

CASA dissertation using Bookdown

Yifan Wu

CASA0010, MSc Smart Cities Dissertation

Supervisor: Dr Duncan Smith

Repository: <https://github.com/Van-Wu1/>

This dissertation is submitted in part requirement for the
MSc in the Centre for Advanced Spatial Analysis,
Bartlett Faculty of the Built Environment, UCL

Word count: unknown ???? update

2025-08-25

Abstract

This study develops a composite framework to evaluate the cycling environment in London, integrating structural rideability, environmental perception, and network centrality into a unified Cycling Quality Index (CQI). Using multi-source datasets, including OpenStreetMap, environmental indicators, and graph-based network measures, the analysis identifies both local and city-wide constraints shaping ridership potential. The results reveal a pronounced core-periphery gradient: central London suffers from fragmented infrastructure, high traffic stress, and environmental burdens, while outer areas contain isolated yet high-quality segments often linked to green corridors. By combining facility-level attributes with spatial accessibility and environmental quality, the CQI provides a holistic measure that highlights critical gaps in network continuity and equity. The study contributes methodologically by operationalising a multi-dimensional index and empirically by mapping London’s cycling disparities. Findings support evidence-based interventions for sustainable mobility, emphasising the need to strengthen central connections, reduce environmental stressors, and extend high-quality cycling corridors into cohesive metropolitan networks.

Declaration

I, Yifan Wu, hereby declare that this dissertation is all my own original work and that all sources have been acknowledged. It is xxx words in length

Yifan Wu, 24th August 2025

Acknowledgements

(Unfinished...)

Contents

1	Introduction	1
1.1	Background and Objectives	1
1.2	Research Question	2
2	Literature Review	4
2.1	Introduction to Urban Cycling Environment Assessment	4
2.2	Review of Existing Evaluation Frameworks	6
2.2.1	Structural Rideability	6
2.2.2	Environmental Perception	8
2.2.3	Network Performance	10
2.3	Composite Index Approaches (CECI Models)	11
2.4	Conclusion	13
3	Data	15
3.1	General Data	15
3.2	Data for Structural Rideability	16
3.3	Data for Environmental Perception	16
3.4	Data for Network Centrality	17
4	Methodology	19
4.1	Network acquisition and pre-processing	19
4.2	Structural rideability (Index1)	20
4.2.1	Model Introduction	20

4.2.2	Data preprocessing	24
4.2.3	Engineered attributes and factor mapping	24
4.2.4	Base score and multiplicative formulation	25
4.3	Environmental perception (Index2)	26
4.3.1	Subcomponents and normalisation	26
4.3.2	Greenery (GVI)	26
4.3.3	Air quality (NO2)	26
4.3.4	Natural-landscape proximity	26
4.3.5	Sub-index formulation	26
4.4	Network centrality (Index3)	27
4.4.1	Graph representation	27
4.4.2	Betweenness	27
4.4.3	Range-limited closeness	27
4.4.4	Sub-index formulation	27
4.5	Composite index, sensitivity, and reporting	28
4.5.1	Merging and keying	28
4.5.2	Composite scoring	28
4.5.3	Sensitivity analysis	28
4.5.4	Quality control and implementation notes	29
5	Results	30
6	Discussion	31
6.1	Research significance	31
6.1.1	Global development goals	31
6.1.2	Local policy	32
6.1.3	Academic research	32
6.2	Limitations	33
6.3	Transferability	34
7	Conclusion	36

7.1	Recommendations	37
7.2	Future Research Directions	38
	Bibliography	38
	Appendix A Research log	39

List of Figures

4.1	PCA biplot (PC1 vs PC2). Arrows denote variable loadings. LTS (stress level) aligns with PC1; fac4 defines PC2, indicating complementary dimensions.	23
4.2	Multicollinearity check (VIF). Bars show VIF for the structural variables used in diagnostics. LTS = 2.03, base index = 4.63, fac1 = 27.73, fac2 = 22.36.	23

List of Tables

Abbreviations

Term	Abbreviation
Air Quality Index	AQI
Bike Composite Index	BCI
Bike Kilometres Travelled	BKT
Bikeability Index	BI
Capital Bikeshare	–
Central Activity Zone	CAZ
Composite Quality Index	CQI
Confidence Interval	CI
Cycling Environment Composite Index	CECI
Defra Data Services Platform	Defra
Digital Elevation Model	DEM
Digital Surface Model	DSM
Digital Terrain Model	DTM
European Union	EU
Geographic Information System	GIS
Global Positioning System	GPS
Greater London Authority	GLA
Green View Index	GVI
Institute for Transportation and Development Policy	ITDP
Kernel Density Estimation	KDE
Level of Traffic Stress	LTS
Mineta Transportation Institute	MTI
Multiple Centrality Assessment	MCA
Nitrogen Dioxide	NO
OSMnx (Python package for network analysis)	OSMnx
Office for National Statistics	ONS
OpenStreetMap	OSM
Ordinary Least Squares	OLS
Principal Components Analysis	PCA
Remote Sensing	RS
Root Mean Square Error	RMSE
Transport for London	TfL
United Nations / World Health Organization	UN / WHO
Variance Inflation Factor	VIF
“15-minute city”	–

Chapter 1

Introduction

1.1 Background and Objectives

Cycling is increasingly recognised as a cornerstone of sustainable transport strategies worldwide. Compared with private cars, it produces negligible emissions, requires less road space, and contributes positively to public health. For cities such as London, promoting cycling is not only a response to the global climate emergency but also a way to address pressing local challenges such as air pollution, congestion, and social inequalities in mobility access.

Over the past decade, London has launched a series of cycling initiatives, ranging from the construction of Cycle Superhighways and Quietways to the more recent introduction of low-traffic neighbourhoods. These investments reflect growing political and societal support for active travel, as well as recognition that cycling can play a meaningful role in shifting travel behaviour. Nevertheless, the actual quality of the cycling environment remains uneven across the city. Central districts often experience heavy traffic, fragmented facilities, and environmental burdens, while peripheral areas may enjoy greener conditions yet lack coherent connections to employment centres or public transport hubs. This unevenness creates barriers for everyday commuting and limits the potential of cycling to

become a mainstream mode of transport.

Traditional assessments of cycling in London have tended to focus on individual aspects, such as the length of designated cycle lanes or the presence of green corridors. While informative, such isolated measures fail to capture the lived experience of cyclists, who are simultaneously affected by infrastructure design, environmental quality, and network connectivity. What is missing is a holistic and systematic framework that can integrate these multiple dimensions into a coherent measure of cycling quality. Without such an approach, it is difficult for policymakers to identify where interventions are most urgently needed, or to evaluate whether current investments are leading to a more equitable cycling environment.

Against this backdrop, the objective of this dissertation is to develop and apply a comprehensive assessment framework for London’s cycling environment. By integrating infrastructural, environmental, and network-based perspectives, the study aims to provide both a methodological advance and an empirical basis for guiding urban mobility planning. Specifically, the research seeks to move beyond piecemeal evaluations and deliver a composite indicator that can reveal the spatial patterns, disparities, and bottlenecks shaping the potential for cycling across the metropolis.

1.2 Research Question

The overarching research question is:

How can the quality of London’s cycling environment be systematically evaluated by integrating structural, environmental, and network-based dimensions, and what disparities and bottlenecks emerge from such an assessment?

This question is pursued through three interrelated aims:

1. Framework development

To design a composite Cycling Quality Index (CQI) that synthesises three complementary perspectives:

- Structural rideability, capturing infrastructure types, geometry, and traffic-related stressors.
- Environmental perception, accounting for greenery, air quality, and proximity to natural landscapes.
- Network centrality, reflecting the connectivity, accessibility, and functional performance of the road system.

2. Empirical application

To apply this index across Greater London, producing fine-grained evaluations at the road-segment level while also generating aggregated profiles at the borough scale. This dual perspective allows for both micro-level diagnosis of problematic street sections and macro-level understanding of wider spatial inequalities.

3. Interpretation and policy relevance

To identify systematic disparities and critical bottlenecks that hinder cycling uptake, and to generate evidence-based insights for policy and planning. By highlighting where cycling conditions are weakest, and why, the findings aim to inform interventions that can enhance continuity, reduce environmental stressors, and extend high-quality corridors into a cohesive metropolitan network.

By addressing these aims, the dissertation contributes in two ways. Methodologically, it demonstrates how diverse datasets can be operationalised into a single evaluative index, offering a replicable framework for other cities. Empirically, it provides a detailed, multi-dimensional portraits of London’s cycling environment. Together, these contributions support the broader goal of enabling more sustainable, equitable, and resilient forms of urban mobility.

Chapter 2

Literature Review

2.1 Introduction to Urban Cycling Environment Assessment

Cycling has increasingly been considered a key component of sustainable urban transport systems because it is a low-emission, space-efficient mode that can deliver environmental, health and social co-benefits (Yanocha & Mawdsley, 2022). Evidence from international studies suggests that shifting a substantial share of trips to cycling can reduce greenhouse-gas emissions, air pollutants and traffic externalities (Yanocha & Mawdsley, 2022). Estimates from the World Health Organization’s Health Economic Assessment Tool indicate that each kilometre cycled yields a societal benefit of about €0.16, whereas each kilometre driven imposes a cost of roughly €0.15 (Yanocha & Mawdsley, 2022).

Research on the health impacts of cycling also suggests significant gains. A large Danish cohort study of more than 52,000 adults followed for 13 years found that habitual cycling was associated with lower risks of cardiovascular disease, respiratory disease, diabetes and all-cause mortality, and these benefits were not materially diminished by exposure to traffic-related air pollution (Logan et al., 2023). Modelling studies have further shown that, even in highly polluted cities,

the health benefits of regular cycling generally outweigh the risks associated with increased inhalation of pollutants (Logan et al., 2023). In the Netherlands, a country with extensive cycling infrastructure, approximately 27 % of all trips are made by bicycle; Fishman et al. quantified the health benefits of this cycling culture at roughly 6,500 deaths prevented per year and about half a year of additional life expectancy per person (Fishman, Schepers & Kamphuis, 2015).

Despite these findings, cycling participation and infrastructure quality vary widely between cities. Transport for London’s international benchmarking study reported that bicycle mode share is about 1 % in New York City, around 2 % in London, and approximately 40 % in Amsterdam (Transport for London, 2013). This heterogeneity highlights the need for systematic assessments of urban cycling environments to inform targeted investments. The FLOW project, an EU-funded initiative, argues that improvements in walking and cycling infrastructure constitute some of the most promising long-term measures for easing congestion because they are relatively inexpensive and can encourage shifts away from car use (Koska & Rudolph, 2016). Surveys conducted for the same project reveal that experts recognise the potential of walking and cycling measures to reduce congestion but note that such measures are implemented infrequently, indicating an implementation gap (Koska & Rudolph, 2016).

Investments in cycling infrastructure also appear to produce wider societal benefits. Research by the Institute for Transportation and Development Policy reports that separated cycle lanes can reduce cyclist injuries and fatalities even as ridership increases, and that improvements in air quality from modal shifts can lead to further reductions in premature deaths (Yanocha & Mawdsley, 2022). In Washington, DC, analysis of the Capital Bikeshare system found that the presence of bikeshare stations was associated with a 2–3 % reduction in traffic congestion on nearby roads (Wichman, 2016). Such results suggest that well-designed cycling interventions can contribute not only to individual wellbeing but also to urban efficiency and economic productivity.

Overall, the literature indicates that cycling can play a meaningful role in reducing environmental impacts, improving public health and enhancing urban liveability. However, the marked disparities in cycling uptake and infrastructure provision across cities underscore the importance of context-specific assessments. Rigorous evaluation of existing cycling environments and careful identification of infrastructural gaps are necessary to design effective policies and investments capable of realising the potential benefits of urban cycling.

2.2 Review of Existing Evaluation Frameworks

2.2.1 Structural Rideability

A critical dimension of cycling environment assessment is Structural Rideability, capturing physical infrastructure attributes that directly influence cycling comfort, safety, and user inclusivity. This concept aligns with established frameworks such as the Canadian Bikeway Comfort and Safety (Can-BICS) Classification System, which systematically evaluates cycling infrastructure based on safety performance and user comfort across different facility types (Ferster et al., 2023). Similarly, the Dutch CROW Design Manual emphasizes five key principles for effective cycling infrastructure: cohesion, directness, safety, comfort, and attractiveness (CROW, 2016), reinforcing the multidimensional nature of structural cycling environment assessment.

A foundational metric within this domain is the Level of Traffic Stress (LTS) (Mekuria et al., 2012), which categorizes road segments by traffic conditions, lane width, speed, and cycling infrastructure presence. Recent developments have extended LTS applicability through open-source mapping platforms, with Wasserman et al. (2019) demonstrating that OpenStreetMap-derived LTS scores achieve 89.9% accuracy compared to field-validated assessments, facilitating scalable urban mapping applications.

While LTS provides a robust framework for cycling infrastructure assessment,

empirical research has identified opportunities for refinement and expansion of its core assessment dimensions. Recent studies have highlighted the critical importance of specific infrastructure characteristics that form the foundation of comprehensive cycling environment evaluation:

1. Physical Infrastructure and Separation

The design and quality of cycling infrastructure significantly influence safety outcomes and user comfort. Research consistently demonstrates that protected bike lanes and physically separated cycle tracks substantially reduce collision risk, with studies reporting injury rate reductions of up to 50% compared to conventional on-road facilities (Harris et al., 2013; Reynolds et al., 2009). The type of physical separation—ranging from painted lanes to grade-separated infrastructure—creates varying degrees of perceived and actual safety, directly influencing cycling participation across different user groups.

2. Traffic Speed and Volume Characteristics

Vehicle speed and traffic volume represent fundamental determinants of cycling stress and safety. Studies indicate that slower traffic speeds are associated with significantly fewer cyclist injuries, with research demonstrating that combined bike infrastructure and traffic calming measures generate substantially higher cyclist comfort ratings (Sanders et al., 2021). The relationship between traffic characteristics and cycling safety forms a core component of stress-level assessment, with speed limits serving as critical design parameters for cycling infrastructure planning.

3. Network Connectivity and Intersection Design

The continuity of cycling networks and intersection treatments constitute essential elements of comprehensive cycling environment assessment. Research demonstrates that cyclist interactions become more severe and less safe at locations with cycling network discontinuities, highlighting the importance of seamless network connections. Intersection and crossing treatments are

fundamental considerations in LTS evaluation, with studies showing that specialized intersection designs can significantly reduce cyclist-motorist conflicts and improve overall network usability. 4. Spatial Configuration and Design Standards

The geometric design of cycling facilities, including lane width, marking clarity, and spatial relationship to vehicular traffic, influences both objective and subjective safety measures. LTS assessment incorporates the number of lanes, effective vehicle speed, and the presence and type of bicycle facility, creating a comprehensive evaluation framework that accounts for the multidimensional nature of cycling infrastructure quality.

These infrastructure dimensions represent the core components of systematic cycling environment assessment, building upon established LTS principles while providing detailed evaluation criteria for evidence-based cycling infrastructure planning and prioritization.

2.2.2 Environmental Perception

Beyond structural attributes, environmental perception represents a critical dimension of cycling environment assessment, significantly influencing cycling comfort, route choice behavior, and overall cycling participation. This subjective yet quantifiable dimension encompasses multiple environmental factors that shape cyclists' psychological and physiological experiences during cycling activities.

1. Visual Greenery and Aesthetic Quality The visual perception of greenery, measured through the Green View Index (GVI), constitutes a fundamental component of environmental cycling assessment. Studies examining cycling patterns demonstrate that eye-level greenness is positively associated with cycling frequency on both weekdays and weekends (Lu et al., 2020; Yang et al., 2023). Systematic reviews confirm that street greenery promotes active travel through the creation of visually attractive, safe, and comfortable environments (Gascon et al., 2024).

2. Air Quality and Pollution Exposure

Nitrogen dioxide (NO₂) serves as a representative indicator of urban air pollution exposure for cyclists. NO₂ is predominantly transport-related, with most emissions from cars, trucks, and buses, directly reflecting traffic-related exposure conditions (Ma et al., 2024). As a regulated pollutant used to assess ambient air quality in urban environments, NO₂ provides a robust and standardized measure for cycling environment assessment. Studies demonstrate substantial spatial variations in NO₂ exposure along different cycling routes, with measurable implications for both physiological comfort and health safety perceptions (An et al., 2018).

3. Natural Features and Landscape Elements

The presence and accessibility of natural features—including parks, green spaces, and water bodies—constitute essential components of environmental perception in cycling assessment. Research consistently demonstrates that exposure to natural environments generates measurable psychological benefits, with studies showing that green spaces boost serotonin and dopamine levels in the brain, contributing to happiness and well-being (Lee & Maheswaran, 2011).

Evidence indicates that people living in proximity to natural spaces have significantly improved mental health outcomes, with benefits persisting up to three years after establishing residence near greener areas (Nieuwenhuijsen et al., 2017). Furthermore, cross-national studies across 18 countries found that frequency of recreational visits to green, inland-blue, and coastal-blue spaces were all positively associated with well-being and negatively associated with mental distress (Hooyberg et al., 2021). The integration of natural features into cycling route evaluation reflects the understanding that proximity to lakes, parks, and natural landscapes enhances overall cycling experience through psychological restoration and mood improvement.

Collectively, these three environmental perception dimensions—visual greenery (GVI), air quality (NO₂), and natural features—provide a comprehensive

framework for capturing the subjective yet quantifiable aspects of cycling environments that significantly influence user behavior, comfort, and participation decisions.

2.2.3 Network Performance

Urban cycling evaluations also rely extensively on network performance indicators such as connectivity, centrality, and infrastructure density. Network performance assessment focuses on evaluating the ease and convenience of movement within the network without necessarily specifying origin-destination pairs, thus offering a versatile tool for urban cycling assessments.

Contemporary approaches to network analysis have been revolutionized by open-source computational tools. Boeing (2017) developed OSMnx, a Python package that enables comprehensive street network analysis using OpenStreetMap data, allowing researchers to download, model, analyze, and visualize urban networks with unprecedented ease and accuracy. This methodological advancement has facilitated large-scale comparative studies of urban network structures across multiple cities and regions.

The theoretical foundations of urban network analysis were established through graph-theoretic approaches that emphasize topological properties. Porta et al. (2006) introduced the primal graph methodology for urban street network analysis, demonstrating that centrality indices effectively capture the structural ‘skeleton’ of urban areas. Their Multiple Centrality Assessment (MCA) framework provides a metric-based approach that investigates multiple centrality indices simultaneously, offering more comprehensive network evaluation than single-index approaches.

Recent research has expanded network performance assessment to incorporate cycling-specific infrastructure elements. Studies demonstrate that bicycle network connectivity significantly influences cycling behavior, with well-connected networks showing higher usage rates and broader demographic participation

(Buehler & Dill, 2016). Furthermore, the integration of bicycle parking facilities and connections to public transportation nodes represents essential components of comprehensive network performance evaluation, as these multimodal connections significantly enhance cycling network attractiveness and usability (Geurs et al., 2016).

Network density and structural coherence also play critical roles in cycling network effectiveness. Research indicates that cycling networks benefit from both high local connectivity and efficient long-distance connections, with network fragmentation representing a significant barrier to cycling adoption (Lowry et al., 2012). The application of graph-theoretic measures such as betweenness centrality and clustering coefficients provides quantitative frameworks for identifying critical network nodes and assessing overall network robustness for cycling infrastructure planning.

2.3 Composite Index Approaches (CECI Models)

Composite indices for assessing cycling environments have emerged as important tools for integrating multiple dimensions of urban cycling conditions into unified assessment metrics. Galarza-Torres et al. (2020) developed an urban Bikeability Index (BI) to assess and prioritise bicycle infrastructure investments, addressing particularities of roads in urban contexts. Their methodology incorporates infrastructure quality, safety considerations, and accessibility factors to guide investment decisions for improved cyclist accessibility.

Hassanpour et al. (2021) proposed a Bike Composite Index (BCI) consisting of two sub-indices representing bike attractiveness and bike safety, estimated using Bike Kilometers Travelled (BKT) and cyclist-vehicle crash data from 134 traffic analysis zones in Vancouver, Canada. This approach demonstrates the practical application of composite indices in real urban environments, providing actionable insights for local planning decisions.

More recently, Félix et al. (2022) developed a method for identifying potential locations for cycling infrastructure improvements using open data in Paris, addressing the need for simple and effective methods to support decision-making in bicycle planning. Their approach integrates spatial analysis with accessibility metrics to pinpoint areas requiring infrastructure enhancement.

Advanced computational approaches have also been employed, with Szell et al. (2022) proposing a framework for generating efficient bike path networks that explicitly considers cyclists' demand distribution and route choices based on safety preferences. This demand-driven design approach represents a significant advancement in evidence-based cycling infrastructure planning.

However, a significant limitation in current theoretical frameworks is the insufficient integration of cycling perceptual factors and subjective safety assessments. Among the 137 indicators identified in bikeability research, only a few relating to air quality were based on cyclist perceptions, highlighting the predominant focus on objective measures. Castro et al. (2023) emphasise that perceived safety is recognised as a key barrier to cycling, yet its constructs are poorly understood, with most assessments focusing primarily on crash and injury risk rather than broader perceptual dimensions. This gap between objective infrastructure provision and subjective cycling experiences represents a crucial oversight, as perceptual factors significantly influence cycling behaviour and route choices.

Despite these methodological advances, current composite index frameworks still encounter limitations in data integration procedures, objective indicator weighting determination, and computational efficiency for large-scale applications. The insufficient incorporation of perceptual and subjective factors further compounds these challenges, highlighting the ongoing need for more holistic, transparent, and computationally robust methods for comprehensive cycling environment assessment.

2.4 Conclusion

Despite the comprehensive literature base, existing cycling environment assessments remain predominantly fragmented, typically addressing only singular dimensions (structural, environmental, or network-based) (Muhs & Clifton, 2015). This fragmented approach limits the effectiveness of urban cycling environment evaluations, preventing the development of comprehensive and integrative insights crucial for urban planners and policymakers (Giles-Corti et al., 2019). Moreover, few existing frameworks have explicitly addressed the synergistic interactions among different cycling environment factors, particularly the intersection between structural and environmental perception variables.

Addressing these research gaps, this study introduces the Cycling Environment Composite Index (CECI), a comprehensive indicator designed to integrate the critical dimensions of structural rideability, environmental perception, and network performance. The CECI differs from prior studies by explicitly synthesizing diverse but interrelated urban cycling determinants, thereby enhancing evaluation comprehensiveness. While aligning conceptually with the “15-minute city” framework—whereby residents can access their daily needs within a 15-minute walk, bicycle or transit ride from their home—the CECI’s broader analytical scope ensures greater adaptability and utility for various urban contexts (Moreno et al., 2021). By explicitly integrating indicators such as GVI, air quality, LTS, network connectivity, and multimodal infrastructure, the proposed CECI methodology provides planners with a nuanced understanding of spatial disparities in cycling environment quality, thereby enabling targeted interventions.

The integration of structural, environmental, and network performance indicators into a single composite index provides a robust, practical framework for assessing urban cycling environments comprehensively. Although developed and initially demonstrated in Greater London, the flexible design and theoretical robustness of CECI facilitate its potential adaptation and application to other urban contexts.

The proposed index thus contributes to the ongoing development of urban cycling environment evaluation methods, potentially offering useful insights for targeted policy interventions and infrastructure investments to support sustainable urban mobility.

Chapter 3

Data

This study draws on a series of open and authoritative datasets to construct a multi-dimensional database for assessing cycling conditions in London. The datasets cover four broad domains: general geographic data, structural rideability, environmental perception, and network centrality. Together, they capture both physical constraints and qualitative attributes of the urban environment. All datasets are openly accessible, ensuring transparency and reproducibility. Prior to analysis, spatial harmonisation was carried out by projecting all data to the British National Grid (EPSG:27700) and clipping them to the Greater London boundary.

3.1 General Data

General-purpose datasets form the foundation of the research. The road network was extracted from OpenStreetMap (OSM) via Overpass Turbo queries. Filters were applied to retain only roads relevant to cycling, excluding private or non-navigable paths. This cleaned network dataset provides the geometric basis for all subsequent stages of analysis, including the calculation of network centrality and the assignment of slope and environmental attributes. To facilitate aggregation, administrative boundaries were sourced from the ONS Statistical

GIS Boundary Files, comprising both Borough-level and Greater London shapefiles.

3.2 Data for Structural Rideability

To capture the influence of terrain on cycling, slope data were integrated. These data were obtained from the Defra Data Services Platform at a resolution of 5×5 metres. Each OSM-derived road segment was intersected with the slope raster to compute mean gradients. This allows for the quantification of physical impedance to cycling: higher slopes are expected to increase exertion and reduce accessibility, making this dataset critical to the structural dimension of rideability.

3.3 Data for Environmental Perception

Cycling experience is also shaped by environmental and perceptual factors. Three distinct datasets were incorporated:

- **Green View Index (GVI):** The GVI dataset originates from the Treepedia project (MIT, 2015), which employs Google Street View images (captured around 2015) to estimate the proportion of visible greenery along urban streets. While the dataset does not reflect more recent greening interventions or modifications to the built environment, it remains widely used as a proxy for visual landscape structure. Urban greenery at the street scale typically evolves slowly, especially in mature built-up areas such as London, and therefore the 2015 imagery still provides meaningful insights into the distribution of visual greenery. As such, despite its temporal limitations, GVI is considered a valid and informative indicator of the visual environment for this research.
- **Air Quality (NO₂):** Data were obtained from the London Air Quality Network, which provides long-term modelled and forecasted concentrations of key pollutants. Nitrogen dioxide (NO₂) was selected as a representative

measure due to its close association with traffic-related emissions and industrial activity. This dataset, available at a 20-metre grid resolution and projected over a 25-year horizon, ensures both spatial detail and temporal consistency.

- **Natural Features:** Additional environmental attributes were derived from OSM, filtered using specific tags { “leisure”: [“park”], “natural”: [“water”, “wood”, “scrub”], “landuse”: [“grass”, “meadow”] }. This allowed the identification of publicly accessible green and blue spaces, including parks, woodlands, grasslands, and water bodies. Such features are strongly associated with enhanced environmental quality and contribute positively to cycling comfort.

3.4 Data for Network Centrality

Finally, the spatial configuration of the road network was modelled to capture structural importance. Using OSMnx, a graph representation of the OSM road network was constructed. Classic measures of network analysis such as betweenness centrality and closeness centrality were then computed. These measures help identify which roads function as critical connectors or bottlenecks, offering valuable insights into where infrastructure upgrades could generate the greatest improvements in overall cycling accessibility.

In sum, this study combines structural, environmental, and network-based datasets to produce a holistic representation of London’s cycling environment. The OSM-derived road network serves as the unifying backbone, onto which slope, greenery, air quality(NO_2), and centrality indicators are layered. This multi-source framework allows for a nuanced assessment that moves beyond simple distance-based measures, accounting for both the physical feasibility and the perceptual quality of cycling in London.

The following table represents the data sources used in this study:

Dataset	Description	A
OSM Road Network	Extracted and filtered cycling-permissible roads	Li
Borough / London Shp	Administrative boundaries at borough and Greater London scale	Pe
Slope Raster	Terrain slope at 5×5 m resolution	R
Green View Index (GVI)	Percentage of visible greenery from street-level imagery	Pe
NO ₂ Concentration	Modelled nitrogen dioxide levels	20
Natural Features	Parks, woodlands, meadows, water bodies from OSM tags	Pe
Road Centrality Data	Betweenness and closeness derived from OSM road network	Li

Chapter 4

Methodology

Figure X summarises the end-to-end pipeline for constructing a link-level Composite Cycling Environment Index (CECI) for London. The workflow proceeds through four modules: (1) network acquisition and pre-processing, delivering a cleaned street graph and processed attributes; (2) a structural rideability module that assigns facility- and safety-related factors to each link and produces Index1; (3) an environmental perception module that aggregates greenery, air quality, and natural-landscape proximity into Index2; and (4) a network centrality module that computes graph-theoretic measures and yields Index3. The three sub-indices are winsorised and scaled to a common range and then combined into the CECI under a baseline equal-weights scheme, with predefined alternatives used for robustness. All operations are undertaken in EPSG:27700. Parameter values and rule tables appear in the Appendix to support reproducibility.

4.1 Network acquisition and pre-processing

1. Extraction and selection

The base street network is obtained from OpenStreetMap via a two-stage Overpass query. An inclusion-oriented stage captures bicycle-permissible

ways across Greater London. A targeted clean removes motorways and ramps, under-construction segments, and links that explicitly prohibit cycling. Direction-dependent rules are implemented for left-hand traffic (`right_hand_traffic=False`).

2. Topological and geometric cleaning

A light-touch sequence improves graph consistency while preserving OSM geometry. Vertices within 2 m are snapped. Connected components with total length < 100 m are removed. Short dead-end stubs < 20 m are trimmed in three passes. Geometries are clipped to the Greater London boundary. Speed attributes recorded in miles per hour are converted to kilometres per hour ($\times 1.60934$).

3. Processed fields

Pre-processing yields link-level attributes used downstream: `proc_highway`, `proc_maxspeed`, `proc_width`, `proc_oneway`, `priority_road`, `motor_vehicle`. These standardise tagging and encode London’s traffic handedness.

4. Normalisation convention

Continuous metrics generated by later modules follow a common scheme: values are winsorised at the 1st and 99th percentiles, then min–max scaled to $[0, 1]$. Edge betweenness is transformed with \log_{1p} before scaling. Normalisation is applied within each sub-index prior to composite scoring.

4.2 Structural rideability (Index1)

4.2.1 Model Introduction

1. Concept and provenance

The structural module quantifies how road-segment design and operating conditions support everyday cycling. It draws on the open-source OSM-Cycling-Quality-Index

developed for Berlin (SupaplexOSM/OSM-Cycling-Quality-Index), which implements a deterministic, rule-based scoring system over OpenStreetMap data. In that reference model, link-level attributes parsed from OSM—such as functional road class and posted speed, cycleway tags (track/lane/opposite), number of lanes, separation and buffer treatments, surface/smoothness, parking configuration, and basic junction handling—are mapped through factor tables to penalties or bonuses. The Berlin implementation produces a link-level cycling quality score and a separate Level of Traffic Stress (LTS) classification using transparent, auditable rules derived from the above tags and thresholds. The intellectual lineage is acknowledged, and licence and repository details are provided in the Code Availability note.

2. Adaptations

The adaptation used in this research retains the central idea that compounding frictions are represented by multiplicative factors applied to a base score derived from functional class and speed, while re-specifying inputs and thresholds where necessary for the London context.

During model transfer, several source-model assumptions interact with country-specific OSM tagging practices and data completeness. In particular, the factor slot denoted as `fac3` in the reference implementation collapses to a near-constant under London’s OSM coverage (the driving attribute is sparsely or inconsistently recorded, yielding no informative variation at link level). To avoid carrying a non-informative term into the structural score, `fac3` is re-specified as a terrain impedance factor derived from segment-level slope.

A further adaptation concerns the role of Level of Traffic Stress (LTS). In the reference framework, Level of Traffic Stress (LTS) and the Cycling Quality Index (CQI) are produced as parallel outputs. To determine whether LTS can be introduced as an additional structural factor within the composite (`fac5` in `Index1`) without undue redundancy, a link-level multicollinearity check was conducted on the set $\{\text{LTS}, \text{fac1}, \text{fac2}, \text{fac4}\}$. Variables were standardised and analysed

Variable	PC1	PC2	PC3	PC4
stress_level	0.771	0.040	0.429	0.469
fac1	0.744	0.348	0.198	0.535
fac2	0.724	0.005	0.679	0.124
fac4	0.246	0.944	0.051	0.216

using principal components analysis (PCA); segments with incomplete structural attributes were excluded to ensure comparability.

PCA indicates a multi-dimensional structure rather than a single dominant latent factor. The explained-variance ratios are 43.31% (PC1), 25.32% (PC2), 17.18% (PC3), and 14.19% (PC4). Loadings show that LTS aligns chiefly with PC1, while fac4 overwhelmingly defines PC2; fac2 contributes strongly to PC1 and PC3, and fac1 loads on PC1 and PC4. The contrasted signs of LTS and fac2 on PC1, together with the near-orthogonal signal of fac4 on PC2, support the interpretation that these variables capture complementary dimensions rather than a single construct. Summary values are reported in Table X, and the geometry of loadings is visualised in Figure A (PCA biplot).

Variance inflation factors (VIF) provide a separate, regression-oriented perspective. LTS exhibits low collinearity ($VIF = 2.03$), with `base_index` at a moderate level ($= 4.63$), while fac1 and fac2 show higher mutual dependence ($= 27.73$ and $= 22.36$). The low VIF for LTS combined with the PCA results indicates that LTS contributes additional information beyond fac1/fac2/fac4 and can therefore be retained as fac5 in Index1. The pairs plot and correlation matrix provided in the Appendix offer further diagnostic context.

Table X. PCA explained variance and loadings (absolute values). (PC1 = 43.31%, PC2 = 25.32%, PC3 = 17.18%, PC4 = 14.19%)

These diagnostics justify retaining LTS as an independent factor in Index1, alongside facility/width (fac1), speed/class (fac2), and separation/surface/priority (fac4), within the multiplicative structural specification.

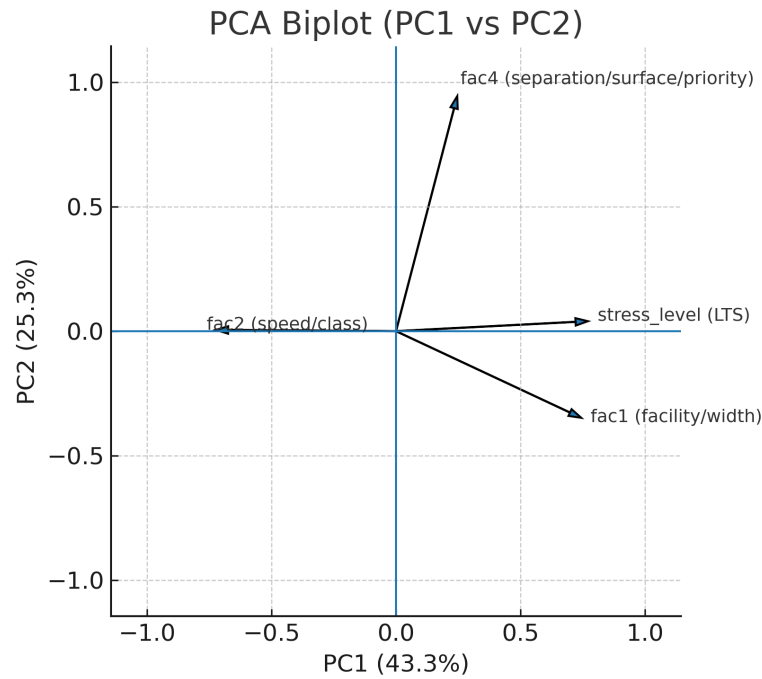


Figure 4.1: PCA biplot (PC1 vs PC2). Arrows denote variable loadings. LTS (stress level) aligns with PC1; fac4 defines PC2, indicating complementary dimensions.

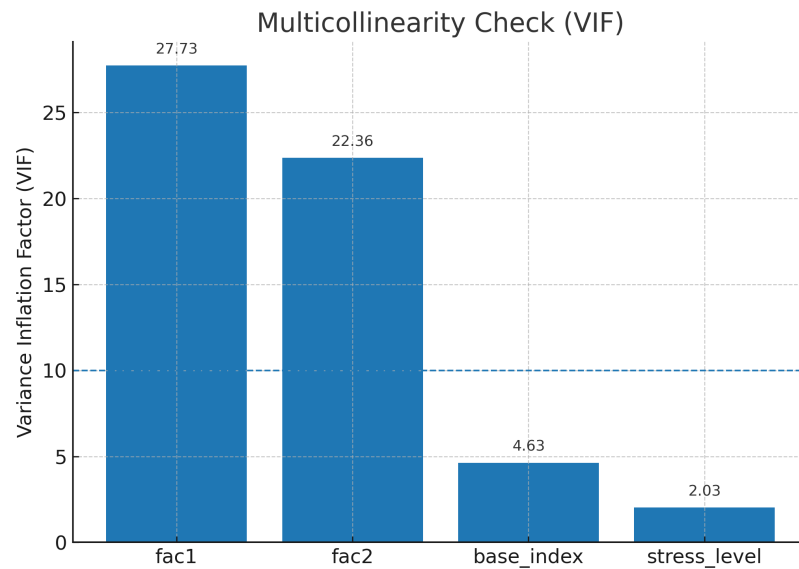


Figure 4.2: Multicollinearity check (VIF). Bars show VIF for the structural variables used in diagnostics. LTS = 2.03, base index = 4.63, fac1 = 27.73, fac2 = 22.36.

Also, there are some adaptations for left-hand-traffic regime, local speed patterns, and tag conventions to fit in with Great London. The structural specification preserves the core Berlin logic while substituting a slope-based `fac3` and justifiably incorporating LTS as `fac5`, yielding a London-tailored yet methodologically transparent module. Full parameter tables appear in the Appendix.

4.2.2 Data preprocessing

A key element of preprocessing is sidepath identification. Because cycle facilities may be encoded on parallel geometries, a sidepath search prevents under-recognition of protection. Each main link is sampled every 100 m, and a 22 m search radius is used to detect cycle-designated paths. Where a sidepath is found, the main link inherits the relevant facility attributes for scoring, reflecting the facilities realistically available along the corridor.

4.2.3 Engineered attributes and factor mapping

Five categories of factors are computed from processed tags and terrain.

1. Facility and width (`fac1`): Functional class together with `cycleway:*` tags and width inform a factor that rewards physically separated facilities and adequate operating width. Where width is unknown, conservative defaults are used (see Appendix). The mapping distinguishes protected tracks, painted lanes, advisory lanes, and mixed-traffic conditions.
2. Speed and classification (`fac2`): Posted speed (`proc_maxspeed`) and `proc_highway` class are combined to reflect exposure to motor traffic. Cut-points are calibrated to London’s distribution of speeds and classes and are interpreted under left-hand traffic for oneway segments.
3. Slope (`fac3`): Terrain impedance is computed by sampling the 5×5 m slope raster along each geometry. Samples are aggregated to a representative

segment value (robust statistic as specified in the Appendix) and then mapped to penalty factors. Missing slope receives $fac3 = 0.9$ as a conservative treatment that avoids unjustified improvement.

4. Separation, buffer, surface, and priority ($fac4$): Presence and width of buffers, type of kerb or median separation, parking configuration, surface/smoothness classes, and priority on major roads contribute additional penalties or bonuses. Directionality is evaluated with respect to left-hand traffic (e.g., parking or buffer on the rider's side). Rule tables provide deterministic mappings from tag combinations to factors.
5. Level of Traffic Stress ($fac5$): An LTS class is assigned from the joint configuration of speed, lanes, facility type, and separation (with junction handling consistent with local practice). Classes are translated into factors using the codebook values reported in the Appendix.

4.2.4 Base score and multiplicative formulation

A base score $Base(highway, maxspeed)$ represents the nominal quality of a segment absent additional penalties or bonuses. The structural sub-index is:

$$Index_1 = Base \times fac_1 \times fac_2 \times fac_3 \times fac_4 \times fac_5$$

followed by linear rescaling to $[0, 100]$. The multiplicative form ensures that simultaneously adverse attributes (e.g., steep gradient, high speed, lack of separation) reduce scores more than any single attribute alone, reflecting the compounding nature of perceived stress and safety constraints. Conversely, coherent design packages (e.g., protected track with buffer on a low-speed street) achieve proportionally high values.

4.3 Environmental perception (Index2)

4.3.1 Subcomponents and normalisation

The environmental module operationalises three link-level components and combines them after normalisation.

4.3.2 Greenery (GVI)

Link values are obtained by buffering each link midpoint by 30 m (or by regular along-link samples with the same radius) and averaging nearby GVI observations. Instances with no observations inside the buffer are recorded for transparency.

4.3.3 Air quality (NO2)

The gridded NO2 surface (cell size 20 m) is sampled every 10 m along each link and averaged. The normalised value is inverted as $(1 - \text{NO2_norm})$ so that higher values consistently denote better conditions.

4.3.4 Natural-landscape proximity

Polygons representing natural features are dissolved by type to avoid double counting. A multi-ring proximity function is computed around each link using the levels (50 m, 0.1), (40 m, 0.2), (30 m, 0.3), (20 m, 0.5), (10 m, 0.7), (1 m, 1.0); weights increase with proximity.

4.3.5 Sub-index formulation

After normalisation to $[0, 1]$, the three components are combined as:

$$Index_2 = 0.5 \times GVI_{\text{norm}} + 0.3 \times (1 - NO_{2,\text{norm}}) + 0.2 \times Natural_{\text{norm}}$$

then rescaled to $[0, 100]$.

4.4 Network centrality (Index3)

4.4.1 Graph representation

A primal, undirected graph is constructed from the cleaned OSM network. Edge weights equal metric length in metres.

4.4.2 Betweenness

Edge betweenness is estimated using approximate sampling with $K = 1200$ source pairs. Values are transformed with \log_{1p} and min-max scaled.

4.4.3 Range-limited closeness

Node closeness is computed within 2 km and 5 km network radii; edge-level closeness is the mean of endpoint values. A multi-scale composite is formed as:

$$C_{\text{multi}} = 0.6 \times C_{2\text{km}} + 0.4 \times C_{5\text{km}}$$

then scaled to $[0, 1]$.

4.4.4 Sub-index formulation

The centrality sub-index balances through-flow importance and local reachability:

$$Index_3 = 0.4 \times Betweenness_{\text{norm}} + 0.6 \times C_{\text{multi,norm}}$$

rescaled to $[0, 100]$.

4.5 Composite index, sensitivity, and reporting

4.5.1 Merging and keying

Outputs are joined by a stable key consisting of the OSM id and a hashed `edge_uid` derived from endpoint coordinates. Geometry is inherited from the structural layer to maintain alignment.

4.5.2 Composite scoring

The composite index is the arithmetic mean of the three sub-indices under a baseline equal-weights assumption:

$$CECI = \frac{Index_1 + Index_2 + Index_3}{3}$$

This choice balances design quality, environmental comfort, and network position at evaluation stage. Length-weighted summaries may subsequently be produced for Borough or MSOA units for descriptive purposes; these aggregations are intentionally separated from model construction.

4.5.3 Sensitivity analysis

To examine robustness to weighting assumptions, three alternative schemes are predefined:

- Scheme A (Index1/Index2/Index3): 0.5 / 0.25 / 0.25.
- Scheme B (Index1/Index2/Index3): 0.25 / 0.5 / 0.25.
- Scheme C (Index1/Index2/Index3): 0.25 / 0.25 / 0.5.

Comparative statistics and maps are reported alongside the baseline.

4.5.4 Quality control and implementation notes

Post-cleaning checks review connectivity and degree distributions. Borough-level sidepath rates are monitored to flag tagging anomalies. Overlay steps (slope, greenery, pollution, natural features) record geometry–attribute counts to confirm coverage before index computation. Normalisation diagnostics document the effect of p1–p99 winsorisation. Processing uses standard Python geospatial libraries (GeoPandas, Shapely, rasterio, OSMnx) and graph tools with fixed random seeds. Parameter files and scripts are available in the project repository.

Chapter 5

Results

Chapter 6

Discussion

Short introduction to the chapter, reviewing the previous chapter and detailing what this one aims to achieve and build upon.

6.1 Research significance

6.1.1 Global development goals

This study contributes to the wider agenda of sustainable development by addressing how active travel can be systematically evaluated and promoted. Cycling is explicitly connected to several United Nations Sustainable Development Goals (SDGs), including SDG 3 on good health and well-being, SDG 11 on sustainable cities and communities, and SDG 13 on climate action. By operationalising a composite measure of cycling quality, the research provides a framework for monitoring progress toward these goals. The findings highlight the need to improve accessibility, reduce exposure to environmental hazards, and ensure equity in mobility provision—key principles that resonate with international objectives for low-carbon, inclusive urban futures.

6.1.2 Local policy

At the local level, the dissertation provides evidence directly relevant to London’s ongoing transport and environmental strategies. The identification of low-quality segments in the Central Activity Zone suggests that current investments in cycling infrastructure have yet to resolve issues of continuity and environmental stress in the city’s busiest areas. The mapping of high-quality but disconnected corridors in outer boroughs also underscores the importance of linking peripheral areas to central employment hubs. These insights can inform initiatives such as the Mayor’s Transport Strategy, the Healthy Streets Approach, and borough-level cycling action plans. By integrating infrastructural, environmental, and network perspectives, the CQI framework offers a tool for prioritising interventions where they can deliver the greatest impact.

6.1.3 Academic research

Academically, this study advances the growing field of urban cycling assessment by bridging methodological divides between infrastructure analysis, environmental perception, and spatial network science. While previous work has often focused on one dimension at a time, the composite index demonstrates the value of synthesising diverse indicators into a single evaluative framework. The approach responds to calls in the literature for multi-scalar, holistic assessments of urban mobility systems and contributes empirical evidence on London, a city that is often studied but rarely through an integrated, composite perspective. Beyond cycling research, the framework also illustrates how geospatial data, environmental metrics, and graph-theoretic measures can be combined in the analysis of urban accessibility.

6.2 Limitations

The estimation of edge betweenness via stochastic source–target sampling imposes an inherent constraint on strict reproducibility. The sample size KK is chosen to balance run-time and stability, yet the procedure remains probabilistic; small numerical deviations and local rank shifts can arise across runs, and may be accentuated by differences in software versions, parallel execution order, or hardware. Under these conditions, the index is most appropriately interpreted through spatial patterns and relative ordering rather than exact equality of segment-level values across replications.

Temporal alignment across inputs constitutes a further limitation. The component datasets are not fully contemporaneous—street-level greenery is inferred from imagery circa 2015, modelled NO_2 fields originate from LAEI 2016 projections, and the OSM network is continuously updated. This asynchrony introduces the possibility of local timing mismatches, particularly where recent greening initiatives, traffic-calming schemes, or redesigns have occurred. As a result, fine-grained differences should not be read as current, time-specific conditions in all locations but rather as a structural and perceptual baseline.

Limitations also arise from cross-city parameter transfer within the structural rideability module. Thresholds and mappings for elements such as effective width, separation, and buffer rules derive from a framework calibrated in Berlin. Although the metadata in both contexts are sourced from OSM via Overpass Turbo, country-specific tagging practices, attribute completeness, and data conventions differ. During transfer, this heterogeneity can increase the frequency of missing or partially specified attributes. To avoid optimistic bias where attributes are absent, a conservative penalty mechanism is adopted that applies uniform downward adjustments under defined missingness conditions; this improves internal consistency but may produce a systematic downward shift in areas with sparser tagging.

The dependence on OSM further entails issues of coverage and tagging consistency. Volunteer-contributed data are uneven across space and over feature types, and fine-grained fields of particular relevance to cycling—cycleway subtype, buffer presence and width, surface and smoothness—may be omitted or inconsistently recorded. Such variation can propagate to factor assignment and to the stability of LTS classification, especially in neighbourhoods where infrastructure is evolving faster than tagging activity.

Design choices embedded in the network analysis also shape results. Centrality is computed on a primal, undirected, length-weighted graph, which abstracts from signal delay, turn costs, grade at junctions, and micro-routing preferences; the selected closeness radii and the approximate betweenness sampler represent defensible but not unique specifications. Similarly, distributional treatment through p1–p99 winsorisation and min–max scaling reduces the influence of extremes but compresses tail variation, with potential consequences for the relative prominence of isolated high- or low-scoring segments.

In addition, the composite weighting schemes encode normative assumptions regarding the relative salience of structural quality, environmental comfort, and network position. Alternative analytical or policy perspectives might reasonably prioritise these dimensions differently, yielding variation in spatial rankings. Taken together, these considerations indicate that the index should be interpreted as a comparative, spatial diagnostic rather than as a definitive, time-specific measure of present conditions.

6.3 Transferability

A further significance of this research lies in its transferability to other urban contexts. The CQI framework is built on widely available datasets and reproducible geospatial methods, making it adaptable to cities beyond London. While local calibration would be necessary to reflect different infrastructural

standards, environmental conditions, or policy priorities, the general approach of integrating structural rideability, environmental perception, and network centrality is applicable to diverse settings. This opens avenues for cross-city comparisons and benchmarking, enabling researchers and policymakers to assess cycling environments in a consistent way. By doing so, the framework supports broader efforts to mainstream active travel within sustainable mobility agendas globally.

Chapter 7

Conclusion

This dissertation has developed and applied a composite framework for evaluating the cycling environment of London, addressing the interplay between infrastructural conditions, environmental perception, and network connectivity. The research was motivated by the need to move beyond fragmented assessments of cycling quality and to deliver a holistic tool that can highlight disparities and guide targeted interventions. Through the construction of a Cycling Quality Index (CQI), the study has provided new insights into the spatial distribution of cycling opportunities and constraints across Greater London.

The findings revealed a marked core–periphery gradient, with central districts characterised by high traffic stress, fragmented facilities, and environmental burdens, while outer areas contained more isolated but higher-quality segments often associated with parks and green corridors. Importantly, the analysis showed that no single dimension is sufficient for evaluating cycling quality: infrastructural improvements may be undermined by environmental pressures, while attractive green routes may fail to deliver connectivity. By integrating these factors into a unified index, the study contributes both methodologically, by operationalising a replicable evaluation tool, and empirically, by providing one of the most detailed portraits of London’s cycling environment to date.

7.1 Recommendations

Based on the findings, several recommendations are offered:

1. Strengthen central network continuity Investments should prioritise filling infrastructural gaps and reducing bottlenecks in central London, where demand is highest but network fragmentation and environmental stress are most acute. Upgrading crossings, junctions, and corridor continuity can deliver substantial improvements in accessibility.
2. Integrate cycling with environmental planning
Air quality management and urban greening strategies should be coordinated with cycling investments. Targeted planting, green buffers, and pollution mitigation measures along key corridors can enhance both the environmental perception and health outcomes for cyclists.
3. Connect peripheral high-quality segments
Outer boroughs often contain strong but disconnected cycling corridors. Linking these segments into longer, continuous routes would enable greater use of peripheral assets and support more equitable access to cycling across the metropolitan region.
4. Embed equity into cycling strategies
Borough-level variation demonstrates that cycling provision remains uneven. Policies should therefore focus not only on expanding total infrastructure but also on ensuring that improvements address underserved communities and neighbourhoods.
5. Adopt multi-dimensional evaluation tools
The CQI framework developed here can support monitoring and prioritisation. Incorporating such composite indicators into policy evaluation would enable decision-makers to capture the full range of factors shaping cycling quality rather than relying on isolated metrics.

7.2 Future Research Directions

While this study provides a comprehensive assessment, further work is required to refine and extend the framework:

- **Dynamic and temporal analysis:** Incorporating peak-hour traffic conditions, seasonal variations in air quality, and temporal changes in greenery would provide a richer understanding of cyclists' experiences.
- **User-centred validation:** Surveys, sensor data, or participatory mapping could be employed to ground-truth the CQI and ensure that it aligns with the perceptions of diverse user groups.
- **Cross-city comparison:** Applying the framework to other urban contexts would test its adaptability and create benchmarks for evaluating cycling environments internationally.
- **Integration with multimodal accessibility:** Future work could explore how cycling quality interacts with public transport nodes, last-mile connections, and emerging forms of micro-mobility.

In conclusion, this dissertation has demonstrated that a multi-dimensional approach is essential for understanding and improving cycling environments. By providing a replicable framework and empirically mapping London's disparities, the research offers both methodological and policy contributions. Ultimately, enabling cycling to play a greater role in London's mobility system will require sustained investment, integrated planning, and evidence-based evaluation tools such as those developed in this study.

Appendix A Research log

subsection

sub sub section

Date	Task
31st May 2020	data search, commenced literature review
7th June 2020	revised literature in the direction of x

Appendix B Proposal