The Impact of Terrain on Cycle Hire Journey Frequency

Soki Kimura

Introduction

The rapid spread of cycle hire schemes impact the transport system in cities worldwide, increasing the share of cycling and public transport usage, lowering greenhouse gas emissions, and improving public health (DeMaio, 2009). Since its first implication in Amsterdam in the 1960s, the system has seen technological and methodological improvements over time. The modern system equipped with telecommunication technology has enabled casual usages between different stations, encouraging short, high-frequency and one-way journeys(Beroud and Anaya, 2012; Beecham, 2015). The London Cycle Hire Scheme (LCHS), known as the Santander Cycles, is one of these systems. In service since 2010 (Li *et al.*, 2019), it has 800 docking stations within central London (Transport for London, 2023b).

Literature Review

The detailed datasets of cycle hire schemes such as LCHS around the world has enabled analysis for cycling behaviour from many aspects. Gebhart and Noland (2014) has found uncomfortable weather conditions including cold weather, precipitation, and high humidity reduces bikeshare usage in Washington DC. Wood, Slingsby and Dykes (2011) explored the visualisation of journeys taken by the LCHS, illustrating two major patterns of usage: leisure usage often seen near Royal Parks and commuting to and from major train stations in peak hours.

Meanwhile, research on the relationship between cycling behaviour and the physical environment is limited (Heinen, Wee and Maat, 2010). A small number of past researches (Rodríguez and Joo, 2004; Parkin, Wardman and Page, 2008) show that hilliness has a negative impact on cycling. The terrain is quantified in different ways, one segmenting the actual routes surveyed, while the other calculates the average percentage of slope using raster data.

Drawing on these two streams, this report has conducted an anlysis on the unexplored relationship between cycle hire scheme usage behaviour and the physical environment.

Research Question

Is there a correlation between the quantity or duration of cycle hire trips and the slope of the physical environment?

Hypothesis

There will be less journeys involving uphill routes. This will be measured from two perspectives:

1. Docking stations located in lower elevations have more arriving journeys than departures, while stations with high elevation being the opposite.
2. There will be a smaller number of trips for origin-destination pairs requiring uphill travel.

Data

Data Sources

We will use the LCHS data that Transport for London (TfL) makes available (Transport for London, 2023b, 2023a). The location of the 800 docking stations is available, along with the name, ID, and the number of docks each station has. The ID and name for start and end stations, date and time, duration, and type of cycle (conventional or e-cycle) for every journey taken by the LCHS is available for analysis.

The spatial data is joined by the LIDAR Composite Digital Terrain Model (DTM) (Environment Agency, 2023) with a resolution of 2 m to extract the elevation. Statistical boundaries and congestion charge boundaries used are from the London Datastore (Greater London Authority, 2014, 2019).

Accessibility by other public transport modes also impact the usage of bicycle hire schemes (Gebhart and Noland, 2014). In London, this is evaluated by an Access Index (AI) published by Transport for London (2015b), which is then categorised into Public Transport Accessibility Levels (PTAL) (Transport for London, 2015a). An AI of 2.5 or below will be classified as the lowest band (1a), while an index above 40 will be classified as the highest (6b).

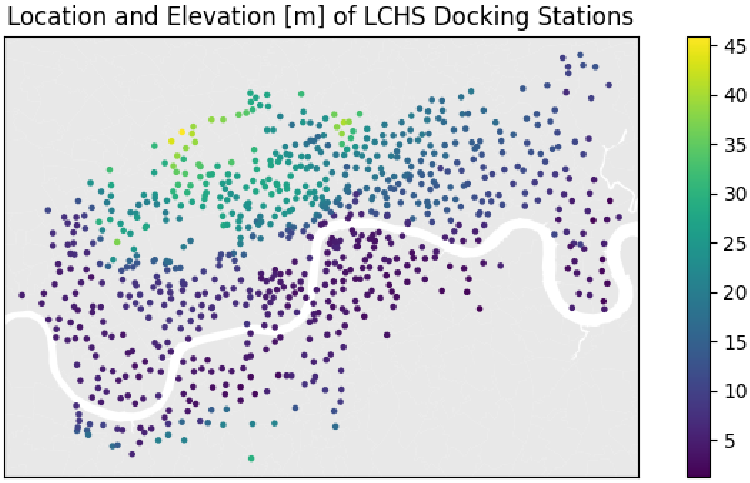

Figure 1: The location and elevation of docking stations in London

Summary of data

The scope of analysis is the journeys taken on the LCHS from December 2022 to November 2023, the most recent one-year period where data is available. There is a total of 8,414,631 journeys taken during this period.

Table 1: Characteristics of journeys made by the LCHS

Statistics	Classic Cycles	E-cycles	Total
Number of Trips	7,808,234	606,397	8,414,631
Mean Travelled Distance [m]	2279.41	3124.62	2340.32
Mean Height Difference of Travel [m]	-0.23	-0.07	-0.22

The negative average in height difference indicate downhill travel is slightly preferred over their uphill counterparts. A one-sample t-test shows this value is significant compared to a mean of 0.

E-cycles have longer distance and smaller difference in height compared to conventional cycles, meaning journeys more physically demanding to the cyclists are taken. This report will attempt to identify the effect of the height difference on the journeys, focusing on journeys made by classic cycles.

Methodology

This research will be conducted in two parts. The first part will consider the usage of individual docking stations, followed by the second part focusing on origin-destination combinations.

Analyse the usage of individual docking stations

First, we will focus on the usage of individual docks to test the first hypothesis. The number of journeys that originated and terminated at each docking station will be counted, and the ratio of departing journeys compared to the arriving journeys (hereinafter "DA ratio") will be calculated as our target variable. This will be compared to the elevation and other characteristics of the station in a regression model. The other dependent variables are the number of ports per station, whether the station is within the central zone, the AI, and the population density of MSOA the station is located within.

Summarizing the data into Origin-Destination pairs

In order to conduct a quantitative analysis on the amount of journeys taken by cyclists, we will summarise the data by grouping them by their origin and destination. MSOAs will be used as the unit of analysis in order to reduce the size of the matrix. There are 160 MSOAs where LCHS operates, therefore the summarised dataset includes 25,600 rows of data. By removing journeys that start and finish within the same MSOA, we have 25,440 rows of data for analysis.

A multiple regression model is conducted to differentiate the impacts of height difference from other explanatory variables summarised by Heinen, Wee and Maat (2010). Other potential explanatory variables in this research are as follows:

- distance of journey (average distance between docking stations)
- population density of MSOA
- access to public transport (average AI value of stations)
- number of docking stations in MSOA
- direction of journey

The characteristics for origin and destination were both considered separately. The direction of journey for the whole dataset will be calculated from the ratio of stations within the central zone.

Analysis of variables

The scatter plot between the distance and the journeys is shown below, along with a log plot and a log-log plot.

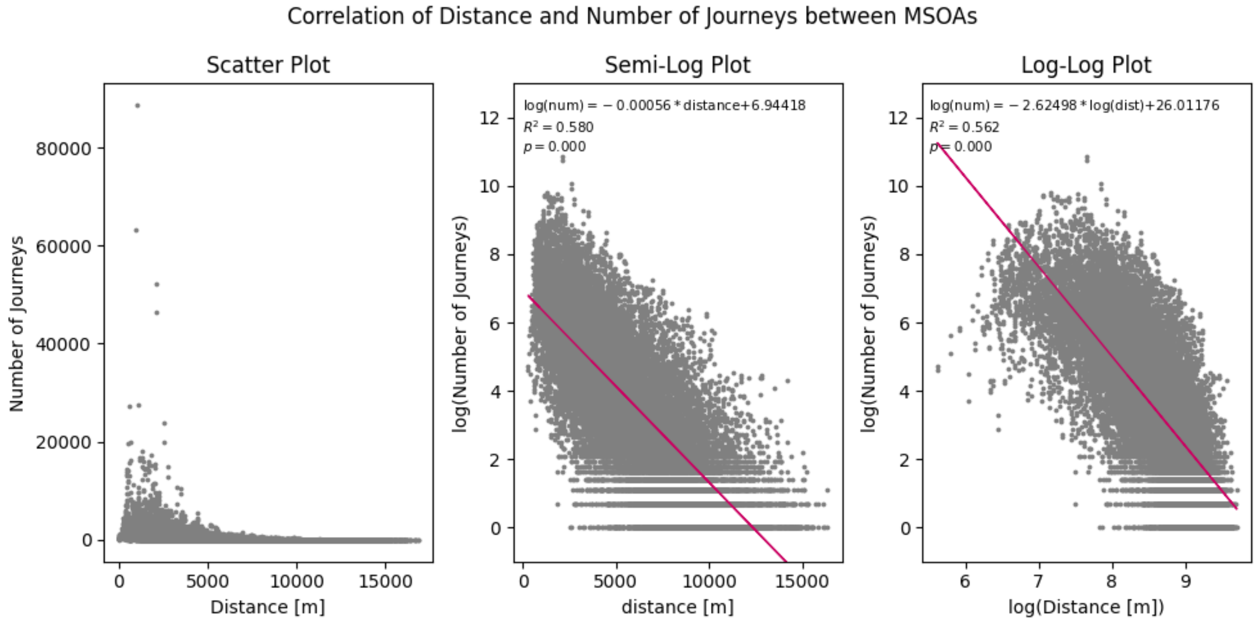

Figure 2: Correlation of Distance and Number of Journeys between MSOAs

The log plot appears to be closest to a linear correlation, while the log-log plot has an upward convex. Thus, the relationship between distance and number of journeys follow an exponential relation. Assuming the other variables have a linear relationship, the model we will use to estimate the number of journeys is as follows:

$$y = \alpha_{\text{dist}} \exp(\beta_{\text{dist}} x_{\text{dist}}) + \sum_i \beta_i x_i + \alpha$$

- x_{dist}: distance of journey
- β_{dist}: coefficient for exponential law (slope of semi-log regression line)
- α_{dist}: constant for exponential law (intercept of semi-log regression line)
- x_i: other explanatory variables
- α: constant for linear regression
- β_i: coefficients for linear regression

By considering $\exp(\beta_{\text{dist}} x_{\text{dist}})$ as one variable `distance_exp` and α_{dist} as its coefficient, the algorithm for a multiple linear regression can be utilised for our proposed model.

Results

Analysis of Individual Points

The relationship between the elevation and cycling activities is shown below. The DA ratio has a linear correlation with the elevation, and is statistically significant explaining 11.0 % of the variance in the DA ratio.

We will consider the multiple linear regression for the DA ratio. The Pearson's correlation matrix below shows no sign of multicollinearity, and alinear correlation between the independent variable and the dependent variables.

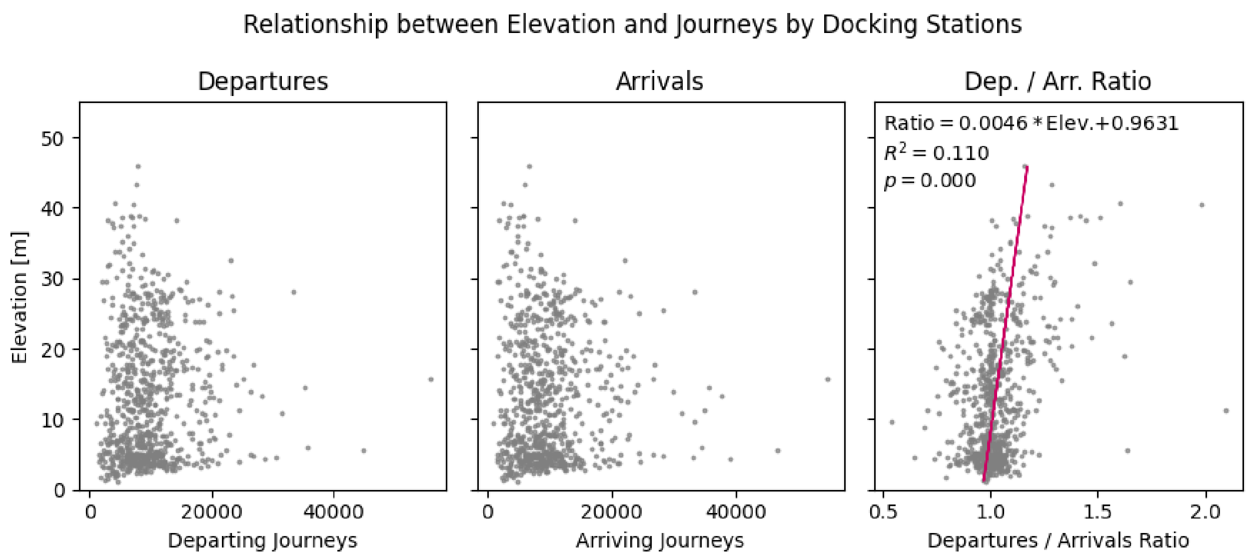

Figure 3: Relationship between elevation and docking station usage

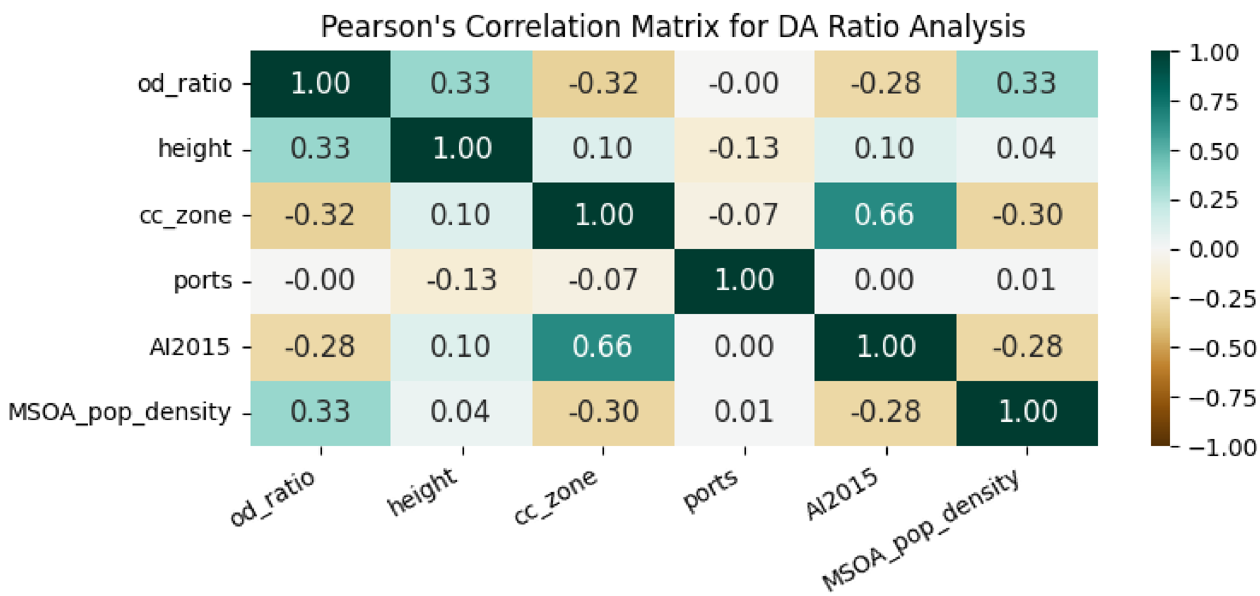

Figure 4: Pearson's correlation matrix for DA ratio analysis

The results of the regression model is as follows.

Table 2: Results of the regression model. Adjusted R-squared value: $R^2 = 0.286$. The height, location, accessibility and the population density are statistically significant, while the number of ports per station is insignificant.

Variable	Coefficient	Standard Error	t	$P > \|t\|$
Constant	0.9246	0.018	52.034	0.000
`Height`	0.0049	0.000	11.783	0.000
`cc_zone`	-0.0567	0.011	-5.268	0.000
`ports`	0.0003	0.000	0.794	0.427
`AI2015`	-0.0005	0.000	-2.757	0.006
`MSOA_pop_density`	4.6302	0.660	7.011	0.000

The factors that lead to more departures than arrivals are the elevation and the population density of the area, while locating in the central zone and the accessibility to public transport lead to more arriving journeys. A 1 m rise in elevation leads to a 0.5 % increase in departures compared to arrivals. stations in the highest PTAL band (AI > 40) will have 2 % more arrivals than departures compared to the lowest band (AI < 2.5).

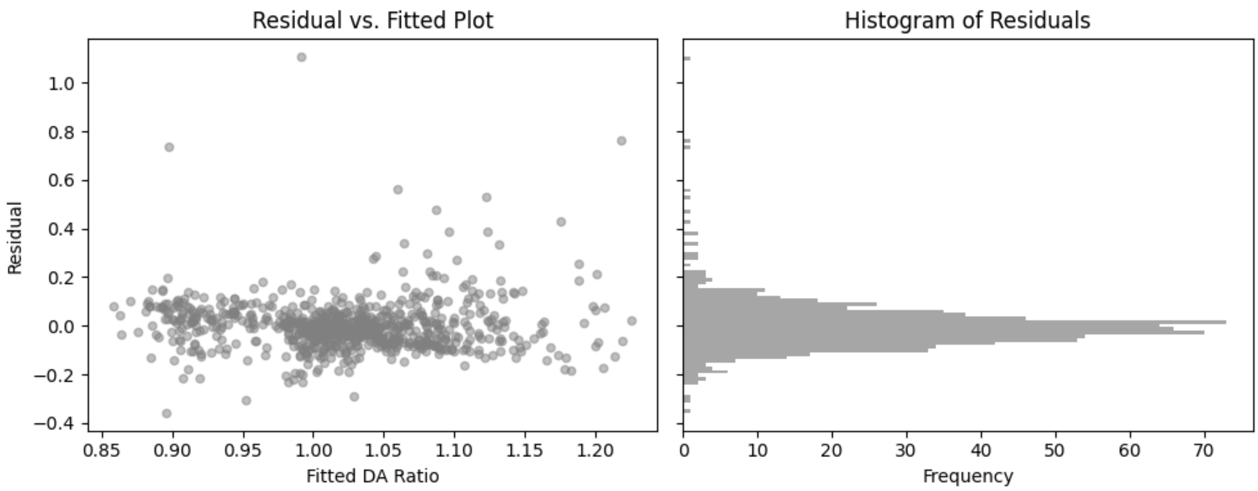

Figure 5: Residual vs. Fitted Plot and the histogram of residuals for the model. The homoscedasticity and the normal distribution of errors can be confirmed.

Analysis of the origin-destination flow

Now, we will analyse the flow of cycles using an origin-destination analysis. The correlation matrix is shown as follows:

There is a collinearity between the number of stations per area, the accessibility by public transport, and whether the docks are in the central area, although this is below the threshold of 5 for a VIF analysis. The results of the regression model is as follows.

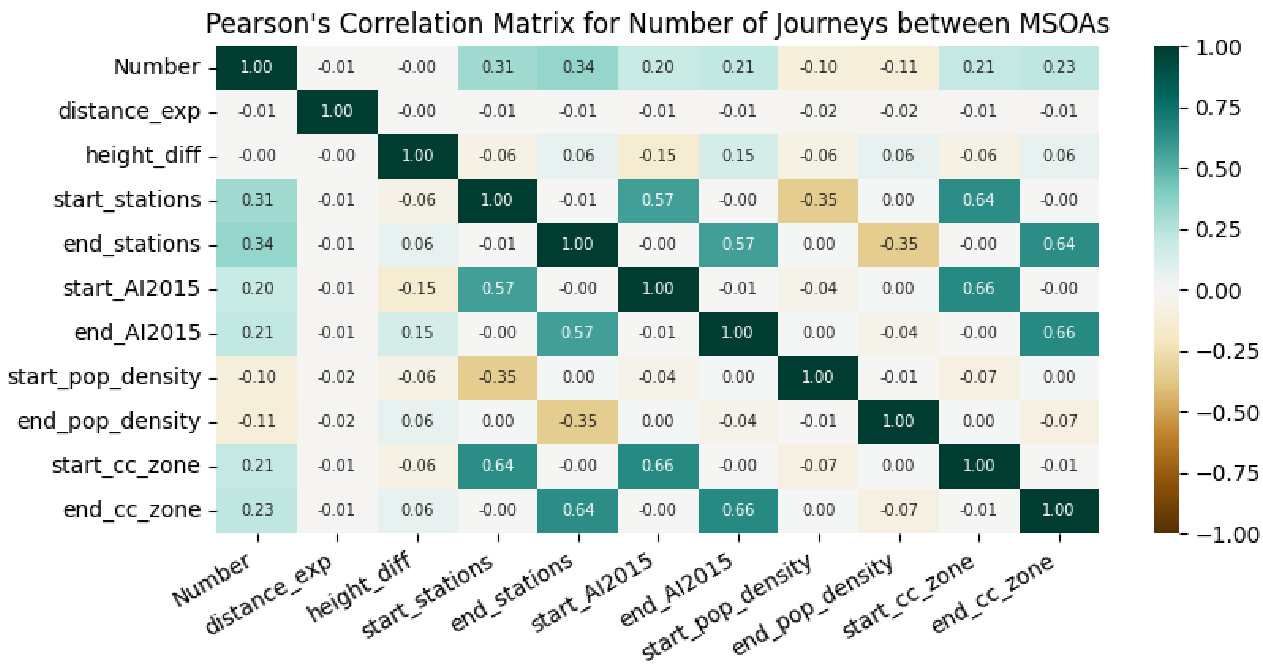

Figure 6: Pearson's correlation matrix for OD analysis

Table 3: Results of the regression model. Adjusted R-squared value: $R^2 = 0.373$. Distance, number of stations and the population density are significant. The difference in height, along with the relationship with the central zone is not a significant factor. AI of destination is statistically significant, while AI of origin is not ($p > 0.05$).

Variable	Coefficient	Standard Error	t	$P > \|t\|$
Constant	-552.5421	28.093	-19.668	0.000
`distance_exp`	2.7091	0.034	80.581	0.000
`height_diff`	-0.4716	0.394	-1.196	0.232
`start_stations`	57.9386	1.517	38.186	0.000
`end_stations`	63.3330	1.517	41.741	0.000
`start_AI2015`	0.6236	0.330	1.887	0.059
`end_AI2015`	0.6825	0.330	2.065	0.039
`start_pop_density`	-0.0026	0.001	-2.569	0.010
`end_pop_density`	-0.0037	0.001	-3.675	0.000
`start_cc_zone`	25.3524	22.087	1.148	0.251
`end_cc_zone`	26.5992	22.087	1.204	0.228

The model shows that there will be more travel between MSOAs if the distance between the MSOAs are smaller, there are more stations within each MSOA, or the population density is smaller.

Discussion

The impact of elevation

From the results, we can conclude that there is some relationship between elevation and LCHS journeys. With focus on the individual stations, stations located in elevated areas see more departures than arrivals, which aligns with previous research and empirical observations where an uphill journey has a negative impact on cycling. On the other hand, the frequency of travel between MSOAs are not impacted by the relative difference in their height.

The first point of discussion is that the relative height difference may not be representing hilliness that negatively impacts cycling, presenting a limitation of this research. Hills encountered en route, the steepness of slope, and the general hilliness of the terrain are not considered, which the methodology to quantify and the correlation with frequency are both potential fields of further research.

The second possibility is that slope may be irrelavant for the overall route selection. London is a flat city with over 75 % of the stations located between 0-20 m in elevation, and the small difference may not be enough to alter mode choice in the macro scale. The difference observed in the scale of individual ports may be caused by choice within the area, where cyclists prefer departing from high stations but return at nearby stations with lower elevation. Further analysis, considering intra-MSOA differences should be conducted to confirm.

Other factors influencing frequency

Through this research, other decisive factors affecting cycling behaviour were discovered. Closer distances between areas saw more journeys between them, which aligns with previous literature. The negative correlation between population density may indicate usage is more frequent in commercial areas than in residential areas. The positive correlation between public transport accessibility for the destination and cycling behaviour indicate bike-and-ride usage is becoming common, which merits public transport users by reducing door-to-door travel time (Martens, 2007). With the AI at the origin not being significant, a *ride-and-bike* behaviour seems to be less common than cycling to public transport.

Interestingly, some factors triggering assymetry among outbound and return trips have been discovered. High population density leads to more departures, while being in the central zone increases arrivals. This may indicate Londoners cycle on their way to the city centre, but use other transport modes on their way back. The reasons, whether it might be the darkness as suggested by Stinson and Bhat (2004), influence of alcohol, or simply not wanting to engage in cycling after a long day, is a potential area for future exploration.

Limitations

This research has not considered all factors that may influence cycling behaviour, such as socio-economic factors, infrastructure, and weather. The variables may be proxies of the actual dominant conditions, in which case my discussions may be drawing incorrect conclusions. The journeys by classic cycles on LCHS may not be the representation of the cycling behaviour in London as a whole, which does not take into consideration private cycles, e-scooters, and other bicycle hire schemes.

Conclusion

In this research, we have created a dataset of journeys taken by the LCHS, with added information on the elevation and accessibility to public transport for origins and destinations of each journey. The correlation between elevation of individual docking stations and departure-arrival ratio has shown uphill travels have negative impact on cycling behaviour, although this pattern could not be observed when considering the macro-scale origin-destination journey frequency. The field of studying cycling behaviour through cycle hire data is growing rapidly (Beecham, 2015), and further research is expected for the better understanding of the effect of the physical environment on cycle hire flows.

Reference

Beecham, R. (2015) 'Using Bikeshare Datasets to Improve Urban Cycling Experience and Research Urban Cycling Behaviour', in Gerike, R. and Parkin, J. (eds). Farnham: Ashgate, pp. 267–283. Available at: https://www.crcpress.com/Cycling-Futures-From-Research-into-Practice/Gerike-Parkin/p/book/9781138546868 (Accessed: 30 November 2023).

Beroud, B. and Anaya, E. (2012) 'Private Interventions in a Public Service: An Analysis of Public Bicycle Schemes', in Parkin, J. (ed.) *Cycling and Sustainability*. Emerald Group Publishing Limited (Transport and Sustainability), pp. 269–301. doi: 10.1108/S2044-9941(2012)0000001013.

DeMaio, P. (2009) 'Bike-sharing: History, Impacts, Models of Provision, and Future', *Journal of Public Transportation*, 12(4), pp. 41–56. doi: 10.5038/2375-0901.12.4.3.

Environment Agency (2023) 'LIDAR Composite Digital Terrain Model (DTM) 2m'. Available at: https://environment.data.gov.uk/survey (Accessed: 29 December 2023).

Gebhart, K. and Noland, R. B. (2014) 'The impact of weather conditions on bikeshare trips in Washington, DC', *Transportation*, 41(6), pp. 1205–1225. doi: 10.1007/s11116-014-9540-7.

Greater London Authority (2014) 'Statistical GIS Boundary Files for London - London Datastore'. Available at: https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london (Accessed: 5 January 2024).

Greater London Authority (2019) 'Ultra Low Emissions Zone 2019 - London Datastore'. Available at: https://data.london.gov.uk/dataset/ultra_low_emissions_zone (Accessed: 5 January 2024).

Heinen, E., Wee, B. van and Maat, K. (2010) 'Commuting by Bicycle: An Overview of the Literature', *Transport Reviews*, 30(1), pp. 59–96. doi: 10.1080/01441640903187001.

Li, H. *et al.* (2019) 'Effects of dockless bike-sharing systems on the usage of the London Cycle Hire', *Transportation Research Part A: Policy and Practice*, 130, pp. 398–411. doi: 10.1016/j.tra.2019.09.050.

Martens, K. (2007) 'Promoting bike-and-ride: The Dutch experience', *Transportation Research Part A: Policy and Practice*, 41(4), pp. 326–338. doi: 10.1016/j.tra.2006.09.010.

Parkin, J., Wardman, M. and Page, M. (2008) 'Estimation of the determinants of bicycle mode share for the journey to work using census data', *Transportation*, 35(1), pp. 93–109. doi: 10.1007/s11116-007-9137-5.

Rodríguez, D. A. and Joo, J. (2004) 'The relationship between non-motorized mode choice and the local physical environment', *Transportation Research Part D: Transport and Environment*, 9(2), pp. 151–173. doi: 10.1016/j.trd.2003.11.001.

Stinson, M. A. and Bhat, C. R. (2004) 'Frequency of Bicycle Commuting: Internet-Based Survey Analysis', *Transportation Research Record*, 1878(1), pp. 122–130. doi: 10.3141/1878-15.

Transport for London (2015a) 'Assessing transport connectivity in London'. Available at: https://tfl.gov.uk/cdn/static/cms/documents/connectivity-assessment-guide.pdf (Accessed: 11 January 2024).

Transport for London (2015b) 'Public Transport Accessibility Levels - London Datastore'. Available at: https://data.london.gov.uk/dataset/public-transport-accessibility-levels (Accessed: 11 January 2024).

Transport for London (2023a) 'TfL Cycling Data'. Available at: https://cycling.data.tfl.gov.uk/ (Accessed: 3 December 2023).

Transport for London (2023b) 'Transport for London Unified API'. Available at: https://api.tfl.gov.uk/BikePoint (Accessed: 29 December 2023).

Wood, J., Slingsby, A. and Dykes, J. (2011) 'Visualizing the Dynamics of London's Bicycle-Hire Scheme', *Cartographica: The International Journal for Geographic Information and Geovisualization*, 46(4), pp. 239–251. doi: 10.3138/carto.46.4.239.