# Methods for the validation of model-based flows allocated to the road network: a case study of cycling

## Abstract

To do:

* Polish the text and add references in the review of methods for model validation.
* Accident data analysis
* Generate PCT estimates for London.
* Get James Howarth to compare with Strava data (sounding out – 2nd paper?)
* Compare with Bike Citizens data (illustrative example of e.g. Hackney)
* Write conclusions
* Submit!

## Introduction

Modelling is an important component of the transport planner’s toolkit, providing insight into transport systems that would otherwise be impossible (Lacono et al. 2007). Models facilitate the estimation of current and future scenarios of movement patterns (Schlesinger et al. 1979); visualisation of key patterns which may otherwise be invisible (Keim, 2001) and evaluation of alternative investment options in 'model experiments'. This can help make important decisions, e.g. where new bus routes should go (Wirasinghe and Vandebona 2011). Modelling and the communication of model results can thus be seen as central to the transport planning process. Indeed, modelling has become so central to transport planning decision making that there have been calls to reassess its role in the sector. A recent report, for example, identifies instances of misuse in transport modelling and proposes greater levels of transparency and public engagement in model development to tackle this issue (Hollander 2015).

A critical but often overlooked component influencing the amount of trust we should have in transport models is validation. There is a strong argument suggesting that all transport models used for policy evaluation should be verified and cross-validated to test the reliability their results, for example by comparing the model’s output under a specific scenario against real-world data (Anderson and Woessner, 1992). Yet this is rarely done, especially in relation to route-allocated model output. This can lead to remarkable failures in the transport decision making process. In one infamous scheme, the Thames Gateway Bridge, the lack of validation in the underlying model was responsible for an entire scheme being scrapped (Hollander 2015).

This paper uses the example of a model estimating the rate of cycling on the transport network to illustrate methods for rout-allocated model validation. The aim is to demonstrate and compare different geographical methods and datasets for evaluating flow estimates. Specifically, the following datasets were used to compare model output with real-world measurements of the transport system:

* Screenline ‘count’ data. This dataset is derived from linear detectors place perpendicular to the road or cycle path. Each cycle that passes over the detector – typically a hollow tube whose compression emits a pulse of air which is converted into an electrical signal – is recorded as the wheels interact with the detector. The method therefore compares model output at specific points on the travel network with the screenline’s recording of the number of cyclists travelling in either direction.
* CCTV cameras. This dataset is conceptually the same as the screenline data but uses CCTV footage rather than linear detectors to measure the flow of cyclists.
* Road traffic casualty data. This is a georeferenced dataset on the location of bicycle crashes in the study area, called ‘Stats19’ (see Lovelace et al. 2015). As with the model output, the data is highly dispersed. However, it is not necessarily always accurately allocated to the road network. Raster-based methods were developed and tested to make the point-pattern Stats19 data commensurable with the curvilinear outputs of the model.

In addition to these three datasets, we propose methods for using of a fourth input data option: crowd-sourced ‘volunteered geographical information’ (VGI) on cyclist route choice. Because of the paucity of appropriate GPS-derived VGI in the study region, this final methods section is presented in the abstract, based on sample data from another city (London).

In addition to showing that model cross-validation *can* be done for route-allocated transport models, this work is motivated by the lack of literature on model validation in the area of active transport. It is therefore anticipated that demonstration of methods and findings will be useful in future research, in terms of concepts, methods and some empirical findings on model validation. Although the methods were developed for a specific purpose and mode – the evaluation of the Propensity to Cycle Tool (PCT) – they are applicable to any transport or agent-based model that generates estimates of travel flow rates on the route network. There is great potential to build on this work to refine, automate (e.g. via open source software packages) and further test and evaluate the different methods available, and to develop new methods not covered in this paper

The paper is structured as follows: the next section provides a brief description of the PCT in the context of route-allocated transport models. This provides a motivation for the subsequent overview of model validation in the transport modelling literature. Because of the close link between datasets and methods for flow estimate validation, the subsequent section describes the input data and methods together. Datasets derived from screen-line counts, CCTV footage, road traffic casualty records and passively collected ‘volunteered geographical information’ (VGI) are described in turn, alongside the methods needed for them to be compared with the PCT model results.

# The Propensity to Cycle Tool

There are many models using various techniques to study where people choose to cycle to work and for leisure as well as studies which collect data via GPS units and analyse the data to find the most popular routes. (Dill and Carr, 2003; Krizek *et al.*, 2007; Menghini *et al.*, 2010; Broach *et al.*, 2012; Ehrgott *et al.*, 2012; Larsen *et al.*, 2013). However at the time of writing there are not any models which aim to predict the amount of cycling in an area as well as the possible response to cycling becoming more favourable amongst the population. Therefore the PCT is aiming to become a niche tool to help planners.

There are some novel approaches in the literature which study the most where spending will have the most effective spending on the road network (Larsen *et al.*, 2013). Larsen *et al.* (2013) used a Geographical Information System (GIS) approach to assess the cycling network in Montreal, using a form of Multiple Criteria Analysis (MCA) with datasets surrounding Origin-Destination data (OD) for cyclists, car trips, suggested cycling routes from the public, and accident data they evaluated the current system and proposed suitable modifications.

The UK is currently facing a large increase in the number of people choosing to travel by bicycle as there are now 600 million more miles ridden per year in comparison to 1993 (Hollingworth *et al.*, 2015). This rise in cycling has resulted in a rise in incidents, Keep (2013) discovered that there was 32% increase in people being killed or seriously injured while cycling between 2002 levels and 2012 levels. Figure 1 shows the number of people killed or seriously injured over a 5 year period in Leeds, there has been an increasing number of incidents possibly due to the increased levels of cycling (Leeds Data Mill, 2015).

Recent studies into cycling infrastructure prove that spending on infrastructure helps to reduce the amount of people killed or seriously injured while cycling as well as encouraging people to take up cycling (Reynolds *et al.*, 2009; Broach *et al.*, 2012). Unfortunately building cycling infrastructure is not cheap, schemes such as city connected, an infrastructure project linking Bradford and Leeds, is set to cost c£30m while a north-south cycle superhighway through London is costing £160m. Therefore tools such as the PCT can help improve the location of cycling infrastructure projects allowing them to have the most effective cost-benefit scenario.

*Figure 1 the number of people killed or seriously injured while cycling in Leeds (Leeds Data Mill, 2015)*

# Review of methods to validate flow estimates

Methods to validate flow estimates have been evolving overtime either by the technique that is used or the data used to validate the model. Possibly the most use method is to use on-street counts of flows (Donnelly et al., 2012; Munuzuri et al., 2012, 2004). However these methods are questionable with as Donnelly et al. (2012) highlights, indicating that the quality and quantity of counts makes the validation data poor and therefore conclusions on the validity of the model are inconclusive. Methods to improve the data collection of on street counts have been improving with the introduction of auto recognition software (Khan and Raksuntorn, 2001). Auto recognition software allows for continuous data capture of cyclists, therefore improving the quality and quantity issues associated with manual on-street data capture. Camera detection technology can also improve the quantity of counts to provide a more holistic overview of the transport network.

Using Volunteered Geographic Information (VGI) from GPS devices and smart phones real world data can be used to help validate models (Strauss et al., 2015). While the use of such systems does not capture 100% of cyclists it can be inferred to be a percentage of all cyclists and therefore correlate to the model data.

# Methods

### Screenline

Measuring daily flows of traffic is usually carried out using the screenline method (Nicolaisen and Driscoll, 2014). Screenline counts are usually carried out manually via pen and paper records, therefore opening them up to human error when recording data. Recently there has been an advance in computing that allows the recognition of shapes; this technology is now being used to record the number of cyclists passing a point. The added advantage of this automation is that data can be collected for every hour of everyday, whereas previously manual data collection could only be carried for as long as the user was willing to pay to record data. Therefore collecting data manually would lead to a sub-optimal dataset.

*Figure 2 the location of each screenline used to validate the PCT*

In this paper the screenline data was collected in April and May 2014. To try to make the data as uniform as possible counts were only recorded on days inside school term time and when the weather was not adverse. On the day of collection data was collected over a 12 hour period (7am – 7pm). Figure 2 shows the distribution of the survey points.

### Auto-recognition camera data

Technology can now be used to help provide higher resolution count data because cameras can now detect the shape of a cyclist. Data can be recorded for the whole day over the whole year, whereas manual methods such as the screenline method only offer data for part of the year and day.

The data collected in this study has been collected from the locations seen in figure 3. The cameras detect the number of cyclists on an hourly rate as well as the direction that they are heading. The data is freely available on the Leeds Data Mill website <http://leedsdatamill.org/dataset/leeds-annual-cycle-growth->.



*Figure 3 the location of each camera used to validate the PCT*

### Road traffic casualty data

Stats19 data are police records of road traffic casualties. They are therefore biased in many ways and represent the most controversial model comparison dataset. The primary reason that Stats19 data should not be used as a proxy for where people *actually* cycle is that the risk of injury is not uniform across the length of the road network. For this reason, models can be used to assess the geographical distribution of cycling risk. Whilst acknowledging that cycling risk is not uniform, it is important to acknowledge that there is always some risk per unit km cycled, whatever the road conditions (ref). This thinking underlies the experimental use of Stats19 data for model validation presented in this paper.

Rasterisation

Because of the highly distributed nature of the Stats19 data and the fact that it measures a fundamentally different thing from the daily cyclist count estimated by the model, its format needed to be transformed before quantitative comparisons could be made. Building on the methods presented by Larson et al. (2013), it was decided to transform both datasets into raster format. This allowed direct comparison between seemingly incommensurable datasets. In addition, the process of ‘rasterisation’ is computationally efficient, accounts for the majority of ‘NA’ cells where there is no model output or crashes and provides a way to acknowledge spatial inaccuracies in the Stats19 data. The process of rasterization involved 3 stages, resulting in the raster data presented in Figure x:

* Create a raster grid within a bounding box surrounding the study area and a cell size pre-determined by the user (50 m was used initially).
* Rasterise the Stats19 data. This was done initially using a simple count method whereby each record added a value of 1 to the cell it was in.
* Rasterise the model output. This was more complicated due to the curvilinear nature of the lines and the fact that the model result cannot easily be converted into a count per raster cell. To deal with this problem we sampled points on each segment of the road network represented by the network, with the number of points set proportional to the length of the segment (representing cumulative risk) multiplied by the estimated number of cyclists on that part of the road network.

The output from this rasterization process are illustrated in Fig. x below.

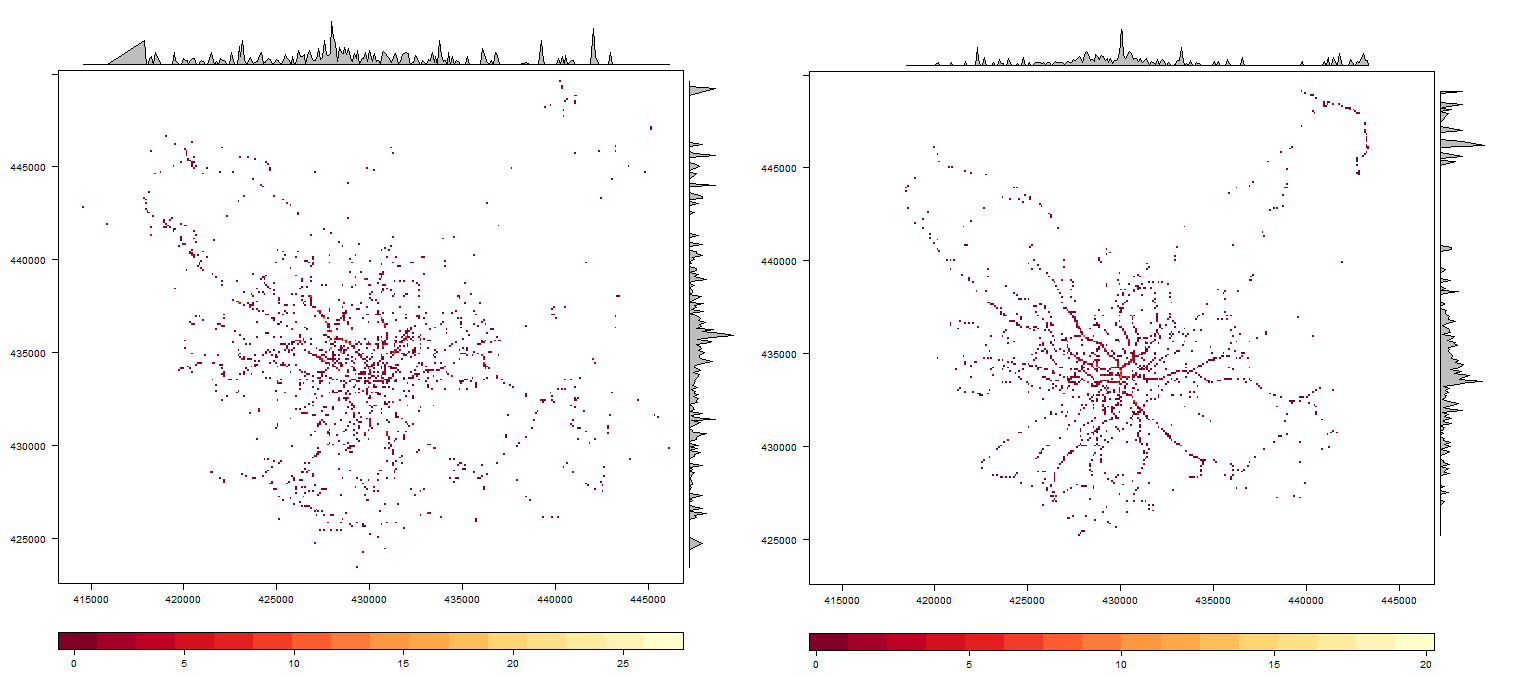


Figure x: the raster output of the Stats19 data (left) and the model output data (right) after the rasterization process.

# Results

## Screenline

Figures 4 and 5 display the data for the screenline counts and the PCT model prediction. Figure 4 shows that the average daily number of cyclists for the screenline data is much higher than the PCT model. However the PCT model only accounts for commute cyclists, therefore this trend is too be expected.

Figure X

Figure X

### Auto-recognition camera data

Fig. 6 illustrates that it is important to take trends in cycling at the local level into account. Note that CCTV footage from point 90810 indicates rapid year-on-year, whereas there are steady and even declining trends for the other locations. Ideally the results should be fitted for the same year of data collection for which the model is calibrated.

Figure X

Figure x

### Road traffic data

After the rasterization process outlined in the previous section, the Stats19 data and the model output were in a commensurable form, with information on the number of casualties and a measure of modelled route-allocated travel activity reported for each identical cell in the raster bounding box. This raster ‘bloc’ contained 184092 cells (274 rows by 337 columns). To ensure that we were not counting Stats19 data outside the geographical bounds of the model output, the results are only reported for each cell on the travel network: 7985 cells, 4.3% of the cells in the raster block (Figure X).

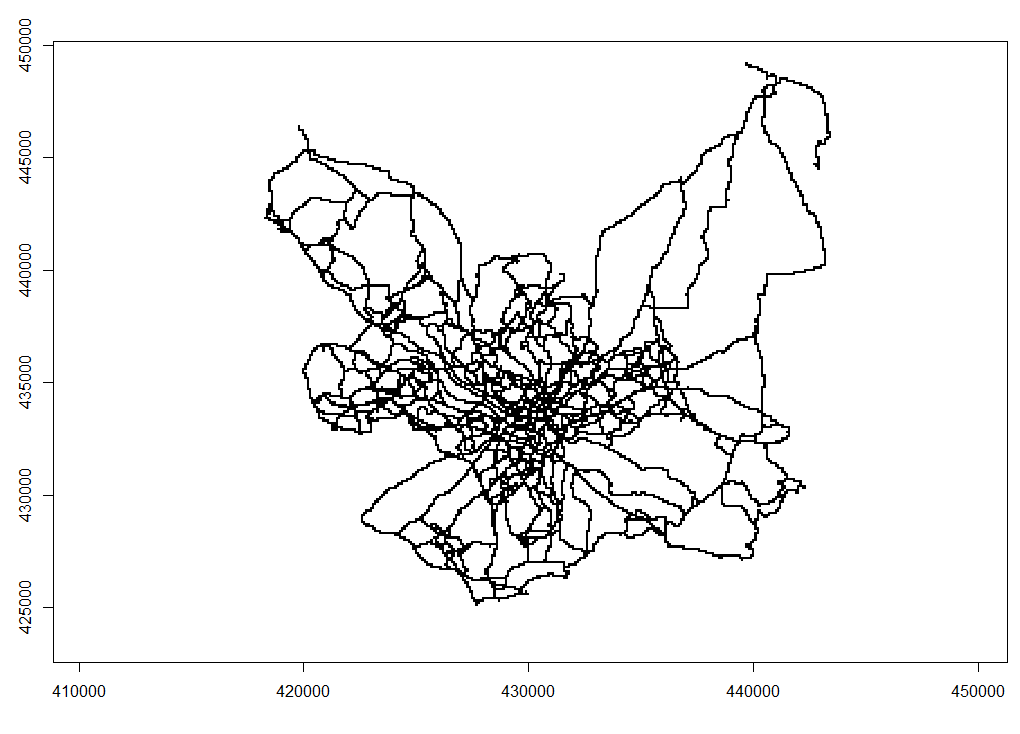


Figure x: the 7985 cells representing the travel network in the model output. The raster output data presented in Fig x was ‘masked’ by this dataset.

Under the ‘base 2011’ scenario almost 15% of the variability in casualty rate per raster cell could be accounted for by the model output (r-squared = 0.141), with a Pearson’s r value correlation value of 0.38 (Figure x).

**Changes over time**

The correlation between the model output and the stats19 data for the ‘Government Target’, representing a future doubling in cycling, was slightly lower (r-squared = 0.116). This provides some suggestion that the model is working as expected.

To assess whether the ‘future scenario’ was more or less representative of recent data, the Stats19 data were split in two, with a cut-off date at the end of 2009. It was found that the correlation between model output for the future scenario declined slightly, with r-squared values of 0.088 and 0.091 for the 2005-2009 and 2010-2014 data respectively. An issue raised by this exercise was the impact of data sparsity on the goodness-of-fit, implying that Stats19 data is too sparse for evaluating the model’s performance at higher temporal resolution (e.g. yearly resolution would only contain 250 collision points). Therefore we caution against drawing conclusions from this exercise.

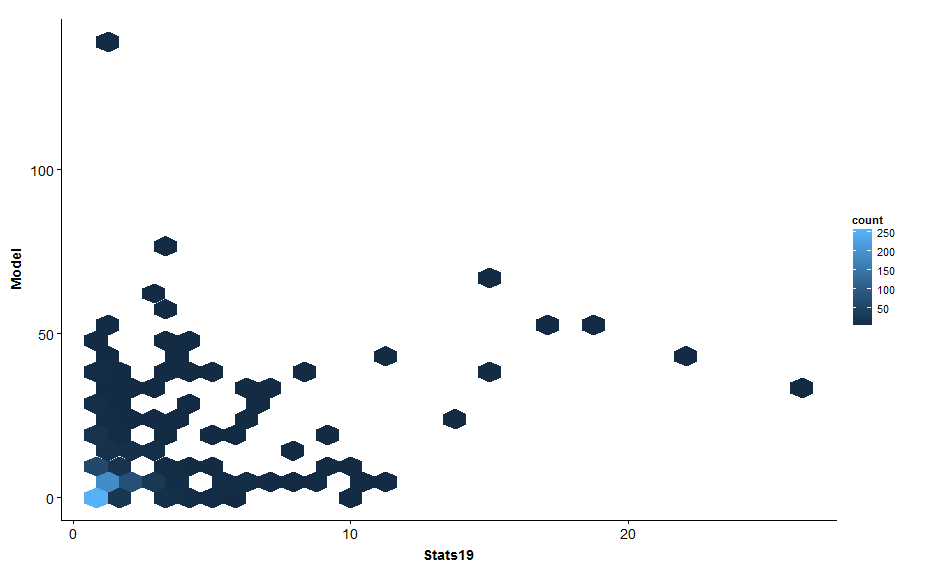


Fig. x Hexagonally binned scatter plot showing the concentration of cell values for the Stats19 data (x axis) and model output (y axis).

Correlation with different cell size

Explore time – only 2011 accidents any better?

Discussion and conclusions

* Main findings of the validation
* Findings from each validation technique
  + Include strength of technique and weakness
* Introduce potential data sources such as strava and discuss the pros and cons of such data
* Rasterisation to standardize diverse datasets + deals with inaccuracy

The screen line data indicated that there was positive correlation between the real world data with an R2 of 0.123. The model predicted less cyclists, but this was down to the fact the model is only counting commute cyclists and not leisure cyclists, therefore the prediction will be less than the real world data. A positive outcome of the study shows that in every location the real world data is closer to the predicted government target when compared against the current scenario. As the model is using 2011 census data and the real world data has been obtained for 2014 this shows there has been overall growth in cycling, this trend has been noted nationally over this period (British Cycling, 2012, 2014).

While the screenline method provided a lot of data to analyse the data was only collected over seven days in April, therefore the data cannot be considered representative of the whole year. Cycling rates tend to fluctuate over the year, particularly in places such as Britain where winters tend to be wet and cold in comparison to the summer season (Lovelace et al., 2015). Unfortunately this data would be unfeasible to collect for a whole year, however carrying out the method for the same period in each season would provide a better outlook on the current rate of cycling in the area.

Data collected from the auto recognition CCTV cameras indicate a similar positive correlation to the screenline data with an R2 of 0.254. When compared with the screenline model estimates for the current cycling rates the auto recognition CCTV data was closer for most of the sites. The model clearly has some issues in certain areas such as site 90421(See Fig x), this site is located next to a filtered permeability access for cycles. At this site the model seems to be observing this as a dead end and predicts routing down other roads instead.

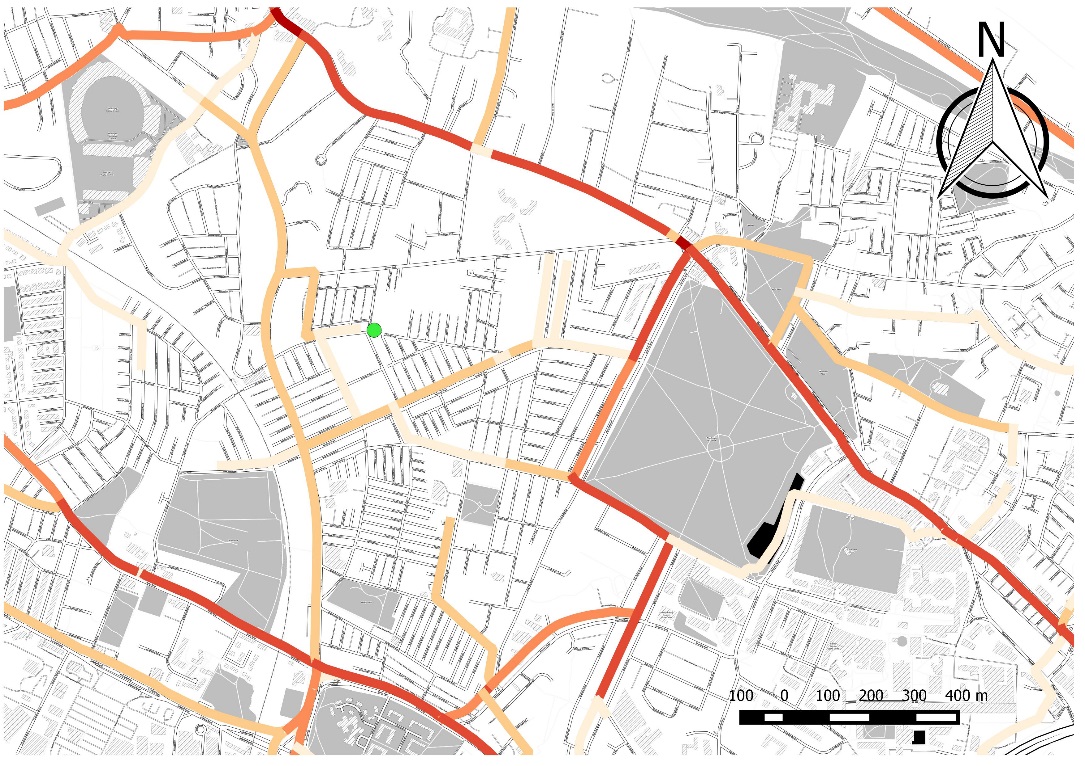


Figure X Camera located next to filtered permeability for cycles

This data is more accurate than the screenline data as it has been collected per hour for every day of the year, therefore providing a true insight into the rate of cycling in the area. Unfortunately it is still not perfect, the placement of the cameras is sporadic and only provides detail of that particular location. Cyclists could well be using other roads and this would be unbeknown. In contrast the screenline method is marginally better has the locations have been carefully selected to include all major roads along known commuter routes.

The stats19 data once again produce a weak positive correlation with an R2 of 0.116, very similar to the other two data sources. Meaning there is a relationship with the number of accidents and the amount of cyclists using the road. This method carries numerous assumptions regarding cycling safety. To imply that this is a causal relationship i.e. the more cyclists using a road will result in more injuries assumes that all roads carry the same risk factor and that cycling safety is uniform across the whole road network. This assumption is simply not true, studies have shown that these facilities help to improve safety (Garrard et al., 2008; Larsen et al., 2013; Reynolds et al., 2009). Similarly areas around junctions are considered to be the most risky areas for cycling (Garrard et al., 2008), therefore an area with numerous junctions is likely to have a greater exposure to risk when compared to a straight section of road.

While the coverage of this data is superior to the other two datasets there is the possibility of under reported incidents, minor bumps will not be included in the dataset as stats19 data only accounts for incidents where the police have been involved. Therefore the data should be regarded as incomplete. The amount of data collected per year is c250 incidents, this is a low number to try and correlate the rate of cycling with the rate of incidents at a small scale across a whole city road network. While the stats19 data is not perfect due to the assumptions and small amount of data it is the most extensive and therefore there is some credibility in the usage of the dataset to estimate the rate of cycling.

In this method the use of rasterization helps to smooth out the data and reduce error. Due to the diverse nature of the stats19 data and the model output any inaccuracy is dealt with by aggregating the data to specific sized cells (Molnar and Julien, 2000). While this is a benefit the raster cell size may affect the results of the study with bigger cells eliminating any trends in the data and smaller cells producing misleading results (Molnar and Julien, 2000).

Amongst cyclists, particularly leisure cyclists, there is a trend to log rides via smartphones and GPS devices and load the rides to specific websites in order to track goals and fitness (Oksanen et al., 2015). Strava, Map My Ride, Endomondo and Bike Citizens are all examples of phone applications and websites where people can load either cycling only or other activities. The data collected by these websites can therefore offer insights into cycling. One particular host, Strava, already make some of this data public via their website. This shows the popular routes in any city of interest (see fig x). Obtaining this data would help to validate the model as it is similar to the stats19 data but is a much larger dataset. For example in 2015 there were over 7 million activities loaded to the site for London, being the largest metropolitan area outside of London, Leeds it would be expected that Leeds would have a large amount of uploaded data as well. This type of data is more commonly known as ‘Big Data’ (Ward and Barker, 2013). Within transport studies Big Data presents a new challenge due to the nature of the data. The skills required to analyses Big Data are different from conventional transport modelling skills as they require a computer science and computer programming background. Besides being a challenge, Big Data presents a new frontier in transport studies, for example the strava insights data for London shows that the most popular day for cycling coincided with a public transport strike. This helps transport modelers understand how people travel in reaction to certain circumstances which previously may have gone undetected as there would be very little data available on how people responded to the event.

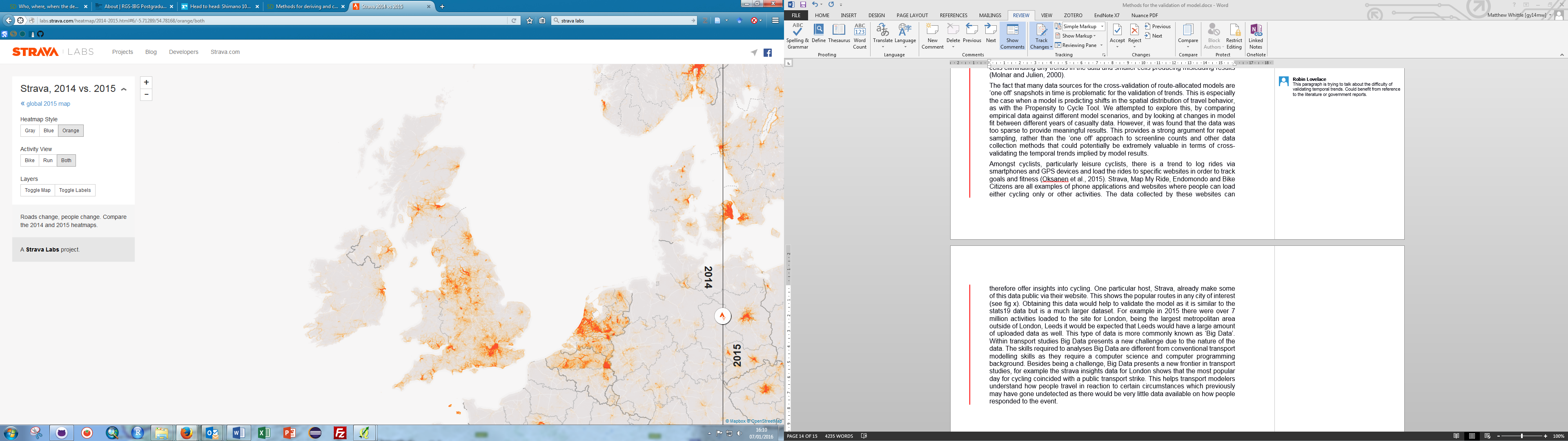


Figure X a map of all the uploaded routes in the UK for 2014 from strava (<http://labs.strava.com/heatmap/2014-2015.html#6/-5.71289/54.78168/orange/both>)

## References

British Cycling, 2014. British Cycling reaches 100,000 membership milestone [WWW Document]. Br. Cycl. URL https://www.britishcycling.org.uk/about/article/20141006-about-bc-news-British-Cycling-reaches-100-000-membership-milestone-0 (accessed 12.9.15).

British Cycling, 2012. Cycling’s legacy is already well underway [WWW Document]. Br. Cycl. URL https://www.britishcycling.org.uk/about/article/bc20120817-about-bc-news-Cycling-s-legacy-is-already-well-underway-0 (accessed 11.30.15).

Donnelly, R., Thompson, R.G., Wigan, M., 2012. Process validation of urban freight and logistics models, in: Taniguchi, E., Thompson, R.G. (Eds.), Seventh International Conference on City Logistics. Elsevier Science Bv, Amsterdam, pp. 400–408.

Garrard, J., Rose, G., Lo, S.K., 2008. Promoting transportation cycling for women: the role of bicycle infrastructure. Prev Med 46, 55–9. doi:10.1016/j.ypmed.2007.07.010

Keim, D.A., 2001. Visual Exploration of Large Data Sets. Commun ACM 44, 38–44. doi:10.1145/381641.381656

Khan, S., Raksuntorn, W., 2001. Accuracy of Numerical Rectification of Video Images to Analyze Bicycle Traffic Scenes. Transp. Res. Rec. J. Transp. Res. Board 1773, 32–38. doi:10.3141/1773-04

Larsen, J., Patterson, Z., El-Geneidy, A., 2013. Build it. But where? The use of geographic information systems in identifying locations for new cycling infrastructure. Int. J. Sustain. Transp. 7, 299–317.

Lovelace, R., Roberts, H., Kellar, I., 2015. Who, where, when: the demographic and geographic distribution of bicycle crashes in West Yorkshire. Transp. Res. Part F Traffic Psychol. Behav. doi:10.1016/j.trf.2015.02.010

Molnar, D.K., Julien, P.Y., 2000. Grid-size effects on surface runoff modeling. J. Hydrol. Eng. 5, 8–16.

Munuzuri, J., Cortes, P., Onieva, L., Guadix, J., 2012. Estimation of Daily Vehicle Flows for Urban Freight Deliveries. J. Urban Plan. Dev. 138, 43–52. doi:10.1061/(ASCE)UP.1943-5444.0000099

Munuzuri, J., Larraneta, J., Onieva, L., Cortes, P., 2004. Estimation of an origin-destination matrix for urban freight transport. application to the city of Seville. Elsevier Science Bv, Amsterdam.

Oksanen, J., Bergman, C., Sainio, J., Westholm, J., n.d. Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data - 1-s2.0-S0966692315001647-main.pdf [WWW Document]. URL http://ac.els-cdn.com/S0966692315001647/1-s2.0-S0966692315001647-main.pdf?\_tid=ec14ce5a-b557-11e5-9082-00000aab0f27&acdnat=1452182673\_aec941975ff93780f9972e6e023b63de (accessed 1.7.16).

Reynolds, C.C., Harris, M.A., Teschke, K., Cripton, P.A., Winters, M., others, 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. Environ. Health 8, 47.

Strauss, J., Miranda-Moreno, L.F., Morency, P., 2015. Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. Accid. Anal. Prev. 83, 132–142. doi:10.1016/j.aap.2015.07.014

Ward, J.S., Barker, A., 2013. Undefined by data: a survey of big data definitions. ArXiv Prepr. ArXiv13095821.

#### Leeds Data Mill (2015). *Cycling accidents in Leeds | ODI Leeds | Publishers*. [online] Available at: http://leedsdatamill.org/dataset/cycling-accidents-in-leeds [Accessed 10 Nov. 2015].