path, and distance travelled. Note that the latter variable will again, on average, tend to correlate (inversely) with urban density as the existence of numerous job opportunities nearby increases the likelihood that a randomly selected individual will both occupy such a job, and cycle to it. Similar logic also applies both to discretionary and recreational trips. It is thus reasonable to assume that both cycle trip generation and distribution correlate with urban density.

Betweenness implicitly scales with density. In the current study, following Chiaradia et al. (2014) we use link weighted betweenness as the density of network links has been shown to correlate strongly with the density of jobs and homes (Chiaradia et al., 2012) and can thus be used to proxy them. The betweenness is spatially localized within a Euclidean buffer, with the buffer representing a maximum trip distance - in SpNA terms, a *radius* - which is fitted by calibration. As a trip is simulated between each distinct pair of links closer than the radius, the denser the network at any given point, the more trips are simulated to and from



that point. This goes some way towards explaining the success of betweenness in modelling pedestrian, cycle and vehicle flows.

Note that the total weight of trips doesn't scale linearly but with the square of urban density. Thus betweenness weighted in this manner quantifies opportunities (potential flows from pairs of links close enough to form a feasible trip) rather than flows between entities with a physical constraint, such as individual households with a travel budget. This is reasonable because betweenness is defined between links which can vary in intensity of land use (i.e. households and jobs per link). The spatial network generates accessibility, and this influences land use through economy of agglomeration – links in denser areas experience more intense land use due to the convenience of their location. Also, scaling effects in cities are known to exist even for individuals (Bettencourt, 2013). Therefore the betweenness model implies an assumption that transport opportunities offered by the network are efficiently exploited through a land use-transport accessibility feedback cycle (Chiaradia et al., 2014). Restricting our consideration to a single mode – cycling – we see that residential self-selection is a special case of this, with cyclists preferring to live on links that are more accessible by bike, and travel to places more accessible by bike.

As well as forming part of the definition of urban density, calibrating the radius parameter also permits competing modes to enter the model, because at higher radius the incentive is to travel by motorized transport rather than cycle. We do not consider any negative effect on cycling of readily available alternatives such as metro transit, in specific locations. While this is not ideal, we note that cycle modelling is still in its infancy and that the Department for Transport (2014a) endorse unimodal approaches for cycling as a simplification.

While we present the model in SpNA terms, it can also be characterized as a direct demand transport model in which trip generation, distribution and mode choice are considered congruent (Cervero, 2006; Lowry, 2014; and de Ortúzar and Willumsen, 2011, chapter 6). Compared to a traditional transport model, the disadvantage is that the lack of individually calibrated stages means we cannot verify each stage of trip generation, distribution, mode choice and route choice individually. In mitigation of this problem, we explicitly test the link between urban density and mode choice.

### 2.3. Route choice model

The design principle guiding the route choice model is that it should make use of publicly available data to capture as many of the factors affecting cyclist route choice as possible. As a starting point we take the commuter-derived figures from Broach et al. (2012) in which a discrete choice model is used to determine factors affecting cyclists' choice of routes in Portland, Oregon. Broach goes on to express the model in terms of equivalent distance, in other words, for each factor, how much extra distance would have an equally deterrent effect on the cyclist. Following our own work with policymakers we have found this choice of words to cause confusion, so instead we call this concept *perceived effort*, as it is the effort of travelling the route measured as the cyclist perceives it rather than as a literal distance. Distance enters Broach's model in log form; thus slope, vehicle traffic and path type have a multiplicative effect on perceived effort. Other factors such as crossings, turns and stop signs are expressed by occurrence per mile, which cancels the multiplication with distance giving them a constant (additive) effect per occurrence.

In the desire to create a model based on publicly available data, and hence more easily usable by practitioners, it is necessary to exclude some factors used by Broach due to lack of data availability. These include traffic signals, stop signs and unsignalled turns and crossings. Following a review of available data on cycle infrastructure in the Cardiff it was found that Open Street Map (2015) contained the most extensive information on traffic free cycle paths (a finding confirmed at UK level by Lovelace, 2015) but patchy information on bike lanes on-road, with not all on-road lanes being recorded, and those that were recorded

being of variable quality. The quality of on-road bike lanes in the UK is subject to extensive criticism with some research questioning whether they improve safety at all (Parkin and Meyers, 2010; Stewart and McHale, 2014). Facilities such as the 'bike boulevards' studied by Broach, as well as being hard to source data for, are also extremely rare in Cardiff and where they do exist, their status as favourable routes is already partially captured in the model due to the absence of traffic on neighbourhood streets. We therefore exclude consideration of on-road bicycle facilities and assume them to be absent. The encouraging effect of off-road facilities is captured in the model as they are completely free from motor traffic, which we simulate explicitly on the roads where it is allowed.

Table 1 summarizes the variables in Broach's model. It can be seen that the attributes with the greatest effect on cycle route choice are slope and vehicle traffic, which are both included in the SpNA model. The largest excluded factor is unsignalled crossing of busy roads (AADT > 20 k) however it would take 4.3 of these per mile to equal the effect of travelling along the same roads, which is included in the SpNA model. Only one section of road in Cardiff is this busy in any case. The excluded factors relating to bridges appear substantial; however Broach notes that these are subject to a high degree of error in that study.

As a high proportion of road links in Cardiff fall into the lowest traffic band in Broach's model (98.6% have estimated AADT < 10,000), applying this model directly would allow little distinction between roads of varying business. We therefore choose to interpolate between classes of vehicle flow by fitting an exponential curve. We plot each 'band' of cost at its lower limit of traffic flow, e.g. the perceived effort of cycling in traffic of 20–30,000 vehicles per day is assumed to apply to 20,000 vehicles per day, as most roads in this each band will have flows close to the lower bound of the band (the distribution of traffic flow over roads tending to exponential tailoff in its upper limits). The exception to this is the 0–10,000 vehicles per day band, for which we take 5000 vehicles per day to be indicative. For zero vehicles per day we take the perceived effort for Broach's traffic free bike path; 840 m for a 1 km trip.

The curve resulting from these assumptions is shown in Fig. 2, and over the range shown can be interpolated to a 'traffic multiplier' expressed as

$$\text{traffic multiplier} = 0.84 e^{k \text{AADT} \times 0.05} = \text{trafficfac}^{0.05} \quad (1)$$

**Table 1**

Equivalent distances for commuting route attributes from Broach et al. (2012) Table 3. Re-sorted by magnitude. AADT = Annual Average Daily Traffic. Reproduced with permission from Elsevier.

Attribute	Distance value (% dist)	Included in SpNA model
Prop. AADT 30 k + w/o bike lane	715.7	Yes
Prop. upslope ≥ 6%	323.9	Yes
Prop. AADT 20–30 k w/o bike lane	140	Yes
Prop. upslope 4–6%	120.3	Yes
Prop. upslope 2–4%	37.1	Yes
Prop. AADT 10–20 k w/o bike lane	36.8	Yes
Unsig. cross AADT 20 k + exc. right turn (mile)	32.2	No
Bridge w/sep. bike facility	– 29.2	No
Left turn, unsig., AADT 20 k + (mile)	23.1	No
Bridge w/bike lane	– 18.2	No
Prop. bike path	– 16	Yes
Prop. bike boulevard	– 10.8	No
Left turn, unsig., AADT 10–20 k (mile)	9.1	No
Unsig. cross AADT 10–20 k exc. right turn (mile)	5.9	No
Turns (mile)	4.2	Yes
Unsig. cross AADT 5–10 k exc. right turn (mile)	4.1	No
Unsig. cross AADT ≥ 10 k right turn (mile)	3.8	No
Traffic signal exc. right turns (mile)	2.1	No
Stop sign. (mile)	0.5	No

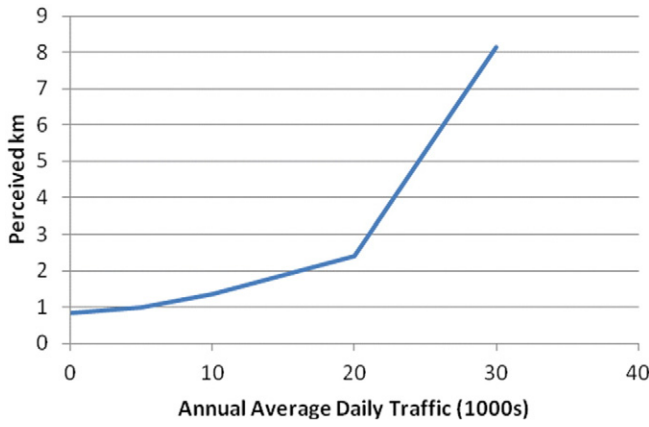


Fig. 2. Perceived effort for cycling in motor vehicle traffic, adapted from Broach et al.

where  $kAADT$  is annual average daily traffic expressed in 1000s. (We define *trafficfac* slightly differently as *traffic multiplier*<sup>1/0.05</sup> for later formulae).

Broach's concept of "turn" relates to junctions on a grid street pattern, i.e. 90° change of direction, which is equivalent to 4.2% of a mile, or 68 m. For irregular street patterns this is modified to an equivalent distance in metres per degree of angular change of 68/90. It should be noted that this incorporates changes of direction along links as well as at junctions, which would not exist in the block street pattern of the original study. Therefore the concept measured is slightly different, but in the original study it would not be possible to distinguish the two. The literature on angular betweenness gives confidence that total change of direction has an effect on route choice.

Noting that slope and traffic have a multiplicative effect on distance, while turns are additive, the above factors can be combined into a hybrid distance metric:

$$\text{perceived effort} = \text{distance} \times \text{slopefac} \times \text{trafficfac}^{0.05} + \text{angular distance} \times \frac{68}{90} \quad (2)$$

where *trafficfac* is as defined above, *distance* is measured in metres, *angular distance* is the cumulative directional change over the whole route measured in degrees, and *slopefac* is the relevant coefficients for slope taken from Broach (converted such that e.g. +323.9% is a ratio of 4.239):

$$\text{slopefac} = \begin{cases} 1.000 & \text{if slope} < 2\% \\ 1.371 & \text{if } 2\% < \text{slope} < 4\% \\ 2.203 & \text{if } 4\% < \text{slope} < 6\% \\ 4.239 & \text{if slope} > 6\% \end{cases} \quad (3)$$

The model is then calibrated to match local data. Multiplicative factors can be calibrated through exponentiation, while additive factors can be calibrated through multiplication. Thus the final model form is

$$\text{perceived effort (calibrated)} = \text{distance} \times \text{slopefac}^s \times \text{trafficfac}^t + \text{angular distance} \times \frac{68}{90} \times a \quad (4)$$

To match Broach we would set calibration parameters  $s = 1$ ,  $t = 0.05$ ,  $a = 1$ . Final parameters are chosen by exploration of parameter space around this point. The sDNA+ configuration used to implement this formula is shown in the appendix; this formula (and

sDNA+) can be extended to take account of further data if available, e.g. on-street cycle lanes, Level of Service, etc.

The estimated vehicle traffic flows used to inform the calculation of *trafficfac* are themselves based on a second sDNA model based on angular betweenness alone, similar to Chiaradia et al. (2014) but calibrated over a range of radii from 10 to 35 km to match motor vehicle trip lengths. The surrounding region is included in the vehicle model so that origins and destinations for vehicle trips can likewise be inferred from urban density. It is necessary to include one-way restrictions in the motor vehicle model to ensure both halves of a dual carriageway are always used by the model, instead of the most direct option being used both ways; if this were not done, any 'unused' sections of dual carriageway would erroneously appear to be attractive traffic-free cycle paths.

#### 2.4. Flow prediction and road safety models

The combination of the mode and route choice models specified above is represented in SpNA terms by a hybrid betweenness measure, computed over a range of radii. We calibrate the model by picking the radius,  $s$ ,  $t$  and  $a$  parameters to give the best correlation between betweenness and measured flows. Note that the calibration process involves only four parameters; origin and destination balancing factors (which would be typical in a transport model) are not employed, so the risk of overfitting the model is minor by comparison. For this reason, we fit the flow model using all available measurements, though the road safety model can still be considered as a test of the flow model against an independent road traffic accident data set.

Recalling that our cycle flow data originates from two separate sources, we post process to fit a variable accounting for differences in collection methodology. However, as the methodology for each measured flow also correlates with the type of location (on road/traffic free), it is not clear to what extent this variable is correcting for data source vs type of cycle path. The model is given in Eq. (5):

$$\text{predicted flow} = \text{source correction} \times \text{Betweenness}^\beta \times \varepsilon \quad (5)$$

where

$$\text{source correction} = \begin{cases} 0 & \text{for Department for Transport (on road)} \\ \alpha & \text{for Cardiff Council (traffic free)} \end{cases} \quad (6)$$

The effect of this is to scale measured flows from Cardiff Council by a factor of  $\alpha$ , while  $\beta$  captures nonlinearities in the relationship between flows and betweenness.

The road safety model is included as a brief demonstration of what is possible with an SpNA model which handles driver and cyclist behaviour differently. As the model discriminates between route choices of cyclists and motor vehicle drivers, independent predictions of flows are produced for each category of user. This enables the construction of a secondary model which predicts the potential for accidents based on the presence of both types of road user in large numbers on the same link. This is a special case of a Hauer model (Turner et al., 2006, sec. 2.3.2) which predicts accident rate  $A$  from flows  $Q_a$  and  $Q_b$  (in our case of motor vehicles and cyclists respectively):

$$A = kQ_a^\alpha Q_b^\beta \quad (7)$$

where we leave the model uncalibrated and set  $\alpha = \beta = k = 1$ , thus the output  $A$  is not a rate but can be interpreted as a 'conflict score' which we pass to a binary classifier to discriminate between low- and high-risk roads, with the threshold chosen manually. The classifier is tested on its ability to discriminate known incident sites from points generated randomly on the network using SANET (Okabe and

Okunuki, 2015), removing those randomly generated points which fall within 30 m of known incident sites. The model handles inaccuracy in the road traffic incident data by classifying each point based on the highest conflict score existing within 30 m of that point.

### 3. Results

For the test of mode choice, the proportion of people choosing to cycle to work correlated with urban density as measured by the number of network links within a 4.5 km buffer, with  $R = 0.61$ . The relationship is shown in Fig. 3. We thus conclude that the study area shows a strong link between urban density and the decision to cycle, which validates use of urban density measured in this manner as the basis of the flow model.

For the motor vehicle sub-model, optimal correlation with measured vehicle flows ( $R = 0.90$ ) is achieved with an angular betweenness radius of 28 km. The resulting flows are shown in Fig. 4.

For predictions of cycle flow, we present in Table 2 an exploration of parameter space surrounding the route choice parameters derived from Broach et al. That is to say, we tweak the model's sensitivity to vehicle traffic, slope and angular distance. The parameters giving the best model fit ( $R = 0.70$ ) are  $a = 0.2$ ,  $s = 1$ ,  $t = 0.04$ . The optimal radius for the cycle flow model is 3 km. This is on a similar scale to the optimal radius for cycle mode choice prediction (4.5 km), though some variation is to be expected as the models are calibrated to a different data sets – cyclist counts and census data respectively. The correlation between mode choice and urban density is also strong on the 3 km scale ( $R = 0.56$ ).

Comparing with Broach's figures derived for Oregon, the inferred effect of slope and road traffic on cyclist route choice is very similar. The effect of directness is reduced, which is to be expected as our model measures all angular distance (i.e. bends in roads) as well as the angular distance encountered at junctions. Additionally, the block street structure of Oregon allows for practical route planning which avoids angular distance – as routes with more turns will typically be no shorter in distance – while the same cannot be said for Cardiff, where cyclists must occasionally overcome their aversion to twisty routes if they wish to pick the shortest path.

Although the difference in model performance for uncalibrated and calibrated models is minor for the measured data points, this should not be taken as an indicator that calibration is only adding marginal quality, as the performance of the model is ultimately likely to be limited by issues in the measurement of pedal cycle flows (Gordon, 2014, and section 2.1 of the current paper). Rather, the clear peaks in model performance for the calibrated values of  $t$ ,  $s$ ,  $a$

and  $radius$  indicate that the selected values are genuinely meaningful, and thus that such calibration is a valid technique to use.

Fig. 5 shows a scatter plot of predictions from the optimal model, against recorded flows which have not been corrected for the source of data. The predicted flows are mapped in Fig. 6. A systemic bias is evident in recorded off-road flows which are higher than predicted; this would be consistent both with the hypothesis that the DfT methodology for recording pedal cycles on-road results in undercounting, and with the hypothesis that off-road paths are more attractive than predicted. Applying the regression model given in Eq. (5) thus improves correlation to  $R = 0.78$  for estimated  $\alpha = 1.87$ ,  $\beta = 0.64$ . This represents a 28% reduction in model error.

The road safety model (Fig. 7) predicted 75% of incident sites and 73% of non-incident sites (i.e. has 75% sensitivity, 73% specificity) based on a threshold conflict score of 42 million.

### 4. Discussion and conclusions

The contribution of this study has been to present a methodology for fine-scale, city-wide modelling of cyclist flows, by combining SpNA with more behaviourally realistic foundations. The model is based on minimal data that is generally available to the public. It provides reasonable correlation with measured flows and is sensitive to the location and nature of changes to infrastructure. A secondary novel aspect is the simulation of link level motor traffic flows to feed into the simulation of cyclist flows, thus accounting for the effect of one on the other.

For the models presented here to be useful in practice they are enhanced by tools for managing any mismatch between measured flows, model predictions and user expectations. To this end, the sDNA + software includes features which allow users to establish why the model predicts that links are, or are not used, when the user thinks or measured flow data shows otherwise. To assist understanding of how new infrastructure might integrate into the existing network, visualisations can be filtered to display only trips that pass through a specific link, thereby showing predicted behaviour associated with new infrastructure.

As with any transport model, if applying this model elsewhere it is best to recalibrate the parameters, but the model structure should generalize without modification: a betweenness model for motor vehicle flows which informs a second betweenness model of cyclist flows based on distance, angular distance, slope and motor traffic. Pilot work in other locations shows smaller towns may not exhibit such strong correlation between urban density and cycle transport demand, e.g. due to site specific factors such as the location of a single large employer. Including these factors would be a direction for future model improvement. Also, as noted in Section 2.2, total betweenness weight scales with the square of urban density: explicit calibration of this scaling effect may also improve the model.

A more accurate assessment of model performance would be achieved in a location where on-road and off-road cycle flows are recorded using the same methodology. While we have fitted a regression parameter to account for data source, it is not clear to what extent this also corrects for cyclists' preference for traffic free paths above and beyond the extent to which the route choice model already accounts for aversion to traffic. If such an effect remains in a future study, then the model is easily modified (1) by recalibrating for greater aversion to traffic; (2) by calibrating a multivariate regression flow model with two classes of cyclist, confident and unconfident, with differing aversion to traffic; (3) by providing additional weight to traffic free paths as leisure destinations in their own right.

Ultimately, we would like to answer the call of Krizek et al. (2009) for medical-grade evidence of the effects of cycling infrastructure on health, and more broadly, urban design on health. SpNA has already shown some promise in this regard when predicting community cohesion mediated by walkability (Cooper et al., 2014). In the case of cycling, the links from models of flow to health cost/benefit ratios are already

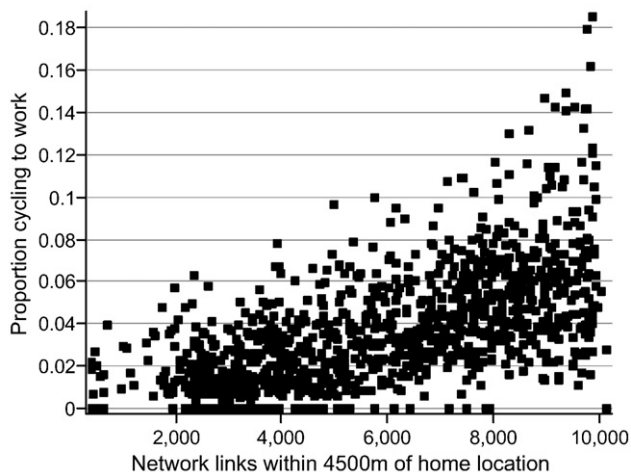


Fig. 3. Relationship between cycle mode choice and urban density for output areas in Cardiff.



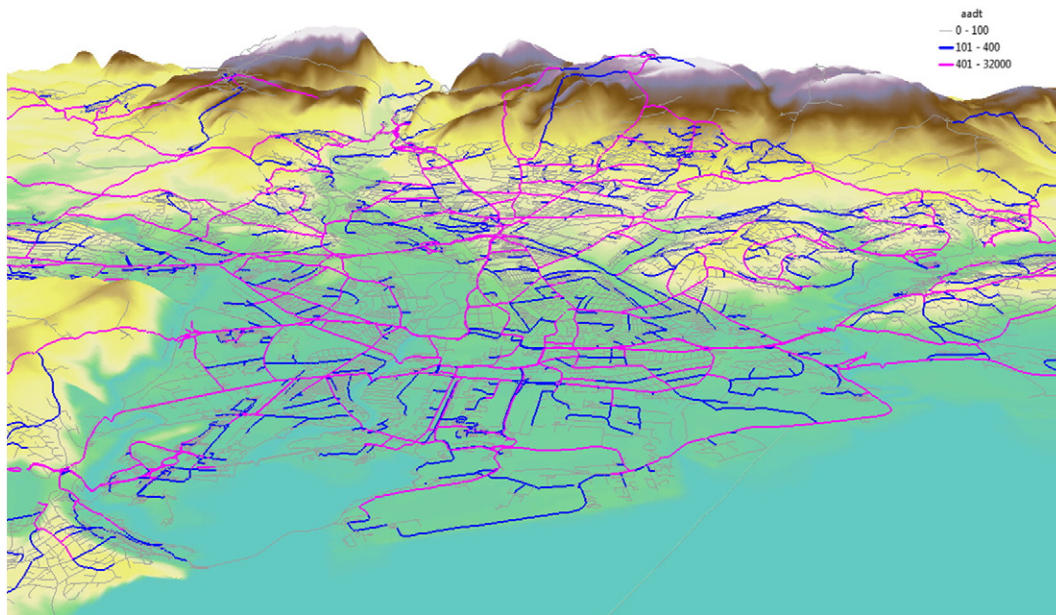


Fig. 4. sDNA estimates of motor vehicle Annual Average Daily Traffic (AADT) used to inform cycle route choice model. 3-d model rendered in ArcGIS ArcScene; vertical exaggeration = 5.

partly quantified by the Health Economic Assessment Tool (or “HEAT”, World Health Organization, 2014). The proportion of new flows generated by infrastructure, and the mean trip distance for cyclists, are both inputs to HEAT which in principle can be predicted by this model. It is also possible to compute the reduction of exposure to road traffic for existing cyclists (a lesser benefit which is missing from the HEAT model) and, if we take cycling culture to be an exogenous factor, we can predict future cycling flows on links in the face of an increased tendency for people to cycle longer distances. However, the approach of computing cost-benefit for cycle projects has come under criticism (Hollander, 2016) so it may be reasonable to stop short of such outputs, instead using SpNA to visualise accessibility and flows through infrastructure as one of many factors informing the decision making process.

In terms of the effect of infrastructure on demand for cycling, this model addresses some questions, but leaves others open – much like the varying results cited in the introduction (Ewing et al., 2014; Goodman et al., 2014; Parkin et al., 2007; Pooley et al., 2011; Winters et al., 2013). As we simulate the effect of urban density on demand, and as new infrastructure can cause dramatic increase in density (e.g. if it connects a dense area to a previously isolated one), we do show an effect for new infrastructure that alters the spatial distribution of urban density. It would be an improvement, however, to test an updated model in which cycle demand is increased by the presence of low-traffic or traffic free paths, irrespective of density. A model demonstrating this effect on smaller than city scale remains elusive, and while Parkin et al. (2007) shows the effect to be present at Local Authority

scale, Parkin’s model is sensitive only to the total length of traffic free paths rather than their location, and thus not useful for planning the path locations.

With the aim of upgrading this effort to be sensitive to the location of cycle paths – and thus useful to the planners designing them – we therefore propose further improvement of the SpNA model presented above by (1) calibrating a trip length distribution rather than a simple maximum length, and (2) measuring such lengths in terms of perceived effort rather than Euclidean network distance – in SpNA terms, a hybrid radius.

## Acknowledgements

Road Traffic Incident data is public sector information licensed under the Open Government Licence v3.0. Basemaps © OpenStreetMap contributors. 3d Terrain and Fig. 1 inset map contain OS data © Crown copyright and database right 2016. Data provided by the City of Cardiff Council used with permission, processed according to the methodology described herein. Usage does not imply endorsement by the Council of technical work undertaken or results produced.

Table 2

Results of calibration: bivariate correlation between hybrid betweenness (radius 3 km) and real flows as a function of traffic, slope and angular distance avoidance factors individually. Both variables are Box-Cox transformed.

t (s = 100, a = 20)	bivariate correlation (r)	s (t = 4, a = 20)	bivariate correlation (r)	a (t = 4, s = 100)	bivariate correlation (r)
0	0.66	0	0.68	0	0.67
0.01	0.67	0.5	0.69	0.2	0.70
0.02	0.69	1	0.70	0.4	0.68
0.03	0.69	1.5	0.69	0.6	0.68
0.04	0.70	2	0.68	0.8	0.67
0.05	0.70			1.0	0.66
0.07	0.67				
0.09	0.64				

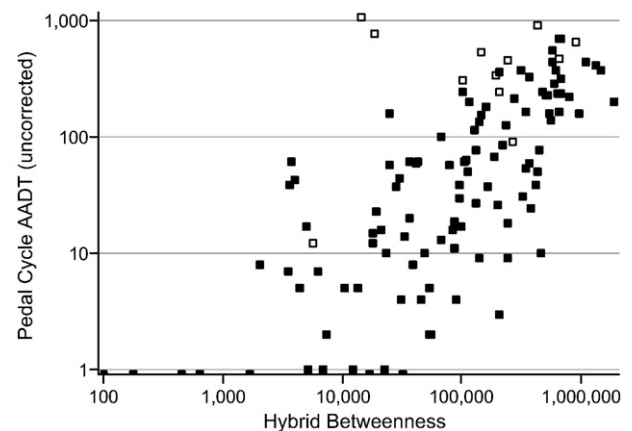


Fig. 5. Bivariate scatter plot showing the relation between Hybrid Betweenness ( $a = 0.2$ ,  $s = 1$ ,  $t = 0.04$ , radius = 3 km) and annual average daily cycle flows, uncorrected for the data source. On-road flows recorded by DfT are shown in black; off-road flows recorded by Cardiff Council in white. Correcting for data source improves correlation, in effect by shifting the white points downwards.

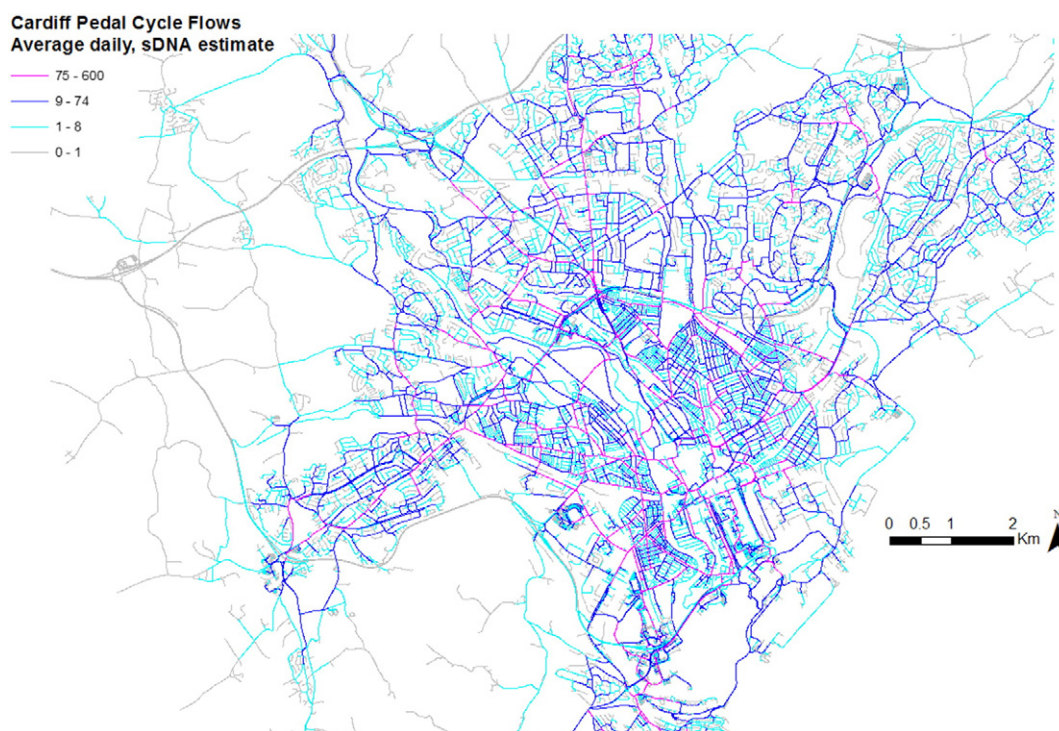


Fig. 6. Map of predicted cycle flows in Cardiff (2-d projection of 3-d model).

#### Appendix A. Configuration of sDNA+

The sDNA+ hybrid metric below reproduces the cycle model derived from fine calibration. As such it uses the following parameters, though these can be changed in the code:  $a = 0.2$ ,  $s = 1$ ,  $t = 0.04$ . It

is necessary to specify perceived effort separately for links and junctions, and also to compute the slope from 3d geometry data. Access to geometry is provided by built-in variables `euc` (Euclidean distance), `ang` (angular distance) and `hg` (height gain). Variables defined during formula execution must be preceded with an underscore (`_`) to

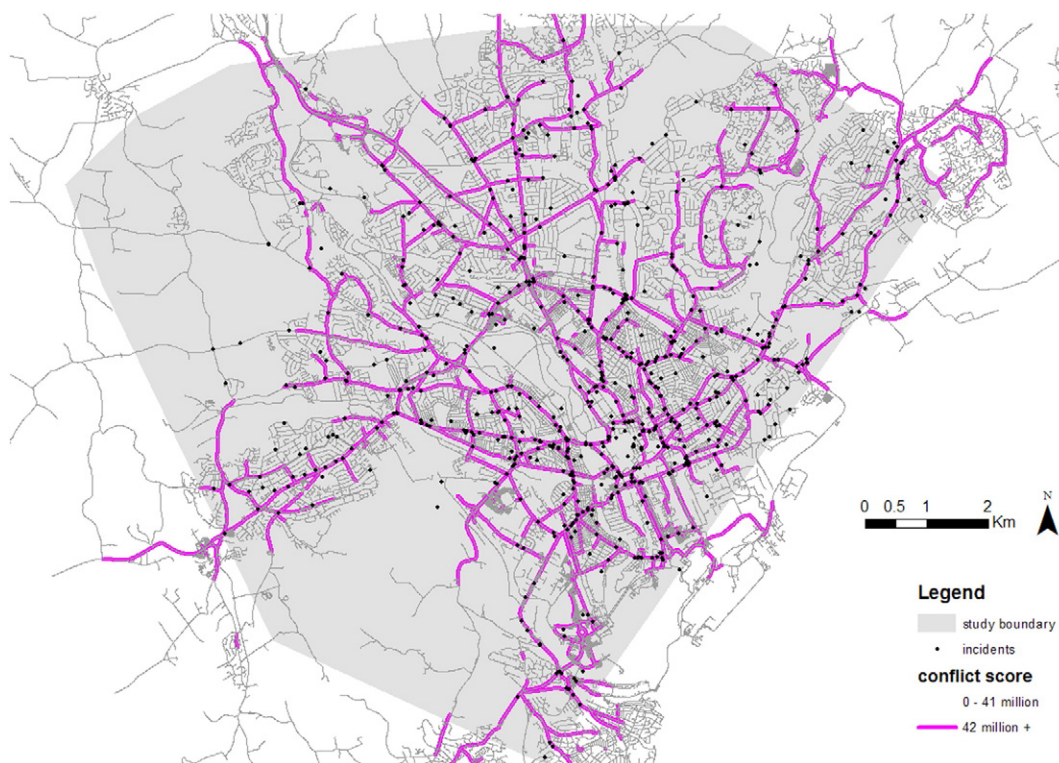


Fig. 7. Model of high-risk roads in Cardiff, overlaid by all road traffic incidents 2005–2012 which involved at least one cyclist and at least one motor vehicle.



distinguish them from pre-existing data attached to the network. For the cycle model, vehicle traffic flow data is pre-existing (having been computed by the angular betweenness model) and referenced in this case as *aadt*.

metric = hybrid; radii = 3000

linkformula =

\_a = 0.3,

\_s = 0.5,

\_t = 0.04,

\_slope = hg/FULLeuc \* 100,

\_slopefac = \_slope < 271:(\_slope < 471.371:(\_slope < 672.203:4.239)),

\_trafficfac = 0.84 \* exp(aadt/1000),

euc \* (\_slopefac^\_s) \* (\_trafficfac^\_t) + \_a \* 68/90 \* ang;

juncformula =

\_a = 0.3,

\_a \* 68/90 \* ang

The configuration for the sDNA vehicle traffic model used was

metric = angular; radii = 28000

To convert the output of this model into *aadt* as referenced in the cyclist model, a conversion factor is calibrated by regression.

## References

- Bettencourt, L.M.A., 2013. The origins of scaling in cities. *Science* 340 (6139):1438–1441. <http://dx.doi.org/10.1126/science.1235823>.
- Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transp. Res. A Policy Pract.* 46 (10):1730–1740. <http://dx.doi.org/10.1016/j.tra.2012.07.005>.
- Brownson, R.C., Hoehner, C.M., Day, K., Forsyth, A., Sallis, J.F., 2009. Measuring the built environment for physical activity. *Am. J. Prev. Med.* 36 (4 Suppl):S99–123, e12. <http://dx.doi.org/10.1016/j.amepre.2009.01.005>.
- Canning, S., Millar, R., Moore, K., 2012. Health impacts of Scotland's canals. Presented at the Scottish Transport Applications and Research conference, Glasgow (accessed 9.15.16). <https://canalrivertrust.org.uk/media/library/6343.pdf>.
- Cervero, R., 2006. Alternative approaches to modeling the travel-demand impacts of smart growth. *J. Am. Plan. Assoc.* 72 (3), 285–295.
- Chiaradia, A., Hillier, B., Schwander, C., Wedderburn, M., 2012. Compositional and urban form effects in centres in Greater London. *Urban Design and Planning – Proceedings of the ICE*. 165, pp. 21–42.
- Chiaradia, A., Cooper, C.H.V., Wedderburn, M., 2014. Network geography and accessibility. *Proceedings of 12th Transport Practitioners' Meeting*. PTRC, London.
- Cooper, C.H.V., 2015. Spatial localization of closeness and betweenness measures: a self-contradictory but useful form of network analysis. *Int. J. Geogr. Inf. Sci.* 1–17 <http://dx.doi.org/10.1080/13658816.2015.1018834>.
- Cooper, C.H.V., 2016. Spatial Design Network Analysis (sDNA) Version 3.4 Manual. (accessed 9.15.16). <http://www.cardiff.ac.uk/sdna/software/documentation>.
- Cooper, C.H.V., Chiaradia, A., Webster, C., 2011. Spatial Design Network Analysis (sDNA). (accessed 9.15.16). [www.cardiff.ac.uk/sdna](http://www.cardiff.ac.uk/sdna).
- Cooper, C.H.V., Fone, D.L., Chiaradia, A., 2014. Measuring the impact of spatial network layout on community social cohesion: a cross-sectional study. *Int. J. Health Geogr.* 13 (1):11. <http://dx.doi.org/10.1186/1476-072X-13-11>.
- Ortúzar, J.D., Willumsen, L.G., 2011. *Modelling Transport*. fourth ed. Wiley-Blackwell, Chichester, West Sussex, United Kingdom.
- Department for Transport, 2011. Road Traffic Estimates Methodology Note. UK. (accessed 9.15.16). [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/230528/annual-methodology-note.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/230528/annual-methodology-note.pdf).
- Department for Transport, 2013. Road Accidents and Safety Statistics. (accessed 1.8.15). <http://data.gov.uk/dataset/road-accidents-safety-data>.
- Department for Transport, 2014a. Active Mode Appraisal (No. A5.1), Transport Analysis Guidance. UK. (accessed 9.15.16). [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/370544/webtag-tag-unit-a5-1-active-mode-appraisal.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/370544/webtag-tag-unit-a5-1-active-mode-appraisal.pdf).
- Department for Transport, 2014b. Traffic Count Data. (accessed 9.15.16). <http://www.dft.gov.uk/traffic-counts/>.
- Ewing, R., Tian, G., Goates, J.P., Zhang, M., Greenwald, M.J., Joyce, A., Kircher, J., Greene, W., 2014. Varying influences of the built environment on household travel in 15 diverse regions of the United States. *Urban Stud.* 52 (13):2330–2348. <http://dx.doi.org/10.1177/0042098014560991>.
- Forsyth, A., Krizek, K.J., 2010. Promoting walking and bicycling: assessing the evidence to assist planners. *Built Environ.* 36 (4):429–446. <http://dx.doi.org/10.2148/benv.36.4.429>.
- Forsyth, A., Krizek, K., 2011. Urban design: is there a distinctive view from the bicycle? *J. Urban Des.* 16 (4):531–549. <http://dx.doi.org/10.1080/13574809.2011.586239>.
- Forsyth, A., Krizek, K.J., Rodríguez, D.A., 2009. Non-motorized travel research and contemporary planning initiatives. *Hot, Congested, Crowded and Diverse: Emerging Research Agendas in Planning*. *Progress in Planning Vol. 71* (4):pp. 170–184. <http://dx.doi.org/10.1016/j.progress.2009.03.001>.
- Goodman, A., Sahlqvist, S., Ogilvie, D., iConnect Consortium, 2014. New walking and cycling routes and increased physical activity: one- and 2-year findings from the UK iConnect study. *Am. J. Public Health* 104 (9):e38–e46. <http://dx.doi.org/10.2105/AJPH.2014.302059>.
- Gordon, G., 2014. Modelling hourly bicycle counts using generalized linear modelling. The Building Futures Group (accessed 9.15.16). <http://www.thebuildingfuturesgroup.com/wp-content/uploads/2014/07/Student-Ambassadors-Report-Gordon.pdf>.
- Handy, S., van Wee, B., Kroesen, M., 2014. Promoting cycling for transport: research needs and challenges. *Transp. Rev.* 34 (1):4–24. <http://dx.doi.org/10.1080/01441647.2013.860204>.
- Hillier, B., Iida, S., 2005. Network and psychological effects in urban movement. In: Cohn, A.G., Mark, D.M. (Eds.), *Spatial Information Theory, Lecture Notes in Computer Science*. Springer, Berlin Heidelberg, pp. 475–490.
- Hollander, Y., 2016. Justifying Investments in Cycling Infrastructure: 10 Lessons Learnt. (accessed 9.15.16). <http://www.ctthink.com/publications.html>.
- Karou, S., Hull, A., 2014. Accessibility modelling: predicting the impact of planned transport infrastructure on accessibility patterns in Edinburgh, UK. *J. Transp. Geogr.* 35: 1–11. <http://dx.doi.org/10.1016/j.jtrangeo.2014.01.002>.
- Krizek, K.J., Handy, S.L., Forsyth, A., 2009. Explaining changes in walking and bicycling behavior: challenges for transportation research. *Environ. Plann. B. Plann. Des.* 36 (4): 725–740. <http://dx.doi.org/10.1068/b34023>.
- Law, S., Sakr, F.L., Martinez, M., 2014. Measuring the changes in aggregate cycling patterns between 2003 and 2012 from a space syntax perspective. *Behav. Sci.* 4 (3):278–300. <http://dx.doi.org/10.3390/bs4030278>.
- Lovelace, R., 2015. Crowd sourced vs centralised data for transport planning: a case study of bicycle path data in the UK. GIS Research UK (GISRUK). Leeds (accessed 9.15.16). [http://leeds.gisruk.org/abstracts/GISRUK2015\\_submission\\_71.pdf](http://leeds.gisruk.org/abstracts/GISRUK2015_submission_71.pdf).
- Lovelace, R., Goodman, A., Aldred, R., Berkoff, N., Abbas, A., Woodcock, J., 2015. The Propensity to Cycle Tool: An Open Source Online System for Sustainable Transport Planning. *arXiv:1509.04425 [cs]*.
- Lowry, M., 2014. Spatial interpolation of traffic counts based on origin–destination centrality. *J. Transp. Geogr.* 36:98–105. <http://dx.doi.org/10.1016/j.jtrangeo.2014.03.007>.
- Manum, B., Nordstrom, T., 2013. Integrating bicycle network analysis in urban design: improving bikeability in Trondheim by combining space syntax and GIS-methods using the place syntax tool. *Proceedings of the Ninth International Space Syntax Symposium*. Sejong University, Seoul.
- McCormack, G.R., Shiell, A., 2011. In search of causality: a systematic review of the relationship between the built environment and physical activity among adults. *Int. J. Behav. Nutr. Phys. Act.* 8 (1):125. <http://dx.doi.org/10.1186/1479-5868-8-125>.
- Office for National Statistics, 2011. *Aggregate Data (England and Wales)*.
- Okabe, A., Okunuki, K., 2015. *Sanet. A Spatial Analysis Along Networks*. Tokyo, Japan.
- OpenStreetMap contributors, 2015. *Open Street Map*.
- Parkin, J., Meyers, C., 2010. The effect of cycle lanes on the proximity between motor traffic and cycle traffic. *Accid. Anal. Prev.* 42 (1), 159–165.
- Parkin, J., Wardman, M., Page, M., 2007. Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation* 35 (1):93–109. <http://dx.doi.org/10.1007/s11116-007-9137-5>.
- Pooley, C., Tight, M., Jones, T., Horton, D., Scheldeman, G., Jopson, A., Mullen, C., Chisholm, A., Strano, E., Constantine, S., 2011. Understanding Walking and Cycling: Summary of Key Findings and Recommendations. Lancaster University.
- Pucher, J., Dill, J., Handy, S., 2010. Infrastructure, programs, and policies to increase bicycling: an international review. *Prev. Med.* 50, S106–S125.
- Raford, N., Chiaradia, A., Gil, J., 2007. Space syntax: the role of urban form in cyclist route choice in central London. *TRB (Transportation Research Record) 86th Annual Meeting Compendium of Papers CD-ROM*. Presented at the Transportation Research Board 86th Annual Meeting, Transportation Research Board, Washington, DC, pp. 07–2738.
- Schwartz, W.L., Porter, C.D., Payne, G.C., Suhrbier, J.H., Moe, P.C., Iii, W., 1999. *Guidebook on Methods to Estimate Non-Motorized Travel: Supporting Documentation*. (accessed 9.15.16). <http://trid.trb.org/view.aspx?id=503335>.
- Stewart, K., McHale, A., 2014. Cycle lanes: their effect on driver passing distances in urban areas. *Transport* 29 (3):307–316. <http://dx.doi.org/10.3846/16484142.2014.953205>.
- Turner, S., Rozenburg, A., Francis, T., 2006. Predicting Accident Rates for Cyclists and Pedestrians (No. 289), Research Report. Land Transport, New Zealand (accessed 9.15.16). <http://www.nzta.govt.nz/resources/research/reports/289/docs/289.pdf>.
- Wardman, M., Tight, M., Page, M., 2007. Factors influencing the propensity to cycle to work. *Transp. Res. A Policy Pract.* 41 (4):339–350. <http://dx.doi.org/10.1016/j.tra.2006.09.011>.
- Wilson, A., Cope, A., 2011. Value for money of walking and cycling interventions: making the case for investment in active travel. *Scottish Transport Applications & Research* (accessed 9.15.16). <http://www.starconference.org.uk/star/2011/angelaWilson.pdf>.
- Winters, M., Brauer, M., Setton, E.M., Teschke, K., 2013. Mapping bikeability: a spatial tool to support sustainable travel. *Environ. Plann. B. Plann. Des.* 40 (5):865–883. <http://dx.doi.org/10.1068/b38185>.
- World Health Organization, 2014. *Health Economic Assessment Tool*. (accessed 12.22.14). <http://heatwalkingcycling.org/index.php>.
- Zhang, Y., Bigham, J., Ragland, D., Chen, X., 2015. Investigating the associations between road network structure and non-motorist accidents. *J. Transp. Geogr.* 42:34–47. <http://dx.doi.org/10.1016/j.jtrangeo.2014.10.010>.