CRANFIELD UNIVERSITY

AVGOUSTA STANITSA

Evidence-based strategies to inform urban design decision-making: the case of pedestrian movement behaviour

SCHOOL OF WATER, ENERGY AND ENVIRONMENT

PhD
Academic Year: 2017 - 2022

Supervisor: Professor Stephen H. Hallett
Associate Supervisor: Dr Simon Jude
April 2022

CRANFIELD UNIVERSITY

SCHOOL OF WATER, ENERGY AND ENVIRONMENT

PhD

Academic Year 2017 - 2022

AVGOUSTA STANITSA

Evidence-based strategies to inform urban design decision-making: the case of pedestrian movement behaviour

Supervisor: Professor Stephen H. Hallett
Associate Supervisor: Dr Simon Jude
April 2022

ABSTRACT

Walking is an essential mode of transportation, and pedestrian movement is a major influencing parameter in city design. Due to the complexity of pedestrian behaviour, new insights concerning the significance of factors affecting walking are challenging to obtain without the use of technology. Furthermore, despite the impact of decision-making in the design of buildings and places, there is currently a limited understanding concerning how urban design decisions are best made. This research aims to "*assess the adoption of, and opportunities deriving from, data-driven innovation techniques in the design of urban spaces, by the analysis of pedestrian movement patterns in urban environments, and to evaluate how the integration of evidence-based strategies can be established in supporting decision-making in relation to future urban designs*".

The research focuses on two groups of stakeholders: Decision-makers in designing buildings and places and End-users undertaking walking activities within urban space. In addressing the aim, a range of research methodologies has been developed and trialled. The work centres on an extended case study concerning a retail high-street locale in London, UK.

This study makes several contributions to the immediate field of urban design research. Firstly, the findings advance the research methods applied to study pedestrian movement in urban environments. Secondly, the results offer real impact in practice by demonstrating the value and importance of adopting data-driven innovation techniques in decision-making processes in urban design via the adoption of a quantitative data-driven, evidence-based methodological framework. Thirdly, the findings support decision-making by presenting a novel methodological framework to assess pedestrian routing in urban environments utilising the classification of pedestrian behaviours and spatial visibility interactions. Finally, this study raises awareness of the critical challenges and opportunities, priorities, and potential development areas for applying evidence-based strategies in informing building and urban design decisions. The research presents a series of recommendations for enhancing data-driven innovation techniques in urban design decision-making processes.

Keywords: urban planning; evidence-based decision-making; data-driven innovations; big data; pedestrian movement; design

ACKNOWLEDGEMENTS

I would like to begin by acknowledging the guidance and support provided by the supervisory team at Cranfield University. I would like to thank my primary supervisor, Professor Stephen H. Hallett, for his patience, continuous contribution, and enthusiasm throughout the work. I would also like to thank Dr Simon Jude who acted as my second supervisor, providing valuable input and reassurance to this work. I would also like to thank my extended supervisory team including Professor Bruce Jefferson and Professor Jim Harris at Cranfield University. The probing questions asked, and feedback provided during my academic reviews, whilst challenging at times, enabled me to grow and develop the scope of the research.

I would like to thank my sponsors DREAM, NERC, ESRC and SNC Lavalin Atkins for funding this research. In addition, I would like to thank my fellow colleagues in SNC Lavalin Atkins, for their ongoing emotional support throughout my studies. I would like to extend my regards to Dr Caroline Paradise who inspired me to apply for this PhD position in the first place and for providing her valuable input in this work. I would also like to thank Paul Medhurst and Dr Arthur Thornton, who supported me in realising this while working in practice. Finally, I would like to express my gratitude to Ruth Hynes, Dr Paul Goodship and Sukhmit Chaggar, whose energy and enthusiasm drove this project forward.

Chapter 5 and 6 of this research was made possible due to the provision of the Wi-Fi location data from The Crown Estate. I would like to thank all those at The Crown Estate who've assisted me in my work and approved the use of their data for the purposes of this research.

Ultimately, I was only able to complete this work, due to the support, and patience of my family and my partner. I'd like to thank my partner Michail, for his continual patience and support, and for being there on my mental and emotional struggle, giving me the motivation and purpose to finalise this work. Furthermore, I'd like to acknowledge the positive encouragement, continuous love and understanding provided by my mother and father, Eleni and Makis. Finally, I would like to thank my brother, Marios, for spending time and effort, while conducting his own PhD studies and working, to provide his input and assistance to this work. Without their love and support I would have not been able to pick myself up and keep moving forward… for that I will be forever grateful.

I wish to dedicate this achievement to my grandparents, Fotis, Thomas, Avgousta and Maria, who acted as role models in life to look up.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

A&MP	Architecture and Master planning discipline
ABM	Agent-based Models
AI	Artificial Intelligence
ARII	Average Relative Importance Index
BD	Big Data
BDAs	Big Data Approaches
BIM	Building Information modelling
CCTV	Closed-circuit television
CH	Calinski-Harabasz coefficient
DDI	Data-Driven Innovation
DMPs	Decision-making processes
DMs	Decision-makers
EFA	Exploratory Factor Analysis
EM	Elbow method
ENG	Engineering Discipline
EPC	Energy performance certificates
EXP	Experienced in DDI
FA	Factor Analysis
GAM	Generalised additive model
GDPR	General Data Protection Regulation
GIS	Geographic Information Systems
GPS	Global Positioning Systems
HPC	High-Performance Computer
ICTs	Information Communication Technologies
IoT	Internet of Things
IT	Information Technology
KMO	Kaiser–Meyer–Olkin test
LD	Lead designer roles

LDel	Listwise deletion
LOS	Levels of service
MAC	Media Access Control
MCAR	Missing completely at random
ML	Machine learning
MoT	Members of the team
NaN	Undefined values
Non-EXP	Non-experienced in DDI
ODI	Open Data Institute
PCA	Principal component analysis
PDPs	Partial dependence plots
PM	Project manager roles
PMD	Project management discipline
POIs	Points of interest
RIBA PoW	Royal Institute of British Architects Plan of Work
rs	Spearman Rank Correlation Coefficient Test
SA	Silhouette Analysis
SLR	Systematic literature review
SPSS	Statistical Package for Social Sciences
UC	User communities
UGC	User-generated content
VGA	Visibility graph analysis
VR	Virtual Reality
WCSS	Within-cluster sum-of-squares

1 INTRODUCTION

1.1 Overview

This research investigates and assesses data-driven innovation techniques in supporting and informing urban design decision-making, providing insights on their practical implementation, challenges, and opportunities. The research examines the role of these techniques by adopting an extended case study undertaken in London, UK, enabling further insights on the classification and assessment of pedestrian movement behaviour.

Chapter 1 introduces the research context and outlines the research question, aim, objectives, and novel contribution. The summary of the chapters is presented, followed by the thesis structure diagram.

1.2 Problem definition

Open urban spaces represent an important asset within cities, providing opportunities for users to engage with their communities and enhance their quality of life (Mouratidis, 2021). Nevertheless, urban growth and development have pressured public urban spaces and, subsequently, their design. The planning and design of a city are influenced by several factors, with mobility being one of the most influential (Mendiola & González, 2021). The act of walking stimulates the complex process of urban design while creation of walkable environments receives great attention due to its various benefits (Schultz, 2014; Choi, 2012). For example, walking has been proven to have significant positive effects on health, while walkable city design can result in energy savings and an enjoyable urban lifestyle (Lawrence & Engelke, 2001; Lee & Moudon, 2004).

Pedestrian movement has attracted much attention across the urban planning disciplines, and a considerable body of research now exists on the spatial orientation of humans within urban environments and the prediction of individuals' movement (Gehl, 2011; Gehl & Svarre, 2013; Zacharias, 2001; Mehta, 2009). However, there remains a lack of objective data in the study of pedestrian movement patterns, and consequently, there is a limited understanding of the role and value of implementing novel technologies in design. In turn, this has resulted in existing design approaches failing to either address user demands effectively or predict their service capacity (Wu, et al., 2017). As a result, understanding of specific aspects of movement, such as urban space recognition and the role of sensorial experience, remains limited, leaving a gap in existing knowledge.

Existing studies have illustrated the necessity of both contemporary data collection and analysis methods, such as objective walking patterns from large-scale monitoring and machine learning analysis techniques, to enable the study of new types of pedestrian movement (Feng, et al., 2021; Lee, 2020). These methods hold the prospect for the collection of new types of pedestrian movement data due to their several advantages, such as increased experimental control and lower implementation costs. They subsequently help to overcome some of the limitations found on traditional approaches (e.g., surveys or observational data collection). Such limitations include the difficulty of recording crowd movements in public spaces via observational data and experiments containing bias (Feng, et al., 2021). The improvement of new technologies, amongst which virtual reality, smartphone monitoring, wearable technologies, etc., has created opportunities to capture new types of information (e.g., large-scale objective movement behavioural data, physiological functions, etc.). Nevertheless, there has been a focus to date limited to understanding pedestrian movement with particular respect to safety and mobility (Karbovskii, et al., 2019; Zhang, et al., 2020; Fernandez-Ares, et al., 2020). This leaves a second knowledge gap; the lack of novel research approaches to study pedestrian behaviour in additional themes. For example, these themes include the exploration of pedestrian movement concerning spatial cognition and the physiological characteristics of the individuals (e.g., emotional responses, motivation, and previous experiences).

A further implication of these technological advancements is that they alter how decision-making processes in urban design are made. Interpreting the information obtained from the application of novel approaches and data types represents a key challenge for decision-makers, adding to the overall complexities of the urban design process (Gandomi & Haider, 2015; Simonet, et al., 2015). Therefore, a third knowledge gap exists in understanding how such approaches may inform urban design decision-making process.

1.3 Research aim, and objectives

This research aims to "*assess the adoption of, and opportunities deriving from, data-driven innovation techniques in the design of urban spaces, by the analysis of pedestrian movement patterns in urban environments, and to evaluate how the integration of evidence-based strategies can be established in supporting decision-making in relation to future urban designs*".

This research critically assesses the role of spatial configuration in the comprehensive approach to pedestrian movement and discusses how it can enhance decision-making process in urban design. Spatial configuration in urban design can be defined as the relative arrangement of parts or elements within the three-dimensional space, entailing aspects of the human spatial experience and behaviour (Hasgül, 2015; Geoghegan, 2001). The influence of urban design attributes in walking patterns is examined by investigating the role of spatial visibility. Spatial visibility is defined as the visible locations in a spatial layout, therefore the area or elements that can be seen by an observer from a given location (Turner, et al., 2001). In addition, the research allows the potential of the integration of new approaches in urban design to be investigated in detail. This research thoroughly demonstrates the spatial visibility impact in pedestrian movement behaviour through the case study approach and highlights where such methods can be adopted in decision-making.

To achieve the stated aim, research objectives (1-3) and sub-objectives, were devised. These are described as follows:

- **Objective 1:** *Undertake a critical analysis of the needs of end-users and decision-makers within the design process for urban systems.*
    - **Sub-objective 1.1:** Assess and evaluate the needs and objectives of stakeholders through a critical review of the current state-of-the-art scientific and practitioner literature.
    - **Sub-objective 1.2:** Identify the influencing factors of the design decision-making process via stakeholder engagement using a structured questionnaire approach.

- **Objective 2:** *Evaluate the potential of data-driven approaches for revealing new insights in geo-temporal human behaviours and their application within the urban planning process in the building design sector.*
    - **Sub-objective 2.1:** Explore the state-of-the-art data collection techniques in the study of pedestrian movement via a systematic literature review, identifying how novel informatics and data-driven technologies inform design decisions and reflecting how humans respond to the built and planned environment and perceive information.
    - **Sub-objective 2.2:** Evaluate the scope and effectiveness of Wi-Fi tracking and Machine Learning techniques for extracting enhanced large-scale information, generating new findings and insights in the study of pedestrian movement via a case study approach.
    - **Sub-objective 2.3:** Appraise data collection, analysis, and visualisation techniques, assessing how they facilitate decision-making and data gathering, and complement traditional urban design approaches.

- **Objective 3:** *Identify the role of evidence-based strategies in supporting robust decision-making in informing urban design approaches in practice.*
    - **Sub-objective 3.1:** Identify the key challenges, opportunities, and priorities for evidence-based strategies in informing building and urban design decisions using semi-structured interviews with stakeholders.

1.4 Methodological approach

This research utilises evidence-based methodologies, while it adopts a range of research approaches to achieve the overall aim. A detailed breakdown of the methodological approach adopted is mapped in Figure 1-1.

A comprehensive systematic literature review is conducted to assess past and current data collection techniques in pedestrian movement behaviour and spatial recognition in urban spaces. Although literature review can be used to theorise the role of novel technologies in upgrading urban areas to meet end-user's requirements, this alone cannot fully explain the impact of spatial configuration on movement patterns. Instead, a methodological framework is required so that the role of novel approaches in decision-making can be identified.

The research also includes a "case study" approach to demonstrate the findings in a "practical" scenario. Two stakeholder groups are considered: *End-users* and *Decision-makers* in designing buildings and places. The thesis focuses on pedestrian movement, recognising the critical impact that walking patterns have on the design of urban environments.

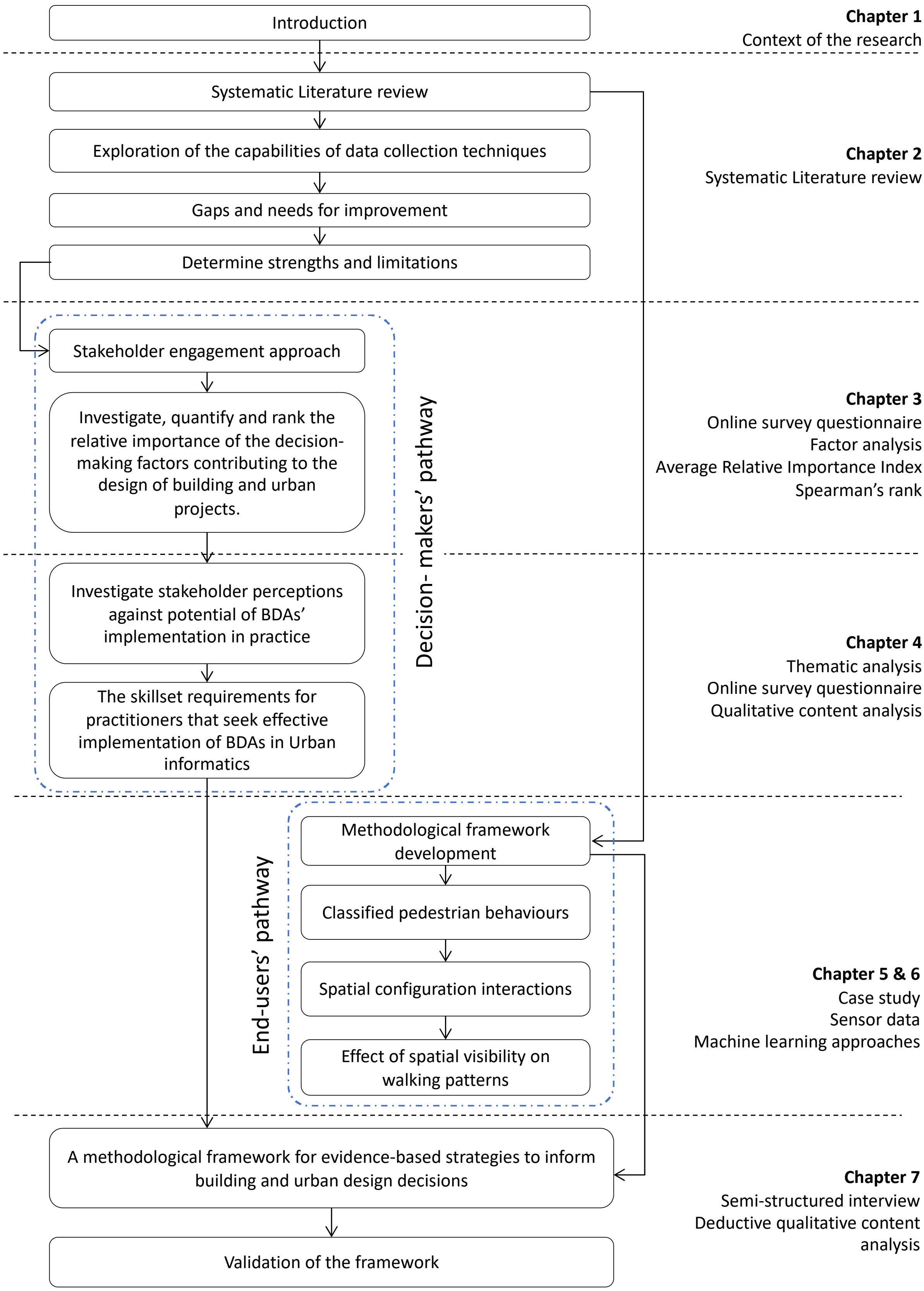


Figure 1-1: Research methodology

1.5 Thesis structure and format

This section provides a summary of each chapter. The work consists of seven chapters, written in manuscript form. The following sections describe each chapter.

Chapter 2: Data collection techniques in the context of pedestrian behaviour in urban spaces: A systematic literature review

Chapter 2 explores the strengths and limitations of data collection techniques in the study of pedestrian movement behaviour, focusing on how pedestrians perceive the urban space around them as they walk through. Four knowledge gaps are identified concerning collection methods: Limited research of pedestrian behaviour in new types of complex set-ups; Collecting and analysing concise high-volume individual objective data sets; Collecting and analysing concise high-volume individual behavioural data sets, and; Representativeness limitations in pedestrian behaviour and urban space captured datasets. Opportunities to cover the gaps identified are discussed.

Chapter 3: Investigating contextual factors influencing decision-making in the design of buildings and places: A survey of stakeholders' perception

Chapter 3 identifies the influencing factors affecting decision-making processes in the design of buildings and places based on stakeholder perceptions. A questionnaire was undertaken to elicit stakeholder priorities when making design decisions. A new means to evaluate the performance of decision-making processes when these are undertaken is provided by developing and applying a quantitative data-driven, evidence-based methodological framework.

Chapter 4: Challenges and applications of Big Data Approaches in the context of Urban Informatics

Chapter 4 identifies the challenges that arise from using Big Data Approaches (BDAs) in the context of Urban Informatics. Big Data is recognised as a new generation of technologies designed to extract value from vast amounts and varieties of data. This chapter is linked to a theme identified in Chapter 3, namely the need for specialised skillsets for effective implementation of BDAs within the design process of buildings and places.

A two-step approach is employed to analyse, synthesize, and present a state-of-the-art of the examined theme. The first step reviews previous literature, focusing on identifying different forms of Big Data challenges and related analytical methods. The second step utilises an online survey questionnaire to investigate stakeholder perceptions of major challenges and the potential of BDAs implementation in research and practice. Four Big Data Approaches have been identified, and the prospect of each to enhance urban design research is discussed.

Chapter 5: Investigating pedestrian behaviour in urban environments: a Wi-Fi Tracking and Machine Learning approach

Chapter 5 is linked to the research gaps identified in Chapter 2. A case study is employed, focussing upon a high pedestrian traffic-dense retail urban area in London, used to investigate and assess pedestrian routing in urban environments. A methodological framework to classify pedestrian behaviours and spatial visibility interaction is provided, utilising machine learning approaches applied to location data derived from Wi-Fi tracking techniques. This approach offers an insightful means to understand pedestrian routing in urban contexts and informs wider wayfinding, walkability, and transportation knowledge.

Chapter 6: Insights into pedestrians' navigation in geo-temporal human behaviours: the case of a retail high-street

Findings from Chapter 5 have helped inform the study described in Chapter 6. This chapter focuses on reviewing key parameters affecting pedestrian movement, identifying the way urban space attributes affect walking activities and types of pedestrian behaviours, and, more specifically, the effect of spatial visibility. Analysis of the impact of the different physical attributes and environmental factors on pedestrian movement has been analysed, providing insights into pedestrian navigation in geo-temporal human behaviours and the urban planning process.

Chapter 7: The challenges of implementing evidence-based strategies to inform building and urban design decisions: a view from current practice

Findings from the previous chapters are synthesised to inform the study in Chapter 7. This chapter explores the key challenges and opportunities, priorities, and potential areas for application of evidence-based strategies to inform building and urban design decisions. These are discussed in relation to key themes, utilising semi-structured interviews undertaken with building and urban design professionals.

Chapter 8: Conclusions and future work

Finally, Chapter 8 summarises research findings in response to the aims and objectives and identifies the limitations encountered. For each study objective, a summary response is offered. Areas of further work are also summarised.

1.6 Research contribution

The novel contributions of this research to the topic area are as follows:

1. This study utilised systematic literature examination to draw from spatial cognition, decision making, and walkability research areas. The work demonstrates that those elements contributing to the sensorial experience of urban spaces should be considered when analysing spatial decision-making and pedestrian routing choices. Further, the work **advances the research methods that can be applied to the study of pedestrian movement in urban environments**.

2. **A quantitative data-driven, evidence-based methodological framework to evaluate the performance of decision-making processes has been developed.** The current analysis captures stakeholder perceptions as to the influencing factors affecting decision-making processes and a quantification as to the way designers make decisions. This research offers a real impact and change in practice by demonstrating the importance of adopting data-driven innovation techniques in decision-making processes in design. The study highlights the need for new metrics, frameworks, and skillsets to improve the ability of designers to extract insights aiming at a better understanding of users' needs.

3. **A novel methodological framework to assess pedestrian routing in urban environments via pedestrian behaviours classification and spatial configuration interactions** to support decision-making. This study utilises an evidence-based method and processes significant amounts of objective movement data to gain insights into pedestrian navigation. More specifically, investigation of the effect of spatial visibility on walking patterns and weather conditions was conducted. The methodological path followed involves the development and application of Machine Learning approaches used with location data derived from Wi-Fi tracking techniques. This research directly contributes to the existing

knowledge surrounding scientific approaches for pedestrian assessment by reinforcing the importance of data-driven environments in supporting improved decision-making in urban design. By understanding the impact of the different physical attributes and environmental factors on pedestrian movement, data-driven design approaches for urban spaces are enhanced, as these factors play an important role on routing choices.

4. This study **raises awareness of the key challenges and opportunities, priorities, and potential areas for evidence-based strategies in informing building and urban design decisions**. Limitations, potential application areas, and perceived constraints for implementing data-driven innovations have been identified, offering opportunities to improve existing design approaches. This contributes in the understanding the implementation potential of strategies that can be applied in decision-making, helping building and urban design organisations make more informed decisions.

1.7 Impact statement

This research focuses on addressing the knowledge gaps identified. Specifically, the study contributes to the knowledge surrounding scientific approaches for pedestrian assessment and modelling, reinforcing the value and importance of data-driven environments in supporting improved decision-making in the design of buildings and their urban space surroundings. Understanding the effects of different physical attributes and environmental factors on pedestrian movement is essential in building data-driven design approaches for urban spaces. This understanding can lead to more accurate and robust models, which decision-makers can apply within the design processes to provide insights into pedestrian navigation in geo-temporal human behaviours and the urban planning process.

The thesis examines pedestrian movement patterns by synthesising the literature and research methods, analytical and visualisation approaches, via a case study approach. This research demonstrates the important relationship between spatial visibility and pedestrian movement behaviours. This is achieved by assessing the walking patterns found in a retail high-street in London. The research further contributes to the literature by offering insights into complex relationships of urban design qualities and street network layout on pedestrian movement.

This research offers real impact and change in practice by demonstrating the importance of adopting data-driven innovation techniques in design decision-making processes via quantitative information, highlighting the need for new metrics, frameworks, and skillsets to understand end-user needs better. Further, it utilises and evolves recognised methods to reveal complex relationships in urban space. Nevertheless, this research differs from previous work by adopting an evidence-based methodology, integrating significant amounts of data to gain insights into pedestrian navigation. One of the key impacts arising from this research is the provision of a framework in which measurable spatial configuration changes and the effect of spatial visibility in dense and crowded urban locations can be recorded robustly. This framework does not rely on potentially biased secondary data sources or the need for long periods to observe changes.

The research outcome is potentially significant in the post-COVID re-evaluation of urban spaces. The findings also offer evidence-based strategies to inform urban design decisions in creating active, walkable environments. Providing a framework in which spatial relationships can be measured would be beneficial to the built environment for planning, designing, and building interventions.

The impact of this research is also reflected through scientific outputs, contributing to the broader field of urban studies (papers; see list of publications in Section 1.8) and engagement with diverse stakeholders (industrial partners: SNC Lavalin Atkins, The Crown Estate, conference presentations, webinars, etc.).

1.8 Papers

- Stanitsa, A., Hallett, S. & Jude, S., In Review. Investigating key factors influencing decision-making in the design of buildings and places: A survey of stakeholders' perception. *Architecture, Structures and Construction.*
- Stanitsa, A., Hallett, S. & Jude, S., In Review. The challenges of implementing evidence-based strategies to inform building and urban design decisions: a view from current practice. *Journal of Engineering, Design and Technology.*
- Stanitsa, A., Hallett, S. & Jude, S., In Review. Investigating pedestrian behaviour in urban environments: a Wi-Fi Tracking and Machine Learning approach. *Multimodal Transportation.*

REFERENCES

Choi, E., 2012. *Walkability as an Urban Design Problem: Understanding the activity of walking in the urban environment (Licentiate dissertation).* [Online] Available at: Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-102182
[Accessed 2021].

Feng, Y., Duives, D., Daamen, W. & Hoogendoorn, S., 2021. Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment,* Volume 187, Article 107329.

Fernandez-Ares, A., Garcia-Sanchez, P., Arenas, M. G., Mora, A. M. & Castillo-Valdivieso, P. A. 2020. Detection and Analysis of Anomalies in People Density and Mobility through Wireless Smartphone Tracking. *IEEE Access,* Volume 8, pp. 54237-54253.

Gandomi, A. & Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management,* 35(2), pp. 137-144.

Gehl, J., 2011. *Life Between Buildings: Using Public Space.* Copenhagen, Denmark: The Danish Architectural Press.

Gehl, J. & Svarre, B., 2013. *How to Study Public Life.* Washington, DC: 2nd ed. Island Press.

Geoghegan, J., 2001. Spatially Explicit Analysis in Environmental Studies. *International Encyclopedia of the Social & Behavioral Sciences,* pp. 14843-14847.

Hasgül, E., 2015. *Space As Configuration: Patterns of Space.* Istanbul, Turkey, University of Derby.

Karbovskii, V., Severiukhina, O., Derevitskii, I., Voloshin, D., Presbitero, A. & Lees, M., 2019. The impact of different obstacles on crowd dynamics. *Journal of Computational Science,* Volume 36, Article 100893.

Lawrence, F. & Engelke, P., 2001. The Built Environment and Human Activity Patterns: Exploring the Impacts of Urban Form on Public Health. *Journal of Planning Literature,* 16(2), pp. 202-218.

Lee, C. & Moudon, A., 2004. Physical activity and environment research in the health field: Implications for urban and transportation planning practice and research. *Journal of Planning Literature,* 19(2), pp. 147-181.

Lee, J. M., 2020. Exploring Walking Behavior in the Streets of New York City Using Hourly Pedestrian Count Data. *Sustainability,* Volume 12, Article 7863.

Mehta, V., 2009. Look closely and you will see, listen carefully and you will hear: Urban design and social interaction on streets. *Journal of Urban Design,* 14(1), pp. 29-64.

Mendiola, L. & González, P., 2021. Urban Development and Sustainable Mobility: A Spatial Analysis in the Buenos Aires Metropolitan Area. *Land,* 10(2), p. 157.

Mouratidis, K., 2021. Urban planning and quality of life: A review of pathways linking the built environment to subjective well-being. *Cities,* Volume 115, Aritcle 103229.

Schultz, H., 2014. Designing large-scale landscapes through walking. *Journal of Landscape Architecture,* 9(2), pp. 6-15.

Simonet, A., Fedak, G. & Ripeanu, M., 2015. Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures. *Future Generation Computer Systems,* Volume 53, pp. 25-42.

Turner, A., Doxa, M., O'Sullivan, D. & Penn, A., 2001. From Isovists to Visibility Graphs: A Methodology for the Analysis of Arcthiectural Space.. *Environment and Planning B,* Volume 28, pp. 103-121.

Wu, H., Liu, L., Yu, Y. & Peng, Z., 2017. Evaluation and Planning of Urban Green Space Distribution Based on Mobile Phone Data and Two-Step Floating Catchment Area Method. *Sustainability 2018,* Volume 10, p. 214.

Zacharias, J., 2001. Pedestrian behavior and perception in urban walking environments. *Journal of Planning Literature,* 16(1), pp. 3-18.

Zhang, P., Li, X.-Y., Deng, H.-Y., Lin, Z.-Y., Zhang, X.-N. & Wong, S.C., 2020. Potential field cellular automata model for overcrowded pedestrian flow. *Transportmetrica A: Transport Science,* 16(3), pp. 749-775.

2 Data collection techniques in the context of pedestrian behaviour in urban spaces: A systematic literature review

2.1 Abstract

Pedestrian movement patterns influence city design and methods of collecting data pertaining to this are key in urban planning. This chapter aims to explore data collection techniques in investigating pedestrian movement behaviour in the study of urban space recognition and the role of sensorial experience on movement patterns, and to determine their opportunities and drawbacks in this context. This paper examines the contexts where these techniques have already been applied via a systematic literature review, analysing journal publications from the Scopus database produced between 2010 and 2021. Four gaps are identified concerning collection methods: Limited research of pedestrian behaviour in new types of complex set-ups; Collecting and analysing concise high-volume individual objective data sets; Collecting and analysing concise high-volume individual behavioural data sets, and; Representativeness limitations in pedestrian behaviour and urban space captured datasets. This study further identifies opportunities to address these gaps utilising a series of emergent technologies.

2.2 Introduction

Walking is an important mode of transportation, and pedestrian movement is a primary influencing parameter in the way urban areas are planned. Walking exposes individuals to the urban space to a greater degree than other transportation modes, such as the car or the train (Loukaitou-Sideris, 2020). Consequently, urban form characteristics, such as aesthetics, feelings of safety, convenience, and comfort, play an essential role in an individual's choice of where to walk (Craig, et al., 2002). Pedestrian movement behaviour is a multi-faceted issue, affecting how people interact with their surrounding environment in a continuous and dynamic way. The significance of the factors affecting walking varies for the different types of walking activities, such as commuting to work or walking for recreational purposes (Gehl & Svarre, 2013). In this chapter, the authors define “pedestrian movement behaviour” as the pedestrian’s movement choices in urban space and where to walk. More specifically, the study focuses on pedestrian movement behaviour in terms of urban morphology; the attributes of which may affect the route choices that pedestrians make.

In the last decades, the desire to encourage walking as a mode of transportation has increased the interest of urban and transport planners (Batty, 1997; Batty, 2001; Liu, et al., 2018). However, implications on understanding movement behaviour might arise due to the large number of variables relating to individual pedestrians and the complexity of the environment surrounding them. Due to the complexity of pedestrian behaviour, new insights are difficult to be obtained without the use of technology. Data collection efforts are critical in understanding the decision-making process and the temporal movement patterns of pedestrians, capturing individual movements in space while undertaking diverse walking activities. This has led to an increased number of studies using a variety of data collection techniques, including field surveys (Natapov & Fisher-Gewirtzman, 2016), pedestrian observations (Özbil, et al., 2015), and interviews (Benachio, et al., 2018; Filingeri, et al., 2017). Although these studies illustrated the benefits of traditional data collection techniques, they have also indicated limitations regarding the types of pedestrian walking behaviour captured and studied by these means (Feng, et al., 2021). For example, the difficulty in recording crowd movements in public spaces via observational data or biased information collected via experimental setups.

The improvement of new technologies has captured the attention in the pedestrian behaviour field, offering the adoption of a wide range of approaches and presenting opportunities to cover existing research gaps. For example, the aesthetic qualities of space remain the most undefined dimensions. Although these are critical influencing factors of pedestrian walking behaviour, their influence is yet to be revealed (Brattico & Pearce, 2013; Brattico, et al., 2013; Chatterjee, 2004). Aesthetic attractiveness is a subjective notion (Storper &

Manville, 2006). It usually includes several factors such as the design of buildings, orientation, and availability of public amenities (Handy, et al., 2002; Harris, 2012). Existing studies have highlighted the lack of incorporation of novel data collection techniques in these research areas, limiting the analysis of pedestrian movement behaviour in complex settings (e.g., high streets).

This chapter aims to explore data collection techniques to investigate pedestrian movement behaviour in the study of urban space recognition and the role of sensorial experience on movement patterns, and to determine their opportunities and drawbacks in this context. This chapter examines the contexts where these techniques have already been applied via a systematic literature review (SLR). The study analysed 227 selected journal publications from the Scopus database covering the period between 2010 and 2021. This study contributes to the existing literature by identifying pedestrian behaviour studies focusing on the data collection techniques used in the analysed contexts. It also highlights critical gaps in employed data collection techniques for pedestrian movement behaviour. This chapter further presents opportunities to bridge the gaps pointed out. The recipients of the findings will be the urban planners, designers, and academics who are interested in delivering urban environments aligned to the end-users needs, utilising novel technologies.

2.3 Pedestrian movement behaviour classification

Data sources for different types of movement behaviour can vary significantly. Therefore, to identify the research focus areas and gaps of existing literature, the different types of behaviour while pedestrians move through space need to be determined. This section provides the pedestrian movement behaviour classifications, which will be used to assess the identified literature.

Batty (2013) noted that urban spaces should be conceived as systems of networks and flows, not only as places in space. Movement results from an alteration of possible forces and circumstances, which amongst these are many physical and psychological factors. These include pedestrian characteristics (i.e., age of the subject, fitness, ethnic origin), environmental surrounding (i.e., weather, terrain condition, safety levels, psychic state of the individual, psychosocial interaction), and traffic flow (i.e., density) (Dridi, 2015; Liang, et al., 2020). Due to this complexity, the approaches suggested to explain pedestrian movement in urban space and the research focus vary among the fields. For example, in health and urban design research, emphasis is placed on the qualities and attributes of urban design, treated in relation to the immediate condition of individual streets. Such studies have sufficiently documented relations among street-level design, pedestrian activity, and environmental correlates of walking (Loukaitou-Sideris, 2020). Research in transportation and planning, though, turned its attention to urban form aspects of walkability (i.e., proximity and distance) and connectivity (directness of travelled route) to reveal their relations with pedestrian movement behaviour (Frank, 2000).

According to Hoogendoorn and Bovy (2004), pedestrian movement behaviour can be classified using a hierarchical structure consisting of three levels being strategic, tactical, and operational level. These levels refer to distinct temporal scales reflecting the choices that pedestrians make. At the strategic level, choice is generic and demonstrates the purpose of the trip. At this level, movement behaviour is usually examined from pedestrian demand perspective (i.e., the volume of pedestrians on a specific road segment or intersection). The tactical level describes pedestrians as a mode/route choice, defining the decision of choosing a route to move from one location to another (Li, et al., 2019). Finally, pedestrians make short-term movement decisions on their way, responding to their surrounding environment at the operational level. This level describes the pedestrian trajectory (i.e., how pedestrians move on the road segment and interact with the crowd). Figure 2-1 visualizes the relative classifications while highlighting the focus areas of this study.

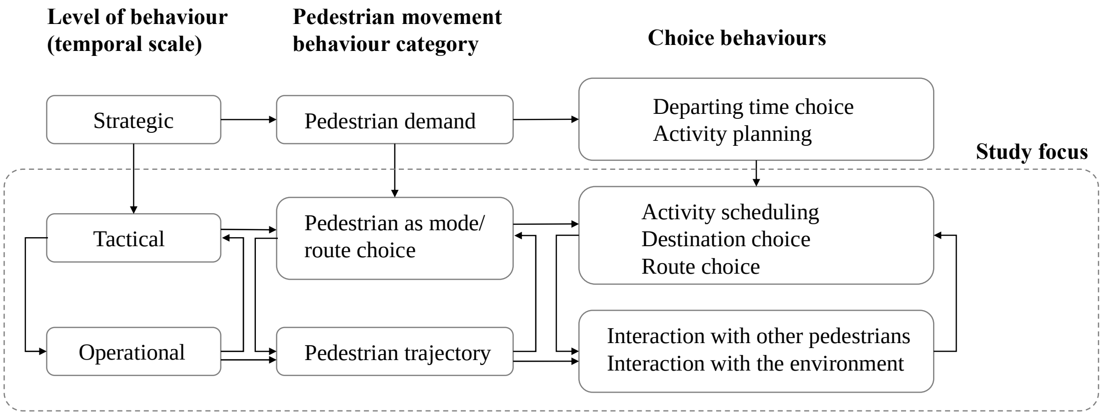


Figure 2-1: Relative classifications of pedestrian movement behaviour and focus areas of the study.

2.4 Materials and methods

A review and critical analysis of peer-reviewed international scientific literature was conducted. A systematic literature review (SLR) was conducted as the preferred research methodology due to the large and complex body of research existing regarding urban space and human behaviour patterns (Stanitsas, et al., 2021).

The first phase of the SLR included tracing and filtering the retrieved documents. Documentation relating to human behaviour in urban spaces was retrieved from an online database (Scopus) in January 2021. A defined set of keywords was applied to article titles, abstracts, and keywords. Due to the high volume of related research, publication date limitations were applied, restricted to those articles published in 2010 or later. Only journal articles published in English were included in the list of potential articles. The keywords used were the combination of terms “pedestrian”, “urban space”, “movement”. This set of references was enhanced employing forward and backward snowballing (Van Wee & Banister, 2016). Therefore, the following keywords were also used for searching: “pedestrian behaviour”, “pedestrian movement” “urban space”; and “field experiment”, “survey”, “virtual reality”, “observation”, “Wi-Fi”, “Bluetooth”, “GPS”, “mobile phones”, “social media”, “IoT”, “data collection” and “wearable technologies”.

Eligible literature included in the review comprised studies addressing: (i) assessment of any of the spatial or personal parameters influencing individual's movement, and (ii) technologies used to assess pedestrian movement and space, including a description of the methodology undertaken (data collection). The initial sample of 1261 documents was further reduced by title, abstract, and availability screening, resulting in a total of 227 journals (Figure 2-2). The second phase of meta-analysis comprised evaluation and synthesis of the documents. A third phase involved a full-text analysis on a further restricted final selection of key documents.

The final analysis included 32 selected documents focusing on pedestrian behaviour's perception or psychological perspective. The categories included constituted part of the following themes: Character of place, Motivation, Emotion, and Previous experience. The selection of these categories was defined to include urban space recognition and the role of sensorial experience on movement patterns in the last decade. The selection excluded non-applicable topics in the research of urban areas or specialised categories, such as personality and intellectual abilities. Figure 2-2 presents the research process, and methodological approach followed in this study. The research was performed as a series of activities organized into phases. Finally, the opportunities and drawbacks arising from their use are discussed.

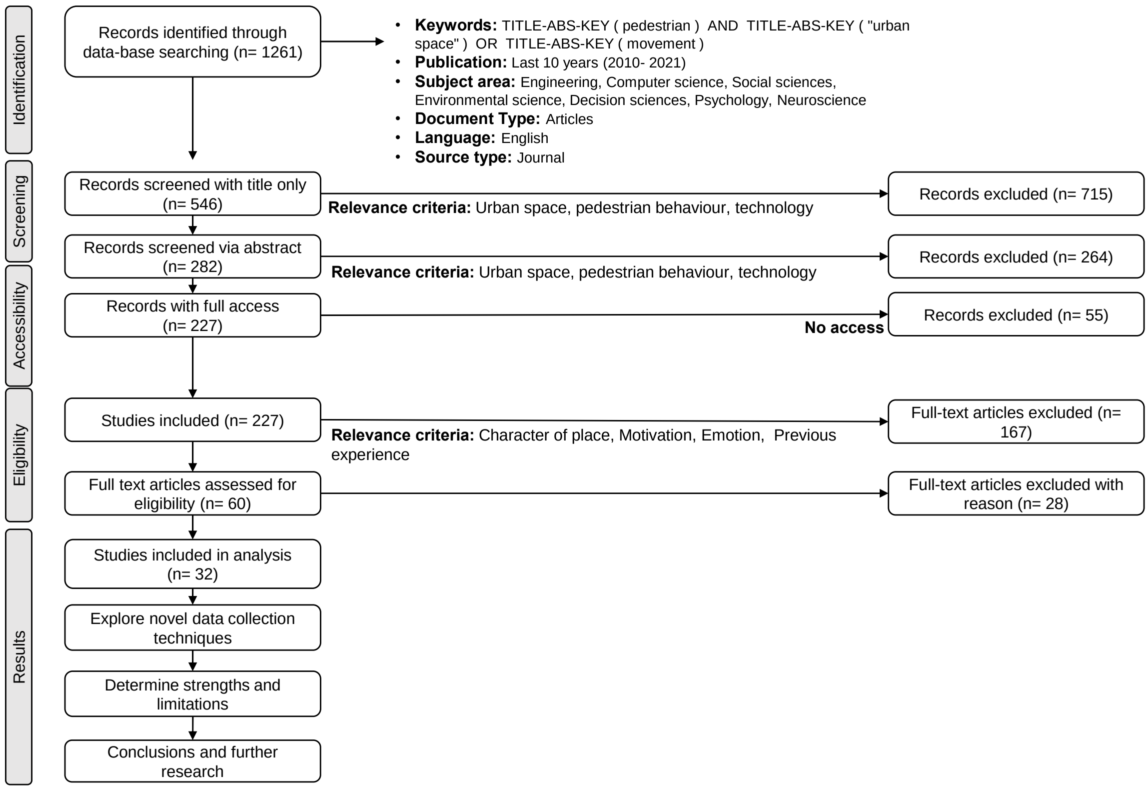


Figure 2-2: Stages involved in the systematic literature investigation

2.5 Exploration of the capabilities of data collection techniques in new types of complex set-ups

Research on urban design and pedestrian movement has been developed steadily in the academic world over the last decade, with an increased interest in the last three years (Figure 2-3). Most of these papers focus on the technical possibilities and drawbacks found in the use of new technologies rather than any revealed insights from collected data concerning pedestrian movement. More specifically, they elaborate on the use and development of novel technologies and their application in predicting pedestrian movement, validation of results, or improved accuracy of data collection techniques (Hanna, 2020; Mills, et al., 2020; Duives, et al., 2020). The studies utilising novel technologies for data collection are centred around crowd behaviour on matters such as safety and mobility (Karbovskii, et al., 2019; Zhang, et al., 2020; Fernandez-Ares, et al., 2020) or space connectivity and configuration (Omer, et al., 2017; Wang & Huang, 2019).

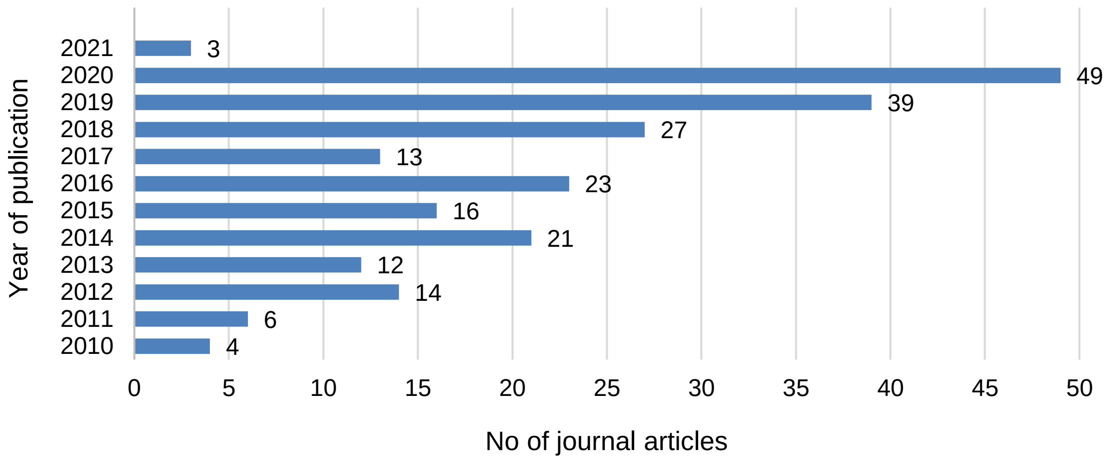


Figure 2-3: Number of publications in the last 10 years (2010- 2021): Urban design and pedestrian movement over time

A recent SLR by Feng et al. (2021) on data collection methods for pedestrian movement concluded that the connection between diverse choice dimensions and the reasoning behind these choices remains as yet little understood. Most of the studies using monitoring techniques to capture objective datasets via novel technologies are applied on a strategic level, addressing research areas such as destination choice and identification of activities. Therefore, the collection of concise behavioural data sets concerning pedestrian choices is still lacking.

This section presents a comprehensive review of studies using data collection techniques to study pedestrian movement behaviour. The authors examine the data sources for both the dependent variables (pedestrian movement) and the independent variables (influencing factors, i.e., streets and other built environment characteristics). There are five frequently adopted data collection techniques, namely: traditional approaches, controlled experiments via the use of virtual reality (VR), wearable technologies, large-scale monitoring, and modelling approaches/ simulations. The data collection techniques are defined, including a summary of the studies utilising this method to study pedestrian behaviour. Opportunities and drawbacks are presented afterward.

2.5.1 Pedestrian movement data collection utilising traditional approaches

Traditional approaches have relied on data from conventional sources such as census or field studies. These techniques are still relevant, and researchers mainly use them when addressing low-volume datasets. In the 32 studies retrieved, the traditional techniques introduced pedestrian behaviour issues by utilising observational data (Choi, 2014; Özbil, et al., 2015), interviews (Benachio, et al., 2018; Filingeri, et al., 2017), and survey information (Kürkçüoğlu & Akin, 2013; Teixeira, 2021).

The literature reviewed illustrates that studies adopting traditional techniques predominantly centre around the themes of emotion and motivation. These studies have explored a variety of parameters, such as preferences and emotional responses towards green spaces (Botes & Zanni, 2020; Qian, et al., 2018) and stationary and sustained activities in pedestrian streets (Ghahramanpouri, et al., 2012). The majority of these studies have used the collected field observations to calibrate simulation models (Capitanio, 2019; Mansouri & Ujang, 2017). Other studies have conducted field experiments as a way to collect on-site qualitative information from the pedestrians participating in these experiments (Nikolopoulou, et al., 2016; Zapata & Honey-Rosés, 2020; Askarizad & Safari, 2020). In both cases, the intention was that the collected data would be used to study people's walking behaviour in natural settings while they move as unobtrusively as possible.

2.5.1.1 Strengths and limitations of traditional data collection approaches to study pedestrian behaviour

The adoption of traditional data collection approaches presents several opportunities concerning data richness and validity. Field observations and surveys can be conducted over long periods of time, and the observer can collect specific characteristics of pedestrians (Feng, et al., 2021). These may include gender, direction, personal items, clothing information, or psychological insights (e.g., preferences and motivations), resulting in rich and detailed information considering the fundamental concepts of influence of human behaviour. In addition, pedestrians do not have the knowledge of being observed. Hence, their response to urban settings is in a more natural fashion. Analysis techniques of traditional data sources rely mainly on statistical models, wherein designers and urban planners have been traditionally trained to undertake such tasks as statistical analysis, survey research, and estimation (French, et al., 2015). Therefore, such approaches can ensure their practical application in research and industry without minimising their research potential.

Drawbacks exist around controllability, data quality, representativeness, and associated costs of the experiments (Hoogendoorn, 2004; Vanumu, et al., 2017). Such approaches require researchers to manually record walking patterns, while in many cases, the techniques used lack the ability to handle dynamic attributes such as capturing preferences or temporal information. Furthermore, such tasks are time-consuming and costly, as they require expensive resources, such as human resources or equipment installation, such as cameras and sensors (Özbil, et al., 2015). In addition, the accuracy of behavioural data relies on several other parameters. These include the setup and the techniques used, such as the granularity of the data, the mechanisms of recording the data and the distribution of their densities, the respondents' internal characteristics, such as past experiences or personal views (Zapata & Honey-Rosés, 2020). Therefore, this often results in unreliable datasets, not suitable for detailed analyses. Nevertheless, such limitations do not apply in the collection of datasets utilising surveys, as researchers have a high degree of experimental control in the techniques used. For example, interview questions are entirely drafted and pre-determined by the researcher. Nevertheless, such approaches may gather insights into behaviours that rarely occur or where an opportunity has not yet been presented (preferences). However, the number of persons observed in such settings is limited. Therefore, the recorded sample is not considered representative of the whole population, and conclusions cannot be drawn as characteristics of the individual behaviours (Millonig, et al., 2009).

2.5.2 Controlled experiments via the use of VR

Conducting controlled experiments by collecting observational data via field studies or sensors' installation can be proven costly (Özbil, et al., 2015; Feng, et al., 2021). Novel technologies offer opportunities for conducting experimental analysis to investigate large complex and realistic scenarios, such as experiments via the use of Virtual Reality (VR). Most such studies focus on studying behaviours in a variety of dangerous situations, e.g., earthquakes, crossing behaviours, or crowd management (Fiset, et al., 2020; Reffat, 2012). Natapov and Fisher-Gewirtzman (2016) report on experiments in VR, focusing on capturing walking patterns to investigate the impact of attractors in city street networks and their influence on pedestrian route choices during explorative walking activities. Such findings can potentially serve as data inputs to calibrate tools and inform design. However, comparative studies are required to validate such results.

2.5.2.1 Strengths and limitations of VR in controlled experiments to study pedestrian behaviour

Controlled experiments in VR present several opportunities, with their main advantage being the controllability of the study. VR experiments have high experimental control. Their environments are built and modified as the researcher dictates, allowing for a quick testing of the effect of specific factors on pedestrian movement (Feng, et al., 2021). Such studies can expose the participants to diverse environments while tracking behaviours that are not possible in real life. Researchers can collect information utilising VR setups that are not easy to observe in natural settings, such as reactions in isolated characteristics found in urban environments (Bhagavathula, et al., 2018). Datasets from such studies are more likely to result in greater accuracies, while their collection is automatic, reducing manual editing and time-consuming tasks. Therefore, such studies are cost-effective, as operational, and logistical costs are significantly reduced (Haghani, et al., 2016). Additionally, representation technologies in practice are widely used throughout the design process, aiming to bring concepts into reality or as a communication language amongst stakeholders (Horne & Thompson, 2008). Therefore, design students develop skills in such technologies from the early days and explore the three-dimensional space, which until recently was satisfied via the use of physical models. This results in an increased number of practitioners and researchers able to utilise VR setups to experiment and collect data against specific research questions, supporting their practical application into design workflows.

There are several drawbacks to VR experiments. Although such technologies are not restricted by physical boundaries (e.g., location and time), they require pre-selection of participants and effects, which might influence the representativeness of the study (Feng, et al., 2021). For example, older people who are not familiar with emergent technologies (Bode & Codling, 2013). However, the most significant disadvantage of such approaches is that virtual environments only imitate real-life situations. Therefore, participants may act differently, and their behaviour can be heavily affected. Walking through an environment is a physical experience, and VR experiments can provide only a limited sense of spatial experience. Therefore, conclusions can be drawn to some extent, but they require cross-validation techniques, utilising alternative sources of information, such as observational data collection, to serve as ground truth datasets.

2.5.3 Pedestrian movement data collection utilising wearable technologies

Due to the technological achievements of the Information and Communications Technology (ICT), creation of new services and interconnected devices has been observed, empowered by Internet of Things (IoT) (Ometov, et al., 2021). During the past decade, the emergence of psychophysiological measurements to inform the broader environmental psychology, cognitive neuroscience, and urban studies has led to low-cost wearable equipment, such as skin conductance and heart-rate monitoring devices (Hammock, et al., 2013; Park, et al., 2015). A few researchers have used wearable devices to identify emotions concurrent with a background of environmental information, capturing physiological signals to evaluate built environments, such as eye movement, blood volume pulse, electrocardiogram, or skin conductance level (Resch, et al., 2020; Kim, et al., 2020; Zhao, et al., 2019). Their findings highlight the opportunities arising from crowdsourced physiological data to gain insights into the way people interact with their environments (Engelniederhammer, et al., 2019). Resch et al. (2020) used a combination of novel data from sensors and traditional datasets to understand pedestrians' feelings and expectations of urban space. Kim et al. (2020) investigated specific behaviours (physical disorders) found in a city to identify their influencing parameters and then evaluate urban settings, promoting neighbourhood walkability and feelings of safety.

Due to the improvement of similar technologies and the addition of precise tracking functions, such as full-body tracking and eye-tracking (Liao, et al., 2016), researchers are now able to collect and analyse various aspects of pedestrian behaviour in great detail (Croft & Panchuk, 2018; Jiang, et al., 2018). These include capturing small actions, such as glances or brief steps, sensory physiological functions, or wide-ranging body movements that are difficult to collect in natural settings (Khan, et al., 2020). Such capabilities can be used as ways to monitor health, decision-making, physical activity, user experience, crowd-sensing, and many others (McCallum, et al., 2018; Rashid & Wang, 2020; Soh, et al., 2015).

2.5.3.1 Strengths and limitations of wearable technologies to study pedestrian behaviour

The most significant advantage of wearable technologies is that researchers can collect detailed individual information, such as heart rate or eye movements, which can be collected in increased granularity and a high level of accuracy. Such approaches present the opportunity of collecting information in participants' natural settings, which is not achievable via laboratory or controlled experiments (Resch, et al., 2020). Therefore, researchers can collect information on participants' experiences when decisions occur, providing them with insights as to why specific actions took place.

Although they have an increased level of controllability, such collection techniques are restricted by ethical considerations (Xue, 2019). Therefore, to be applicable, they require a balance amongst realism, scientific curiosity, and level of invasiveness throughout the experiment (Engelniederhammer, et al., 2019). In addition, wearable devices share the same issues with traditional and controlled experiments concerning representativeness, as wearable devices can result in increased costs. Data resolution and accuracy is strongly correlated to the accuracy of the measurement sensors attached to the wearable device (Pal, et al., 2019), and researchers need to balance the strengths and limitations of the chosen technologies between low-cost and high-cost approaches (Klus, et al., 2019). Therefore, they can be used to study a small sample of the overall population. Finally, intensive pre-processing of wearable tracking data limits their applicability in practice and makes it difficult to achieve a thorough usability evaluation (Liao, et al., 2016).

2.5.4 Pedestrian movement data collection utilising large-scale monitoring

Significant technological improvements have allowed the cities' infrastructure and pedestrian monitoring via smartphones and sensor networks (Wirz, et al., 2013), with early versions of such systems emerging in the 2000s (e.g., (Yang, et al., 2003)). Large-scale monitoring systems to study pedestrian movement include camera-based monitoring, Bluetooth and Wi-Fi sensors, global positioning system (GPS) trackers, and mobile phone data. A wide variety of new approaches has become apparent due to these technological improvements and the collection of high-volume datasets, often referred to as "Big data" (Shi & Abdel-Aty, 2015; Reddy, et al., 2020). Literature analysis revealed that large-scale monitoring is not widely used in scientific areas, such as the influence of the character of a place on pedestrians' motivations, emotions, and previous experiences. Previous studies focus on resolving issues regarding their validity due to the lack of information on these new systems, e.g., Duives et al. (2019). Sophisticated digital sensor systems can actively capture pedestrian movement in larger spatial and temporal scales (Feng, et al., 2021). Although this creates various opportunities to study pedestrian behaviour, most of these studies have turned their focus on the performance of the tracking techniques. However, their findings are limited to the counting of individuals in pedestrian flows (Zaki & Sayed, 2018).

Various studies, from indoor environments to transportation hubs and mass events (Duives, et al., 2020), have applied Wi-Fi tracking techniques, video footage, or traffic cameras to collect large-scale movements or achieve real-time crowd monitoring. Duives et al. (2020) combined video systems and computer vision algorithms to study pedestrian movement in mass events, while Li et al. (2021) used process imaging techniques to analyse pedestrian behaviour in zigzag corridors in the context of safety. However, there is limited research focusing on the underlying characteristics of pedestrian behaviour. One such study used GPS in the context of pedestrian motivation (Orellana & Wachowicz, 2011), enabling the study of movement suspension patterns. Hence, new techniques were used to identify places that attract or restrict human movement, allowing the understanding of pedestrian movement from a behavioural perspective. Several other data collection techniques exist, such as mobile apps like Strava or eco-counters (Inc., 2022; Eco-Counter, 2019). These are tracking applications based on GPS tracking systems and real-time sensors, including heart rate, speed, elevation, and others. They also incorporate social media features such as recording activities, pictures, and text that can enable an in-depth study of pedestrian movement. However, these systems have not been utilised in the identified studies.

2.5.4.1 Strengths and limitations of large-scale monitoring to study pedestrian behaviour

One of the significant advantages of large-scale monitoring is that it enables researchers to study wider crowds in urban space. Such techniques allow researchers to collect comprehensive data samples that represent the pedestrian population well. Compared to other data collection techniques, such data are recorded automatically, reducing manual interaction with the systems. Therefore, researchers can review information in real-life settings as pedestrians have limited knowledge of being tracked, resulting in a high degree of validity (Liang, et al., 2020). Temporal data are traditionally analysed in one-year scales, and now leveraging such techniques can allow for extended periods to be studied or even studied in real-time. Consequently, pedestrian data collected contains rich information, considering the fundamental qualities of pedestrian behaviours (Vanumu, et al., 2017).

However, data collection from large-scale monitoring often lacks controllability and data quality (Feng, et al., 2021). Several exogenous parameters can influence the data collection process, such as sensor setup, camera angle, signal strength, distribution of the nodes, and granularity of the data, significantly decreasing the accuracy and reliability of the collected dataset (Orellana & Wachowicz, 2011). In addition, although such techniques offer solutions to the manual collection of information, there are many restrictions related to installation permissions in the public domain and ethical considerations. Such techniques offer opportunities for studying population patterns in urban spaces. However, the factors influencing pedestrian behaviour cannot be controlled, and the conditions under which the data were recorded are not controllable by the researcher (Liang, et al., 2020). Finally, operational costs and large dataset analysis involve great human resource investments, limiting their practical applications (Feng, et al., 2021). Intense data processing and analysis are involved in such techniques, which require specific skill sets in data mining, machine learning, network analysis, pattern recognition, and visualisation techniques, limiting the number of people who can participate in their implementation (Liang, et al., 2020).

2.5.5 Modelling approaches/ simulations

A great variety of research approaches has been proposed for modelling pedestrian behaviours. The most prominent approaches can be classified into macroscopic, focusing on the temporal evolution of the crowd densities, and microscopic models, investigating behaviour and decision-making of individuals (Millonig, et al., 2009; Wang, et al., 2014). The macroscopic models ignore the individual characteristics of the agents and treat pedestrians as a crowd. Therefore, these models lack on illustrating the effect of environmental change on pedestrian flows and thus do not represent real-life scenarios well (Johansson, et al., 2012). However, microscopic models can better represent pedestrian behaviour in various situations. Such pedestrian models can be categorised even further based on the employed methods, such as discrete or continuous, rule-based, or force-based. The type of the microscopic models used for analysis is closely related to the research question due to their capacity to analyse the dependent (individuals) or independent (space) variables concerning pedestrian behaviour (Millonig, et al., 2009). For example, social force models can demonstrate how others influence individual behaviours and the surrounding environments (Kolivand, et al., 2020), while cellular automata models produce results when describing behavioural rules and spatial relationships (Zhang, et al., 2020). However, such modelling approaches are dependent on the modeller's assumptions and intuition, which may not reflect realistic pedestrian results. Furthermore, these models either focus only on aspects of safety and mobility, or they require rich data sets to cut down the computational power and validate results. Therefore, they heavily rely on experimental studies or monitoring techniques to collect such data.

Finally, most studies researching urban design and walkability rely on the assumption that street configuration is the most important influencing factor of pedestrian movement (Wang & Huang, 2019; Hillier & Hanson, 1984). Many researchers utilise configuration analysis to better estimate pedestrian flow volumes and route choices (Capitanio, 2019; Mansouri & Ujang, 2017; Özer & Kubat, 2015). This approach translates complex street networks into behavioural principles of the individuals' preference for high street network legibility (Boumezoued, et al., 2020). Nevertheless, this lacks qualitative information, such as aesthetics of chosen routes, safety feelings, light conditions, and others, and impacts due to increased time spent in the area or route direction changes.

2.5.5.1 Strengths and limitations of modelling approaches & simulations to study pedestrian behaviour

One of the most significant advantages of modelling approaches and simulations is related to the high controllability of the study. For example, the potential to simulate individual differences, psychological aspects, and social interactions utilising microscopic approaches (Seitz, et al., 2017). The researcher can alter the modelling effects and test diverse scenarios to understand human behaviour better. Therefore, such approaches can be proven cost-effective and directly applied in decision-making processes to inform urban design optioneering.

However, simulation approaches fall short of a comprehensive theory of environmental psychology. Therefore, additional sources of information are an inevitable requirement of successful modelling in the study of pedestrian walking behaviour. Empirical data and observations, usually obtained through traditional collection methods, are utilised to gather temporal and individual attributive data, resulting in increased costs of such studies. Like VR techniques, simulations imitate real-life scenarios. Knowledge-driven behaviour modelling uses multiple sources of behavioural data as input and offers opportunities to the modeler to create meaningful simulation models representing existent scenarios. In many cases, researchers tend to incorporate factors or behaviours that do not respond to reality (Song, et al., 2021). Other challenges and drawbacks of these approaches are related to the decisions of the modelers prior to simulations. Human behaviour is of increased complexity, and analysts predict and replicate such behaviours within their models, often not representative of the general population. In addition, the movement of individuals cannot be considered continuous over space, as they are entitled to the freedom of revisiting places or changing their movement decisions continuously in time, adding to the modelling exercise complexity. Therefore, the outcomes of these approaches are hypothetical and do not always represent reality.

2.6 Identifying literature research gaps

In this chapter four research gaps have been identified through the SLR process: Gap 1: Limited research of pedestrian behaviour in new types of complex set-ups; Gap 2: Collecting and analysing concise high-volume individual objective data sets; Gap 3: Collecting and analysing concise high-volume individual behavioural data sets; and Gap 4: Representativeness limitations in pedestrian behaviour and urban space captured datasets.

2.6.1 Gap 1: Limited research of pedestrian behaviour in new types of complex set-ups

This study illustrates that although pedestrian behaviour has been extensively researched, themes such as evacuation behaviours, pedestrian walking dynamics, and group behaviours have been investigated in aspects of safety and mobility. Other perspectives, such as emotional responses, individuals' motivations, and previous experiences, have received far less attention. Examples of these new perspectives are recent studies around streetscape features relating to comfort and pleasurability (Capitanio, 2019), emotions against the background of environmental information (Resch, et al., 2020), or familiar and unfamiliar spaces (Phillips, et al., 2013). These studies highlight that pedestrian behaviour changes differ based on diverse parameters relating to the physiological characteristics of the individuals. Emotional responses are part of the individual or collective subjective experiences and constitute a motivational factor for behaviour and choice.

There is a lack of research approaches that explore the relationship between pedestrian movement, emotions, and spatial cognition. For example, Capitanio (2019) concluded that streetscape features are why residents prefer enjoyable routes over shorter ones. However, these findings depend on simulation results and observational data collected on-site. As most of these studies have been conducted utilising field observations or controlled experiments, they often suffer from the lack of controllability, accuracy, and reality (Feng, et al., 2021). In addition, limited behavioural responses participants of experimental setups follow the given instructions. Therefore, such approaches limit the potential of the collected data, requiring new data collection methods.

2.6.2 Gap 2: Collecting and analysing concise high-volume individual objective data sets

Field observations offer opportunities to capture data that will not be influenced by the researchers, hence more likely to result in unbiased behavioural data, which can ensure a relatively high degree of validity. However, due to the way these data are collected in the context of the studies discussed, for example via manual recording, this results in a relatively small sample, which cannot be representative of the population. New data sources can provide more detailed and real-time data at various scales. An example is GPS-tracking devices, which enable the collection of long-term geo-referenced trajectory data of individuals. In addition, new types of digital sensors enable crowd monitoring, such as Wi-Fi sensors, mobile phone data, and camera-based monitoring systems. However, most of the studies employing these techniques are limited to the study of emergency behaviours and movement dynamics at various spaces, with little attention given to the internal relationships between different choice dimensions or other underlying characteristics such as emotional responses.

Despite limitations concerning the accuracy of such recorded data and related ethical considerations, these approaches present numerous opportunities to researchers and promise to reveal insights on the way people interact with the built environments. Thus, there is a need to overcome the key challenges described. New methods are required to allow researchers to capture and study high-volume objective data sets from a behavioural perspective.

2.6.3 Gap 3: Collecting and analysing concise high-volume individual behavioural data sets

Traditional data collection methods, such as surveys and field observations, do not lend themselves well to capture sufficient behavioural data to improve our understanding of pedestrian choices. Improvements of sensor systems embedded in everyday life, such as mobile health and assistive technologies, present opportunities to collect personalised data, contributing towards the research of quality of life and well-being. In addition, new models of citizen participation in problem-solving emerge, along with new types of data, such as those captured by social media. However, these new models and related methods are still in their infancy. They are limited to the researchers' capabilities to apply novel processing techniques to convert qualitative data to quantitative information of high-volume data sets. Hence, there is a need for novel methods to allow researchers to capture and analyse high-volume behavioural data sets that include individual attributive information, such as psychological data, experiences, and personal characteristics.

2.6.4 Gap 4: Representativeness limitations in pedestrian behaviour and urban space captured datasets

Controlled experiments and survey methods mainly applied in the context of understanding behavioural responses and walking patterns are usually collected in specific contexts with a focused type of participants (i.e., elderly population, students, etc.). Therefore, it can be argued that these methods cannot represent real-life scenarios or can be generalised to other contexts (Feng, et al., 2021). In addition, pedestrians are highly dependent on the external environment. However, information capturing fine qualities found in urban spaces is still lacking from available data sets. An example of these types of data sets trying to capture urban space qualities in a more detailed level is the Points of interest (POIs), where studies have defined them as points visited frequently or where the commuter stops (Millonig, et al., 2009).

Nevertheless, several other aesthetic qualities are found in the physical space, such as buildings' materiality and form, vegetation, and other visual information. Such attributes are not effectively captured in available data sets to enable researchers to link them with the internal appreciation systems of humans. For example, orientation describes the ability to not getting lost within an urban space, while mystery refers to not having a clear or defined path. In design, this can be achieved by balancing small streets and big ones. Therefore, additional studies are required with various heterogeneity of participants while implementing spatial information with qualitative attributes is needed to unlock the potential of research from a behavioural perspective.

2.7 Synthesis of the findings

The process of identifying strengths and limitations of the data collection techniques, the literature research gaps and proposed opportunities arising through this analysis are presented in Figure 2-4, revealing the inner connections and highlighting the findings of the study.

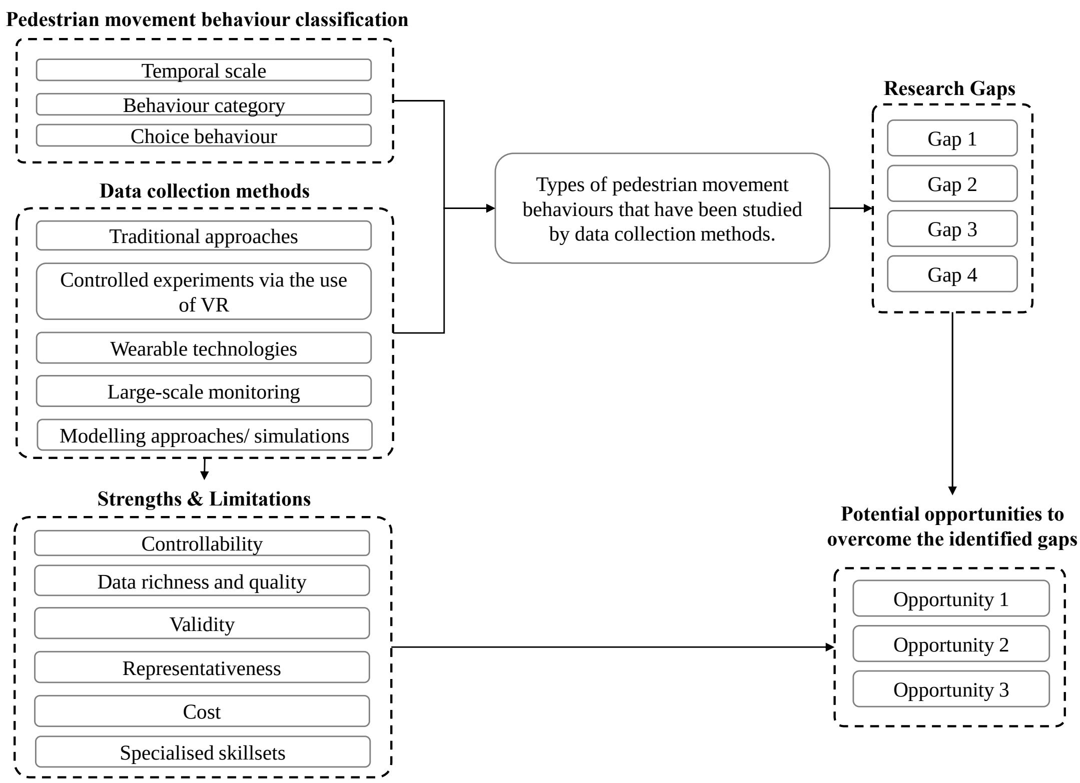


Figure 2-4: Synthesis of the research findings

Table 2-1 provides an overview of the strengths and limitations of the different data collection methods identified for collecting pedestrian movement behaviour data. The content of the studies reviewed is used to discuss the strengths and limitations of the data collection methods, which were discussed from perspectives of controllability, data richness and quality, validity, representativeness, cost, and specialised skillset.

Table 2-1: A comparison of the strengths and limitations of different data collection methods.

Data collection methods	Controllability	Data richness and quality	Validity	Representativeness	Cost	Specialised skillset
Traditional approaches (field observations, surveys, interviews, questionnaires)	• External factors cannot be controlled.	• Can be conducted over long periods of time. • The observer could collect specific characteristics of pedestrians. • Experimental setup influences the accuracy of behavioural data.	• Pedestrians do not have the knowledge of being tracked, hence, their response to urban settings is in a more natural fashion. A relatively high degree of validity.	• Number of people to be tracked in such settings is limited. The collected information may not be representative of the population.	• Expensive resources: manpower, or installation equipment. • Time-consuming and challenging to obtain ethical approval.	• Designers and urban planners have been traditionally trained to undertake such tasks.
Controlled experiments via the use of VR	• High experimental control.	• Increased accuracy and automatically collect data.	• Questioned due to knowledge of the equipment. • Limited sense of spatial experience. • Require cross-validation techniques.	• Can be conducted at a variety of locations and times, increasing the heterogeneity of sampling. • Highly influenced by the selection of participants.	• Manpower resource has lower costs. • Potentially expensive equipment resources. • Can be used repeatedly.	• Specialised skills but part of the designers and urban planners' educational curriculum.
Wearable technologies	• Increased level of controllability.	• Detailed individual information. • Increased granularity and high level of accuracy. • Collecting information in participants natural settings, which is not achievable via laboratory or controlled experiments.	• Knowledge of being tracked, hence, their response to urban settings can be influenced. A relatively good degree of validity.	• Small sample of the overall population.	• Restricted by ethical considerations. • Expensive resources: installation equipment.	• Specialised skills: intensive pre-processing.

Large-scale monitoring	• External factors cannot be controlled.	• Data are recorded automatically. • Longer periods to be studied or even studied in real-time. • Experimental setup influences the accuracy of behavioural data.	• Pedestrians do not have the knowledge of being tracked, hence, their response to urban settings is in a more natural fashion. A high degree of validity.	• Study wider crowds in larger contexts.	• Operational costs and large dataset analysis involve great investments in manpower.	• Specialised skills: intensive processing.
Modelling approaches/ simulations	• High level of controllability.	• Low accuracy and biased	• Simulation approaches fall short as a comprehensive theory of environmental psychology. • Require cross-validation techniques. • Limited sense of spatial experience. • Biased modelling inputs from modellers.	• Often not representative of the general population.	• Cost-effective, but when coupled with traditional collection methods, increase the costs. • Directly applied in decision-making processes.	• Specialised skills: software knowledge and post-processing.

2.8 Exploring the opportunities to overcome the identified gaps

In this section, three opportunities concerning the application of new types of analysis and data collection techniques are proposed by the authors, to bridge the research gaps identified, as these have emerged from the identified literature reviewed. The opportunities are: (i) Employing large-scale pedestrian movement monitoring, (ii) Internet of Things (IoT) systems employment and (iii) Leveraging cross-discipline incorporation.

Opportunity i: Employing large-scale pedestrian movement monitoring

Combining theory-driven models and data-driven approaches is necessary for understanding human and natural processes in cities. In the past years, improvements in novel technologies have created new data-rich environments, encouraging researchers to replace their traditional approaches. New approaches allow the study of pedestrian movements in large complex environments in more detail and, more specifically, unravel insights in situations where research is limited. Furthermore, they provide an extensive understanding of the decision-making processes of pedestrian behaviour in urban settings.

However, there are some challenges concerning large-scale monitoring systems. The first lies in the potential infringement of fundamental rights, the right of privacy. Therefore, employing technical equipment that protects sensitive data “by design” is necessary. The second challenge lies in the analytical capabilities of the researchers, as these new systems generate information at an unparalleled pace. Finally, new approaches may be limited due to high installation and maintenance costs of digital sensor systems, such as Wi-Fi tracking devices or video-based technologies, introducing challenges in supplementing the urban data infrastructure.

Opportunity ii: Internet of Things (IoT) systems employment

New streams of data alter the way cities are defined and operate, changing the focus for policymakers, citizens, and private stakeholders, from the long term to the short term. New models of citizen participation with problem-solving have stimulated research into a range of social issues, utilising information from data available on the internet. Such approaches are also evolving, linking these digital interfaces with infrastructure assets, and providing real-time monitoring and prediction opportunities. These types of models present an opportunity for the IoT data, such as information deriving from mobile phones (e.g., running applications), wearable devices (e.g., sports watches), and other opportunities linked to IoT, such as social media (e.g., Twitter). These types of data can be studied in pedestrian movement, emotional responses to the built environments, and choice behaviour. Consequently, IoT can reveal new insights in movement behaviours regarding types of movement and locations where accessibility is limited.

Similar to opportunity (i), privacy issues present as one of the key limitations regarding IoT. Therefore, the development of standards featuring protection protocols for the use of such data is necessary, while the active participation of the researchers is encouraged. In addition, other key limitations include the application of partially or fully automated techniques, such as using “machine-learning” algorithms. Their effective implementation lies with the researchers; employment of such approaches presents additional challenges, which are further discussed below.

Opportunity iii: Leveraging cross-discipline incorporation & up-skilling

Big Data implementation in urban research and practical applications should not be considered a distinct technology phase. New types of information and models are becoming more and more complex and involve even further the intuition and assumptions of the researcher. Therefore, to apply these novel techniques in such context, a better understanding of the decision-making processes of individual pedestrians is required and analytical capabilities to develop the models that can assess such types of information. Nevertheless, this may involve knowledge from the social and psychological fields. Thus, cross-disciplinary collaboration is necessary to further develop similar research approaches in the future and alter traditional education programmes to match new trends and to up-skill researchers to undertake such analysis.

2.9 Conclusion

This chapter aims to explore data collection techniques to investigate pedestrian movement behaviour in the study of urban space recognition and the role of sensorial experience on movement patterns, and to determine their strengths and limitations in this context. This chapter examines the contexts where these techniques have already been applied via an SLR.

In-depth analysis of the capabilities of data collection techniques reveals a series of gaps that provide valuable insights to urban planners, designers, and academics for improving urban design processes. Opportunities to fill the identified gaps are thoroughly discussed. This study revealed that novel technologies could address the research gaps identified in three ways. The first opportunity applies to the study of pedestrian behaviour in urban space by large-scale pedestrian movement monitoring and gaining an extensive understanding of the decision-making processes of pedestrian behaviour in urban settings. The second opportunity is employing the IoT systems to track pedestrian dynamics and collect behavioural data to inform objective movement patterns, unravelling new insights concerning pedestrian choice and emotional responses. The third opportunity is leveraging cross-discipline incorporation and up-skilling to create individuals equipped with the necessary skillsets to address future challenges in the new data-rich paradigm.

Implications of this study include the fact that the journal articles studied were restricted to available in online academic search databases, with full access, in English language. Furthermore, the fast-paced technological changes have altered the way data are being collected in the last decade. Nevertheless, these advancements are still in their infancy.

Future research can identify more opportunities regarding new types of analysis and data collection techniques. Researchers could then map the methods contributing to pedestrian movement analysis to create the "*researchers' toolbox*". Finally, future studies should focus on gathering and analysing more wide-ranging data concerning urban space recognition and the role of sensorial experience on movement patterns.

REFERENCES

Özbil, A., Yesiltepe, D. & Argin, G., 2015. Modeling Walkability: the effects of street design, street-network configuration and land-use on pedestrian movement. *A|Z ITU Journal of Faculty of Architecture,* 12(3), pp. 189-207.

Özer, Ö. & Kubat, A., 2015. Measuring walkability in Istanbul Galata Region. *ITU A|Z,* 12(1), pp. 15-29.

Askarizad, R. & Safari, H., 2020. The influence of social interactions on the behavioral patterns of the people in urban spaces (case study: The pedestrian zone of Rasht Municipality Square, Iran). *Cities,* Volume 101, Article 102687.

Batty, M., 1997. Predicting where we walk. *Nature,* 388(19 - 20).

Batty, M., 2001. Agent-based pedestrian modelling,. *Environment and Planning B: Planning and Design,* Volume 28, p. 321 – 326.

Batty, M., 2013. *The new science of cities..* Cambridge, MA: The MIT Press.

Benachio, J., Haveriku, X., Zaluski, P.D., Chen, H.-H. & Dietrich, U., 2018. "Slow your motions" interventions in urban spaces towards a livable neighborhood: Case study of Hamm-Nord, Germany. *WIT Transactions on Ecology and the Environment,* Volume 217, pp. 843-854.

Bhagavathula, R., Williams, B., Owens, J. & Gibbons, R., 2018. The Reality of Virtual Reality: A Comparison of Pedestrian Behavior in Real and Virtual Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 62(1), pp. 2056-2060.

Bode, N. & Codling, E., 2013. Human exit route choice in virtual crowd evacuations. *Animal Behaviour,* 86(2), p. 347–358.

Botes, C. & Zanni, A., 2020. Trees, ground vegetation, sidewalks, cycleways: users' preferences and economic values for different elements of an urban street—a case study in Taipei. *Environmental Economics and Policy Studies,* Volume 23, p. 145–171.

Boumezoued, S., Bada, Y. & Bougdah, H., 2020. Pedestrian itinerary choice: between multi-sensory, affective and syntactic aspects of the street pattern in the historic quarter of Bejaia, Algeria. *International Review for Spatial Planning and Sustainable Development,* 8(4), pp. 91-108.

Brattico, E., Bogert, B. & Jacobsen, T., 2013. Toward a Neural Chronometry for the Aesthetic Experience of Music. *Frontiers in Psychology ,* Volume 4, p. 206.

Brattico, E. & Pearce, M., 2013. The neuroaesthetics of music. *Psychology of Aesthetics, Creativity, and the Arts,* 7(1), p. 48–61.

Capitanio, M., 2019. Attractive streetscape making pedestrians walk longer routes: The case of Kunitachi in Tokyo. *Journal of Architecture and Urbanism,* 43(2), pp. 131-137.

Chatterjee, A., 2004. Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of Psychology and the Arts,* 4(2), pp. 56-60.

Choi, E., 2014. Walkability and the complexity of walking behavior. *A/Z ITU Journal of the Faculty of Architecture,* 11(2), pp. 87-99.

Craig, C., Brownson, R., Cragg, S. & Dunn, A., 2002. Exploring the effect of the environment on physical activity: A study examining walking to work.. *American Journal of Preventive Medicine,* Volume 23, p. 36–43.

Croft, J. & Panchuk, D., 2018. Watch Where You're Going? Interferer Velocity and Visual Behavior Predicts Avoidance Strategy During Pedestrian Encounters. *Journal of Motor Behavior,* 50(4), pp. 353-363.

Dridi, M., 2015. Simulation of High-Density Pedestrian Flow: A Microscopic Model. *Open Journal of Modelling and Simulation,* Volume 3, pp. 81-95.

Duives, D., van Oijen, T. & Hoogendoorn, S., 2020. Enhancing crowd monitoring system functionality through data fusion: Estimating flow rate from wi-fi traces and automated counting system data. *Sensors (Switzerland),* 20(21), pp. 1-25.

Duives, D., Wang, G. & Kim, J., 2019. Forecasting pedestrian movements using recurrent neural networks: An application of crowd monitoring data. *Sensors,* 19(2), p. 382.

Eco-Counter, 2019. *Eco-Counter.* [Online]
Available at: https://www.eco-counter.com
[Accessed 25 02 2022].

Engelniederhammer, A., Papastefanou, G. & Xiang, L., 2019. Crowding density in urban environment and its effects on emotional responding of pedestrians: Using wearable device technology with sensors capturing proximity and psychophysiological emotion responses while walking in the street. *Journal of Human Behavior in the Social Environment,* 29(5), pp. 630-646.

Feng, Y., Duives, D., Daamen, W. & Hoogendoorn, S., 2021. Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment,* Volume 187, Article 107329.

Fernandez-Ares, A. et al., 2020. Detection and Analysis of Anomalies in People Density and Mobility through Wireless Smartphone Tracking. *IEEE Access,* Volume 8, pp. 54237-54253.

Filingeri, V., Eason, K., Waterson, P. & Haslam, R., 2017. Factors influencing experience in crowds – The participant perspective. *Applied Ergonomics,* Volume 59, pp. 431-441.

Fiset, F., Lamontage, A. & McFadyen, B., 2020. Limb movements of another pedestrian affect crossing distance but not path planning during virtual over ground circumvention. *Neuroscience Letters,* Volume 736, Article 135278.

Frank, L., 2000. Land Use and Transportation Interaction: Implications on Public Health and Quality of Life. *Journal of Planning Education and Research,* 20(1), pp. 6-22.

French, S. P., Barchers, C. & Zhang, W., 2015. *How Should Urban Planners Be Trained to Handle Big Data?.* Chicago, Springer.

Gehl, J. & Svarre, B., 2013. *How to Study Public Life.* Washington, DC: 2nd ed. Island Press.

Ghahramanpouri, A., Lamit, H. & Sedaghatnia, S., 2012. Behavioural observation of human stationary and sustained activities in pedestrian priority streets of johor bahru. *Journal of Construction in Developing Countries,* 17(2), pp. 105-116.

Haghani, M., Sarvi, M., Shahhoseini, Z. & Boltes, M., 2016. How simple hypothetical-choice experiments can be utilized to learn humans' navigational escape decisions in emergencies. *PLoS One,* Volume 11, p. 1–24.

Hammock, M., Chortos, A., Tee, B.-K. & Tok, J.-H., 2013. 25th Anniversary Article: The Evolution of Electronic Skin (E-Skin): A Brief History, Design Considerations, and Recent Progress. *Advanced Materials,* 25(42), pp. 5997-6038.

Handy, L., Boarnet, M. G., Ewing, R. & Killingsworth, R., 2002. How the built environment affects physical activity: Views from urban planning. *American Journal of Preventive Medicine,* 23(2), pp. 64-73.

Hanna, S., 2020. Random walks in urban graphs: A minimal model of movement. *Environment and Planning B: Urban Analytics and City Science,* 48(6), pp. 1697-1711.

Harris, R., 2012. *Introduction to decision making.* [Online]
Available at: https://www.virtualsalt.com/crebook5.htm

Hillier, B. & Hanson, J., 1984. *The Social Logic of Space.* Cambridge: Cambridge University Press.

Hoogendoorn, S., 2004. *Walking behavior in bottlenecks and its implications for capacity.* Washington DC, Transportation Research Board (TRB), pp. 1-13.

Hoogendoorn, S. & Bovy, P., 2004. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological,* Volume 38, pp. 169-190.

Horne, M. & Thompson, E., 2008. The role of virtual reality in built environment education. *Journal for Education in the Built Environment,* 3(1), p. 5–24.

Inc., S., 2022. *Strava.* [Online]
Available at: https://www.strava.com
[Accessed 25 02 2022].

Jiang, K., Ling, F., Feng, Z., Ma, C., Kumfer, W., Shao, C. & Wang, K., 2018. Effects of mobile phone distraction on pedestrians' crossing behavior and visual attention allocation at a signalized intersection: An outdoor experimental study. *Accident Analysis and Prevention,* Volume 115, p. 170–177.

Johansson, A., Batty, M., Hayashi, K., Al Bar, O., Marcozzi, D. & Memish, Z.A., 2012. Crowd and environmental management during mass gatherings. *The Lancet infectious diseases,* 12(2), pp. 150-156.

Kürkçüoğlu, E. & Akin, O., 2013. The effects of water elements in urban space perception: A case study in Üsküdar Municipality Square. *A/Z ITU Journal of the Faculty of Architecture,* 10(1), pp. 159-175.

Karbovskii, V., Severiukhina, O., Derevitskii, I., Voloshin, D., Presbitero, A. & Lees, M., 2019. The impact of different obstacles on crowd dynamics. *Journal of Computational Science,* Volume 36, Article 100893.

Khan, S., Parkinson, S., Grant, L., Liu, N. & Mcguire, S., 2020. Biometric Systems Utilising Health Data from Wearable Devices: Applications and Future Challenges in Computer Security. *ACM Computing Surveys,* 53(4), pp. 1-29.

Kim, J., Yadav, M., Chaspari, T. & Ahn, C., 2020. Saliency detection analysis of collective physiological responses of pedestrians to evaluate neighborhood built environments. *Advanced Engineering Informatics,* Volume 43.

Klus, L., Lohan, E., Granell, C. & Nurmi, J., 2019. *Crowdsourcing Solutions for Data Gathering from Wearables.* Finland, Tampere University.

Kolivand, H., Rahim, M.S., Sunar, M.S., Fata, A.Z.A. & Wren, C., 2020. An integration of enhanced social force and crowd control models for high-density crowd simulation. *Neural Computing and Applications,* Volume 33, p. 6095–6117.

Liang, S., Leng, H., Yuan, Q., Wang, B.W. & Yuan, C., 2020. How does weather and climate affect pedestrian walking speed during cool and cold seasons in severely cold areas?. *Building and Environment,* Volume 175, Article 106811.

Liao, H., Dong, W., Peng, C. & Liu, H., 2016. Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. *Cartography and Geographic Information Science,* 44(6), pp. 474-490.

Li, H., Thrash, T., Hölscher, C. & Schinazi, V., 2019. The effect of crowdedness on human wayfinding and locomotion in a multi- level virtual shopping mall. *Journal of Environmental Psychology,* Volume 65, Article 101320.

Liu, Z. X., Wang, X. Y., Wang, J.Q., Wang, F. & Liu, Y.Q., 2018. Pedestrian movement intention identification model in mixed pedestrian-bicycle sections based on phase-field coupling theory. *Advances in Mechanical Engineering,* 10(2).

Li, X., Ye, R., Fang, Z., Xu, Y., Cong, B. & Han, X., 2021. Uni- and bidirectional pedestrian flows through zigzag corridor in a tourism area: a field study. *Adaptive Behavior,* 29(3), pp. 1-16.

Loukaitou-Sideris, A., 2020. Special issue on walking. *Transport Reviews,* 4(2), pp. 131-134.

Mansouri, M. & Ujang, N., 2017. Space syntax analysis of tourists' movement patterns in the historical district of Kuala Lumpur, Malaysia. *Journal of Urbanism,* 10(2), pp. 163-180.

McCallum, C., Rooksby, J. & Gray, C., 2018. Evaluating the Impact of Physical Activity Apps and Wearables: Interdisciplinary Review. *JMIR Mhealth Uhealth,* 6(3:e58).

Millonig, A., Brandle, N., Ray, M., Bauer, D. & Van der Spek, S., 2009. Pedestrian Behaviour Monitoring: Methods and Experiences. In: *Ambient Intelligence and Smart Environments. Volume 3: Behaviour Monitoring and Interpretation – BMI.* Amsterdam, The Netherlands: IOS Press, p. 11.

Mills, N., De Silva, D. & Alahakoon, D., 2020. Generating Situational Awareness of Pedestrian and Vehicular Movement in Urban Areas Using IoT Data Streams. *IEEE Internet of Things Journal,* 7(5), pp. 4395- 4402.

Natapov, A. & Fisher-Gewirtzman, D., 2016. Visibility of urban activities and pedestrian routes: An experiment in a virtual environment. *Computers, Environment and Urban Systems,* Volume 58, pp. 60-70.

Nikolopoulou, M., Martin, K. & Dalton, B., 2016. Shaping pedestrian movement through playful interventions in security planning: what do field surveys suggest?. *Journal of Urban Design,* 21(1), pp. 84-104.

Omer, I., Gitelman, V., Rofè, Y., Lerman, Y., Kaplan, N. & Doveh, E., 2017. Evaluating Crash Risk in Urban Areas Based on Vehicle and Pedestrian Modeling. *Geographical Analysis,* 49(4), pp. 387-408.

Ometov, A., Shubina, V., Klus, L., Skibińska, J., Saafi, S., Pascacio, P. & Flueratoru, L., 2021. A Survey on Wearable Technology: History, State-of-the-Art and Current Challenges. *Computer Networks,* Volume 193.

Orellana, D. & Wachowicz, M., 2011. Exploring patterns of movement suspension in pedestrian mobility. *Geographical Analysis,* 43(3), pp. 241-260.

Pal, D., Tassanaviboon, A., Arpnikanondt, C. & Papasratorn, B., 2019. Quality of Experience of Smart-Wearables: From Fitness-Bands to Smartwatches. *IEEE Consumer Electronics Magazine,* 9(1), pp. 49-53.

Park, J., Fairweather, M. & Donaldson, D., 2015. Making the Case for Mobile Cognition: EEG and Sports Performance. *Neuroscience & Biobehavioral Reviews,* Volume 52, pp. 117-130.

Phillips, J., Walford, N., Hockey, A., Foreman, N. & Lewis, M., 2013. Older people and outdoor environments: Pedestrian anxieties and barriers in the use of familiar and unfamiliar spaces. *Geoforum,* Volume 47, pp. 113-124.

Qian, C., Zhu, D., Zhou, Y. & Chen, J., 2018. Measurements of pedestrian friendliness of residential area: A case study in Hexi District of Nanjing. *Sustainability (Switzerland),* 10(6), p. 1993.

Rashid, M. & Wang, D., 2020. CovidSens: A Vision on Reliable Social Sensing for COVID-19. *Artificial Intelligence Review,* Volume 54, pp. 1-25.

Reddy, G., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G. & Baker, T., 2020. Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access,* Volume 8, pp. 54776-54788.

Reffat, R., 2012. An Intelligent Computational. *Transportation Science and Technology Real-time Virtual Environment Model for Efficient Crowd Management,* 1(4), p. 365 – 378.

Resch, B., Puetz, I., Bluemke, M., Kyriakou, K. & Miksch, J., 2020. An interdisciplinary mixed-methods approach to analyzing urban spaces: The case of urban walkability and bikeability. *International Journal of Environmental Research and Public Health,* 17(19), p. 6994.

Seitz, M. et al., 2017. Parsimony versus Reductionism: How Can Crowd Psychology be Introduced into Computer Simulation?. *Review of General Psychology,* 21(1), pp. 95-102.

Shi, Q. & Abdel-Aty, M., 2015. Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies,* Volume 58, pp. 380-394.

Soh, P., Vandenbosch, G., Mercuri, M. & Schreurs, D.-P., 2015. Wearable Wireless Health Monitoring: Current Developments. *IEEE Microwave Magazine,* 16(4), pp. 55-70.

Song, X., Chen, K., Li, X., Sun, J., Hou, B., Cui, Y., Zhang, B., Xiong, G. & Wang, Z., 2021. Pedestrian Trajectory Prediction Based on Deep Convolutional LSTM Network. *IEEE Transactions on Intelligent Transportation Systems,* 22(6), pp. 3285-3302.

Stanitsas, M., Kirytopoulos, K. & Leopoulos, V., 2021. Integrating sustainability indicators into project management: The case of construction industry. *Journal of Cleaner Production,* Volume 279, Article 123774.

Storper, M. & Manville, M., 2006. Behaviour, Preferences and Cities: Urban Theory and Urban Resurgence. *Urban Studies,* 43(8), p. 1247–1274.

Teixeira, C., 2021. Green space configuration and its impact on human behavior and URBAN environments. *Urban Climate,* Volume 35, Article 100746.

Van Wee, B. & Banister, D., 2016. How to write a literature review paper?. *Transport Reviews,* 36(2), pp. 278-288.

Vanumu, L., Ramachandra Rao, K. & Tiwari, G., 2017. Fundamental diagrams of pedestrian flow characteristics: A review. *European Transport Research Review,* 9(49).

Wang, S.-M. & Huang, C.-J., 2019. Using space syntax and information visualization for spatial behavior analysis and simulation. *International Journal of Advanced Computer Science and Applications,* 10(4), pp. 510-521.

Wang, W., Lo, S., Liu, S. & Kuang, H., 2014. Microscopic modeling of pedestrian movement behavior: Interacting with visual attractors in the environment. *Transportation Research Part C: Emerging Technologies,* Volume 44, pp. 21-33.

Wirz, M. et al., 2013. Probing crowd density through smartphones in city-scale mass gatherings. *EPJ Data Science,* 2(1), pp. 1-24.

Xue, Y., 2019. A Review on Intelligent Wearables: Uses and Risks. *Hum. Behav. Emerg. Technol.,* 1(4), pp. 287-294.

Yang, H.-D., Sin, B.-K. & Lee, S.-W., 2003. Automatic Pedestrian Detection and Tracking for Real-Time Video Surveillance. In: *Kittler, J., Nixon, M.S. (Eds.), Audio- Video-Based Biometric Pers. Authentication. AVBPA 2003. Lecture Notes in Computer Science.* Berlin Heidelberg: Springer, pp. 242-250.

Zaki, M. & Sayed, T., 2018. Automated Analysis of Pedestrian Group Behavior in Urban Settings. *IEEE Transactions on Intelligent Transportation Systems,* 19(6), pp. 1880-1889.

Zapata, O. & Honey-Rosés, J., 2020. The Behavioral Response to Increased Pedestrian and Staying Activity in Public Space: A Field Experiment. *Environment and Behavior,* 54(1), pp. 36-57.

Zhang, P., Li, X.-Y., Deng, H.-Y., Lin, Z.-Y., Zhang, X.-N. & Wong, S.C., 2020. Potential field cellular automata model for overcrowded pedestrian flow. *Transportmetrica A: Transport Science,* 16(3), pp. 749-775.

Zhao, Y., Lu, T., Su, W., Wu, P., Fu, L. & Li, M., 2019. Quantitative measurement of social repulsive force in pedestrian movements based on physiological responses. *Transportation Research Part B: Methodological,* Volume 130, pp. 1-20.

3 Investigating key factors influencing decision-making in the design of buildings and places: A survey of stakeholders' perception

3.1 Abstract

Much attention is currently being paid to academic and practitioner literature to understand decision-making in building and urban design processes. However, deeper analysis concerning how decisions are best made or how these should be evaluated and optimised is needed. The use of technology in design increases human-to-machine interactions, altering existing decision-making processes. Understanding how novel technologies affect decisions motivates the development of the process, tools, and metrics. This chapter aims to investigate, quantify, and rank the relative importance of decision-making factors contributing to building and urban space design. A survey was conducted to gain insight into stakeholder perceptions concerning the factors influencing decision-making processes in the design of buildings and places. The research utilised Exploratory Factor Analysis, Average Relative Importance Index, and Spearman's Rank Correlation Coefficient to identify the key factors and their relative importance influencing decision-making processes in design. Ten distinct factors were identified in total, of which four were ranked as highly important for all stakeholder types, namely: *Potential for Dynamic Operation, Recency of tools, Thoroughness and Control*. This study provides a new means to evaluate the performance of decision-making processes when these are undertaken by developing and applying a quantitative data-driven, evidence-based methodological framework. The recipients of the findings will be the urban planners, designers, and academics interested in improving existing approaches in design and final decision outcomes utilising novel technologies.

3.2 Introduction

More than half of the world's population currently live in urban areas (United Nations, 2019). Cities have a multifaceted role in societies, and their design is essential in enabling vibrant and sustainable environments for their inhabitants (Kuddus, et al., 2020). Our world generates data at an unprecedented pace, deriving from diverse sources and devices, from anywhere and at any time (Yi, et al., 2014). These may include socio-economic, spatial, environmental, and several other types of urban data.

The overwhelming growth and continuous production of unstructured data significantly affect how people understand and communicate new knowledge from these data. Interpretation of urban data and related knowledge extraction is a key challenge for decision-makers in designing buildings and places (Gandomi & Haider, 2015; Simonet, et al., 2015). As more urban data become available, opportunities for evidence-based approaches and frameworks to guide decision-making emerge. The increased amount and availability of generated information, often referred to as Big Data (BD), and the improvement of new technologies and analytical techniques help businesses develop strategies. The use of data and analytics to drive the improvement of existing processes is referred to as Data-Driven Innovation (DDI). Such strategies have a critical role when developing business opportunities (Zillner, 2021; OECD, 2015). Thus, organisations gain insights and improve their decision-making processes (DMPs) in various sectors, such as healthcare, energy, infrastructure, and construction (Yi, et al., 2014; Castelli, et al., 2020). For example, analysing data from transportation and mobility through tracking technologies denote moving entities' spatial positions, supporting DMPs of traffic managers and urban designers (Andrienko & Andrienko, 2008).

Harris (2012, pp. 1, para 2) provides two definitions for decision-making:

"*Decision-making is the study of identifying any alternatives based on the values and preferences of the decision-maker*"; and

"*Decision-making is the process of sufficiently reducing uncertainty and doubt about alternatives to allow a reasonable choice to be made among them*".

Decision-making associated with urban design processes is challenging due to data processing and the identification of appropriate metrics. Although the experience of a city is a subjective notion, the characteristics shaping its quality are mainly objective, such as the physical aspects of a city, air pollution levels, transportation and mobility, green spaces, and others (Ferreira, et al., 2015). However, urban designers and architects still rely on their expertise and precedent information when making design decisions. The complexity of the decision-making process introduces differing levels of expertise and worldviews, which can be inherently subjective. All these challenges combined make decision-making difficult.

This study aims to investigate, quantify, and rank the relative importance of the decision-making factors contributing to the design of the building and urban projects. The key influences affecting DMPs are gathered from literature review, and two research hypotheses are tested via analysis of survey data collected: 1) the quality of decision-making varies across the different roles, 2) the earlier the collection of information, alternatives, values, and preferences happens, the better and more informed decisions will be made.

The findings of this research are beneficial to design and construction organisations, practitioners, researchers, and stakeholders in understanding factors affecting DMPs. This study provides a new means to evaluate decision-making performance by developing and applying a quantitative data-driven, evidence-based methodological framework. The current analysis built an applicable framework that assessed the importance of 32 variables in total, quantifying the way designers make decisions and capturing the social aspects introduced by the technological advancements, and increasing data availability and their use in the design process to be evaluated and optimized.

3.3 Decision-making in building and urban design

The design of a building or its surroundings is based upon the synthesis of ideas that are constrained by multiple parameters (Kensing, et al., 1998; Kim & Ryu, 2014). A “design- team” comprised of various specialists, such as architects, civil and structural engineers, traffic managers, and others, often referred to as a “multi-disciplinary” team, is assigned at the start of every project (Denton, 1997). These specialists develop a process of communication among individuals and within diverse groups. Each group is subject to several influences which collectively determine the distribution of the individual communication patterns. The patterns of communication, and hence the final decisions, are influenced by four principal variables: The type of decisions (decision criteria); The individual who made the decisions (the decision-maker); The time and circumstances (decision environment); The time and decision-support required while making the decisions.

Efforts in understanding the human information processing in decision-making in design were initially introduced in the 1980s (Ernest, 1982; Stauffer & Ullman, 1991). These first studies revealed that design processes were characterised by either comparing alternatives to criteria, such as requirements or constraints, or by evaluation, in which the next steps were formed by it. In parallel to these experiments, efforts in designing computer-based tools began, seeking to process design information to understand better designers’ intent, such as the rationale behind their decisions (Chen, et al., 1990; Ullman & Paasch, 1994). These first experiments were unsuccessful due to the multiple factors revealed and the associated complexity. Based on this experience, Ullman and Paasch (1994, p. 1) defined design as “*the evolution of information punctuated by decisions*”.

In that same period, the concept of “Knowledge Society” also began to emerge, recognising that information alone would not bring about significant change (Böhme & Stehr, 1986; Lamberton, 1994; Lytras & Sicilia, 2005). Instead, the key to effective decision-making is how people transform information into knowledge and subsequently manage that knowledge (Kirkman, et al., 2002). Therefore, it can be argued here that the potential value of information is revealed when leveraged to drive DMPs. The increased amount of information and advanced BD analytics can reveal insights concerning the way people interact with the built environment, enabling improved decision-making in critical development areas, such as the design of a city.

3.3.1 Decision makers

The seminal work of Pettigrew (1990) posed the question as to whether the decision problem itself shapes the process to a greater extent than the organisational context through which the process progresses. Rajagopalan et al. (1997) suggested a priority for examining the degree of influence by which variation of DMPs are affected by variations in organisational, environmental, and managerial factors. Adding to this complexity, BD now introduce new challenges, focusing on methodologies, technologies, and tools, in which the decision-maker exists in each procedure of BD analysis (Wang, et al., 2016). More recent studies have attempted to present improved environment-related decision-making as an approach that incorporates intentional processes by which specialists (e.g., ecologists), stakeholders, and decision-makers (DMs) work collectively (Enquist, et al., 2017). Researchers indicate that planning and design decisions are by nature complex and subject of conflict due to the increased number of involved stakeholders (Lahdelma & Salminen, 2000; Hall & Davis, 2007). These stakeholders have different preferences and value judgements, operating at several levels. Individuals affect the outcome of decision-making. Decision-making is concerned with how people interpret problems, form goals, and combine information to arrive at solutions (Bruch & Feinberg, 2017). Individuals differ in terms of capacity to undertake such tasks, resulting in either poor or effective decisions. Poor decisions can result in errors and inadequate use of resources, while effective ones can lead to quick and efficient achievement of goals.

Factors contributing to this include cognitive biases, previous experiences, or even level of commitment, values, and beliefs (Hilbert, 2012; Ghattas, et al., 2014; Gal & Pfeffer, 2008). Cognitive biases represent thinking patterns based on observations that lead to memory errors or inaccurate judgements (West, et al., 2008). When cognitive biases influence decision-making, information perceived as uncertain may be dismissed based on observations and prior knowledge. However, although this may lead to poor decisions, in many cases, this may also enable individuals to make efficient decisions with the assistance of heuristics (Shash, 1993). Heuristics can be described as the mental capacity to solve problems based on previous experiences.

Nevertheless, there are internal analytical processes in the DMPs that can lead to differing interpretation and assimilation and differing speeds of decision-making, even when these are based on the same data (Papadakis, et al., 1998; Kirsi, 2011; Santagata & Yeh, 2016). Confidence in choice-making or the level of risk that a decision-maker is willing to take varies. In some cases, the age of the decision-maker can be proved to be significant, as described by Finucane et al. (2005), noting the decline in cognitive functions as age progresses and the associated potential overconfidence in the ability to make good decisions (De Bruin, et al., 2007). All these factors relating to the individual, such as beliefs and values, bring additional perspectives to decision-making.

3.3.2 The decision environment

An additional influential parameter is the decision environment. Every decision is made within an environment defined as "the collection of information, alternatives, values, and preferences available at the time of the decision” (Harris, 2012), para. 12. According to their study, Qiu et al. (2013) established that diverse stakeholders might have similar preferences as to what constitutes the provision of good quality urban green spaces, with differences in some aspects within these spaces. For example, a study of ecology students found that they had higher tolerance towards the view of natural processes such as decay compared to designers (Qui, et al., 2013). Another study revealed that providers and researchers involved with landscape planning appreciated the provision of rare species, which proved, however, of less value to the local users, such as people who live in the area or were regular users of the spaces (Tempesta & Vecchiato, 2015). An earlier study also established that although local users are more likely to visit formal or well-designed “artificial” spaces, the providers involved with landscape planning emphasised the presence of natural green spaces (Hofmann, et al., 2012). Further to these parameters, social pressures, such as the approval or disapproval of people in the immediate environment, such as friends and colleagues, may affect decision-making and the ranking of factors to a given situation (Pereira, et al., 2016).

Although these differences are present, the reasons for their existence are still not clarified, given that an individual’s preferences are shaped by a wide range of subjective factors (Chan, et al., 2016; Cooper, et al., 2017). The decision environment itself can be of great complexity. Changes in the decision environment impact DMPs, introducing risk and uncertainty. “Risk” in decision-making implies that the possible outcomes of a decision are known or can be determined based on the probability of each outcome. As the decision environment continues to expand following the integration of new information and alternatives, decision-making occurs as close to its deadlines possible (Harris, 2012). This presents benefits, such as the emergence of more information and alternatives. However, it also offers several risks. For example, the decision-maker feels overwhelmed by the breadth of available information.

Therefore, subjectivity forms an essential component in decision-making. Hence, there is the risk of the decision-maker's preferences changing throughout the process, leading either to more informed decisions or equally to poorer ones (Carroll, 1987; Davis, et al., 1992; Hassenzahl, et al., 2010; Malone, 1981; Venkatesh, et al., 2003).

3.3.3 Decision criteria

The criteria affecting decision-making are established and used to evaluate alternative courses of action in DMPs, and they will, in turn, affect the outcome of a decision. Different criteria are suitable in different situations. In group decision-making, decision-makers will explore the criteria, for example, by applying different weightings, while the specific criteria chosen can help identify areas of exploration. Although these differences in perception and reaction to the criteria are present, the reasons for their existence are still not clarified, given that an individual's preferences are shaped by a wide range of subjective factors (Chan, et al., 2016; Cooper, et al., 2017).

The creation and prioritisation of metrics for the decision-making process is needed to minimise the project risks involved in urban design (Madu & Georgantzas, 1991). Project risks involve the responsibility of meeting cost, quality, and time objectives. Hansen and Ahmed (2004) reviewed existing literature in design decision-making, concluding that the design process constitutes a series of decisions taken repeatedly. They presented a conceptual model in which all the decision episodes were recorded within a node, based on four distinct phases: Evaluation, Validation, Navigation & Unification, based on the evaluation and decision-making activities. Later work identified decision-making as a non-linear process, in which most decisions are made by revisiting the choice of criteria and the alternatives multiple times throughout the process, rather than being formed in a context of other decisions not in isolation (Harris, 2012). Janssen et al. (2017) identified 11 factors influencing decision-making based on BD via interviews and literature review. These included process transformation and integration, development of skills, data quality, the flexibility of systems, collaboration, knowledge exchange, and decision-maker quality as some of the key factors identified.

3.3.4 Time

Decisions must be made within a specified time period and a set of circumstances. Decision-making has been characterised by Harris (2012) as a non-linear process. Most decisions are made by revisiting criteria and the possible alternatives multiple times throughout the process. Heuristics are a common decision-making strategy that permits decision-makers to arrive at a correct and viable decision while working with little information. Shah and Oppenheimer (2008) argue that heuristics reduce work in decision-making, as they can be described as mental short-cuts, diminishing the work of retrieving information and streamlining the processes by reducing the amount of integrated information to those necessary for making a choice.

3.3.5 Decision support – theories, tools, and techniques

As settings vary, supporting activities may influence stakeholder perspectives positively and negatively; therefore, it is necessary to have mechanisms in place to support decision-making effectively within the design process. A wide variety of risk frameworks have been produced to date to guide decision-makers through various DMPs in a clear and transparent way that can be replicated (Snyder, et al., 2014; Quinn & Cockburn, 2020). While not entirely removing subjectivity and biases, such frameworks can help based on the assumption that the decision made is entirely or bounded rationally (Wang, et al., 2016; Elgendy & Elragal, 2016). Quantification and standardisation of the quality of decision-making can improve the processes and reduce the existing quality gaps. Additionally, specific skills are required to judge the urgency of DMPs in each of the stages of the process, using the available resources, such as prioritising and recognizing the potential benefits or costs of a choice via data analysis (Dutilh, et al., 2019).

BD approaches enrich the content and scope of any DMP by presenting a plethora of new information, which in many cases, is acquired in real-time. However, solutions need to be implemented to handle and extract knowledge from BD, and decision-makers need to gain valuable insights from rapidly changing information (Elgendy & Elragal, 2016). Hence, BD approaches add to the complexity, effort, and time required to reach a decision, complicating even further the role of the decision-maker.

3.4 Materials and Methods

3.4.1 Methodological Framework

The authors developed and applied a quantitative data-driven, evidence-based methodological framework utilising qualitative information (Figure 3-1). A questionnaire survey was conducted to explore the stakeholders' perception of decision-making within the design process and identify the factors influencing DMPs. The data extracted from the questionnaire survey was analysed using Exploratory Factor Analysis (EFA), Average Relative Importance Index (ARII), and Spearman Rank Correlation Coefficient Test (rs) methods (Prasad, et al., 2018; Olomolaiye, et al., 1987; Jarkas & Younes, 2012; Stanitsas & Kirytopoulos, 2021). These methods are explained in the following sections.

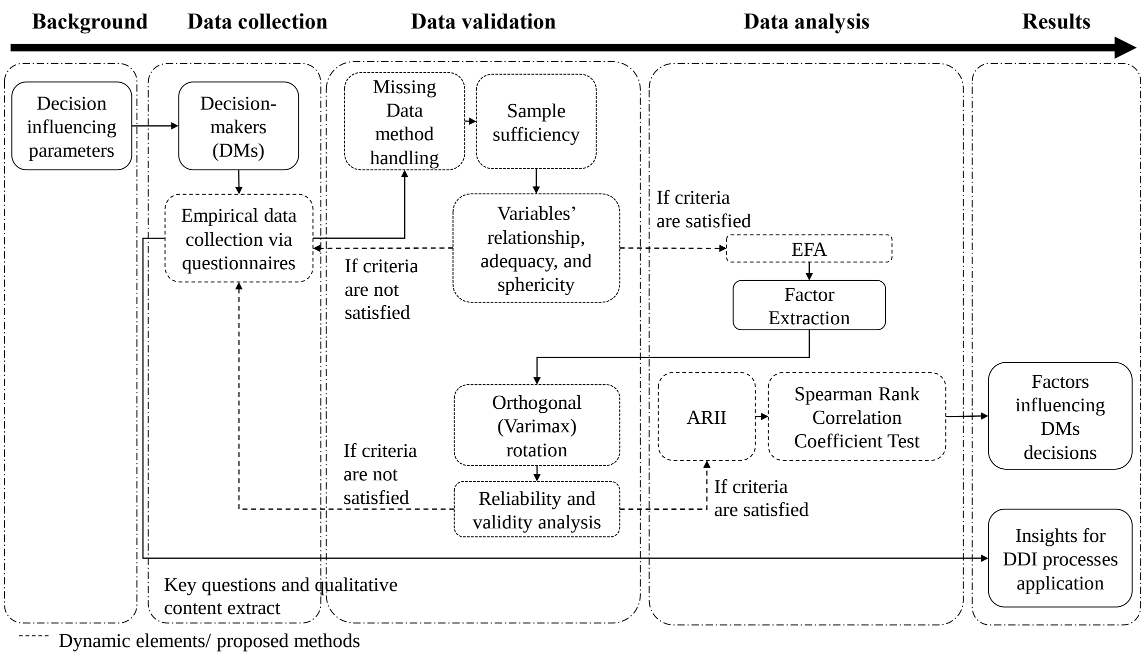


Figure 3-1: Overview of the proposed methodological framework

3.4.2 Empirical data collection and questionnaire design

The chosen data collection research method serves the purpose of gathering data from a large sample on a global scale (Venkatesh, et al., 2003). Questionnaire surveys are proven to provide reliable results for evaluating variables, measuring behaviours of a large sample of individuals of interest in a wide variety of contexts (Hansen & Andreasen, 2004; Ponto, 2015). Furthermore, questionnaire surveys present internal/external validity and, under proper construction, ethical advantages (Madu & Georgantzas, 1991; Stanitsas & Kirytopoulos, 2022).

In an exploratory survey of 136 participants, stratified by their involvement in the design DMP, decision-making quality was measured using an online questionnaire. A pilot questionnaire was issued, completed, and tested by 13 stakeholders before the commencement of the main cohort of participants. The questionnaire was designed for all levels of designers and project managers within the construction industry, and its structure was based on typical decision pathways and the rational model (Harris, 2012). Thirty-eight questions were included, forming four distinct categories in the questionnaire: Part 1: Questions concerning the respondents' role in the organisation and role in the team, Part 2: Evaluative questions concerning decision-making quality, Part 3: Questions concerning the recency of tools to improve design quality, and Part 4: Potential for dynamic operation. Variables of Part 2, 3 and 4 are presented below in Table 3-1 and Table 3-2.

Table 3-1: Review of variables (individual questions asked)

Q1	Do you enjoy making decisions?
Q2	Do you rely on instincts when making decisions?
Q3	Do you consult others, when making decisions?
Q4	Do you stick by your decisions through the design?
Q5	Once you have one option which works, do you leave it like that and proceed?
Q6	Do you remain calm when you have to make decisions in strict timelines?
Q7	Do you feel in control of things, when making decisions?
Q8	How often are your decisions influenced by your ideals regardless of practical difficulties?
Q9	Do you make decisions without considering all the implications?

Q10	Do you change your mind through the process about the decisions already made?
Q11	Do you take the safe option if there is one?
Q12	Do you prefer to avoid making decisions if you can?
Q13	Do you plan well ahead?
Q14	When making decisions do you find yourself favouring first one option then another?
Q15	Do you carry on looking for something better even if you have found a course of action that is just about OK?
Q16	Do you find it difficult to think clearly when you have to decide something in a hurry?
Q17	Do you make up your own mind about things regardless of what others think?
Q18	Do you avoid taking advice over decisions?
Q19	Do you work out all the pros and cons before deciding?
Q20	In your decision making, how often are practicalities more important than principles?
Q21	Is your decision making a deliberate logical process?
Q23	How often do you use the digital tools in a typical week?
Q24	I feel confident using the digital resources given to me.
Q25	To what extent do you think the design process is digitalised, i.e., use of digital tools that optimise design or decrease time spent in repetitive tasks
Q26	To what extent do you agree or disagree that digital technologies have the potential to replace conventional design workshops and the hands-on experience?
Q27	To what extent do you agree or disagree that your design practice would improve if you used a completely digital process?
Q28	Data-driven innovation (DDI) refers to innovative applications derived from data analytics.

Q30	To what extent do you think data-driven decision-making is part of the organisation's culture?
Q31	To what extent do you think the use of data in the design process would benefit your organisation?
Q32	To what extent do you think new technologies, data tools, modelling and visualisation can improve decision-making processes resulting in better designs?
Q33	To what extent do you think monitoring, integrating sensors, citizen science, public engagement and inclusion can improve decision-making processes resulting in better designs?
Q34	To what extend do you think infrastructure, data accessibility and usability for stakeholders can improve decision-making processes resulting in better designs?

Participants represented different hierarchy levels, roles, and experiences in practice and design. They responded to each of the categories using a five-point Likert- type scale (1-5), chosen as it is considered to ease the interpretation of the results indicated by several researchers (Shash, 1993; Gigerenzer & Goldstein, 2002; Hilbig & Pohl, 2008). In addition, the 5-point scale enhances the positive emotional limit for the respondents, who are called upon to evaluate the significance of pre-defined statements. Hence, this approach was selected to ensure the increased contribution of the respondents (Sachdev & Verma, 2004). For questions Q12 and Q16 the ranking is a five-point ranges from very infrequently or never (1) to very frequently or always (5). Due to the negative aspect of the Q12 and Q16 variables and based on the results of the factor analysis, the Likert scale have been inverted to very infrequently or never (5) and very frequently or always (1).

Due to selecting a specific domain (building and places design industry), this research did not follow a population-based sampling (random group of participants). Still, it used convenience sampling (targeted participants based on professional relativity and participatory willingness) (Brodaty, et al., 2014). This method was selected to avoid complications of dealing with a randomised sample and to obtain primary data and trends relevant to DMPs within the design process.

Three additional questions were asked at the end of the questionnaire to gather further insights concerning the Data-Driven Innovation (DDI) processes application (Table 3-2). Participants were asked questions about where the application of these processes would be useful. Furthermore, questions were targeted to understand their perception of when and how this could be applied. Participants had the opportunity to select more than one possible answer, and no ranking was required. Finally, the option to include comments at the end of the questionnaire was provided to the participants. Their responses have been used, via direct quotation of participants, to provide a detailed understanding of key themes (Ernest, 1982).

Table 3-2: Additional questions included in the online survey.

Q29	Which of the design sectors in your organisation are involved in using data technologies and data analytics?
Q35	What would you feel about a completely digital design process for the built environment and its surroundings?
Q36	What would you feel about a completely digital design process combined with data analytics approaches for the built environment and its surroundings?

Finally, Q22 was initially included in the questionnaire survey aiming to gather insights of the types of tools decision-makers use for the designing process of the built environment and its surroundings. Q22 was a multiple-choice type of question, however, it has been excluded from the analysis due to the lack of valuable responses and insights.

Participant profiles were identified through LinkedIn and consultancy organisation companies following a direct approach. A questionnaire survey was issued to the selected stakeholders (respondents), composed of experts, and consisting of academics and industrial practitioners with previous experience in the fields of project management, construction, and design. One thousand five hundred (1,500) invitations were sent out to the selected experts (stakeholders), inviting them to complete the online questionnaire. The survey was conducted from December 2019 to March 2020.

3.4.3 Respondents' profile

Usable questionnaires were returned by 136 respondents, comprising 14 stakeholders' categories. The stakeholder categories, along with the respondents' professional background, is summarised in Figure 3-2. The identified stakeholders were located worldwide to obtain a diverse sample and be able to export a representative conclusion based on the following criteria: frequency of decision-making; being part of the stakeholder groups; and level of experience in the use of DDI.

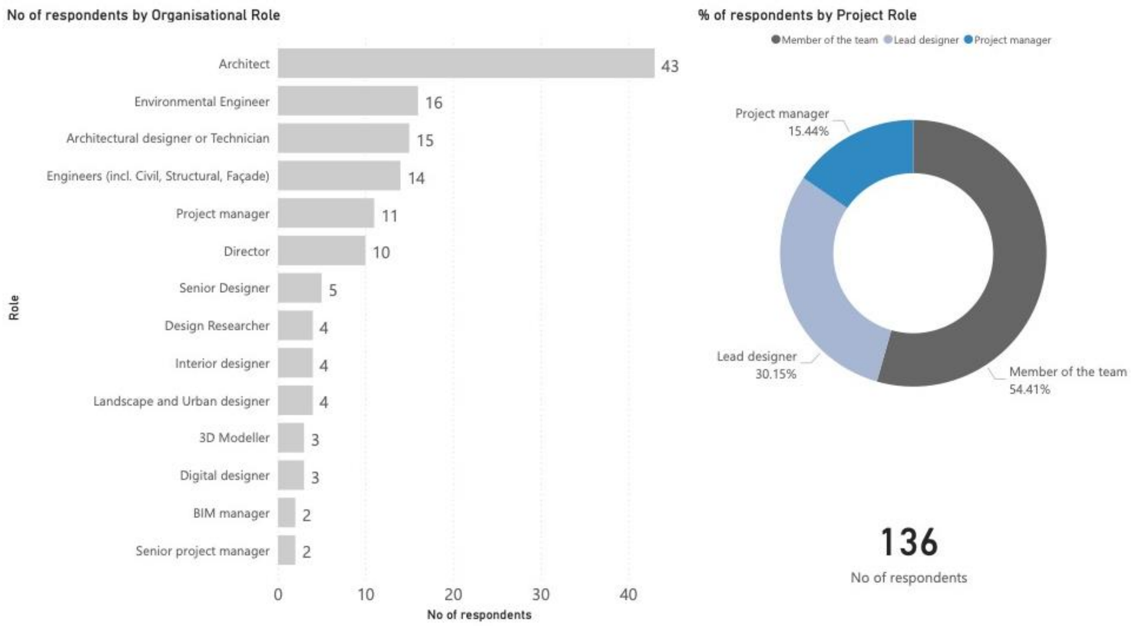


Figure 3-2: Respondent's professional background reflecting their organisational role and their project roles

3.4.4 Missing data

The chosen mechanism of addressing missing values is Missing Completely At Random (MCAR) (Little & Rubin, 1987). The assessment of the structure of missingness considers two factors: (1) the percentage of missingness and (2) missing imputation methods. As described by Chen (2012) and Dong and Peng (2013), the different missing imputation methods do not have any significant impact when the percentage of missingness is low. In this study, the percentage of missing values is less than 5% of the total sample. Therefore, a listwise deletion method, fully deleting the missing values, was performed.

3.4.5 Data Analysis and Validation

The responses elicited from the questionnaire survey were analysed with a statistical software "Statistical Package for Social Sciences (SPSS)" from IBM (IBM Corp., 2019). An understanding of the underlying characteristics of multiple variables is required. Therefore, Factor Analysis (FA) was used as the chosen statistical method due to the benefits that this technique offers when dealing with multiple variables (32 in total) (Alkarkhi & Alqaraghuli, 2019). More specifically, FA helps categorise these variables into distinct groups, making the analysis easier (Ghosh & Jintanapakanont, 2004). The chosen factor analysis type is Exploratory Factor Analysis (EFA). This method is used to reveal the fundamental concepts amongst a large set of factors, thus helping to understand the underlying structure of complex data (Prasad, et al., 2018). The key objective of the EFA is to determine the minimum number of factors needed to produce correlations among the observed variables. The application of EFA is relevant in this study as most of the variables involved cannot be quantified. Therefore, this method proves to be particularly useful in this study, as the qualitative approach is a suitable technique for collecting data, while quantitative analysis supports improved reporting (Oller, 2014). Variables such as best practice strategies in DMPs need to be measured as observed variables.

Two tests were performed to identify whether EFA provides a suitable type of analysis and if the sample of data is appropriate: the Kaiser–Meyer–Olkin (KMO) test, used to determine sample sufficiency, and Bartlett's sphericity test was performed to examine the variables' relationship, adequacy, and sphericity (Delmonico, et al., 2018). The correlations of the variables result from the use of the orthogonal (Varimax) rotation and oblique rotations that allow the variables analysed to correlate and finally decide on the factor space (Field, 2009). The total number of factors was concluded by studying the Kaiser's criterion and the eigenvalues (Hinkin, 1998). Field's (2009) work indicates that the Kaiser's criterion is more accurate when the sample size exceeds 250. Due to the lower number of responses for this study, the Kaiser's criterion and the scree plot were considered.

The reliability value (Cronbach's alpha) and the KMO test were performed for all the factors, following Karekla and Michaelides (2017) and Almeida et al. (2016) research "path" for the unifactorial structure of factors is a way to verify construct validity. However, when there is a small number of items on the scale (fewer than 10), Cronbach's coefficient alpha values can be quite small. For this study, both Cronbach alpha values and the mean inter-item correlation for the items were calculated.

The participants expressed opinions on the level of agreement for each variable transformed on numerical scores, with values from 1 to 5. For this type of analysis, the mean and standard deviation of the variables were not considered suitable to determine the overall ranking (Chan & Kumaraswamy, 1997). Therefore, a calculated weighted average for each factor, divided by the upper measurement scale, was adopted to provide an importance index (Shash, 1993; Ghosh & Jintanapakanont, 2004). ARII values have been calculated for the defined factors to be considered with reliability in the collected data using, including all ten factors initially and the average value for each factor (ARII) (Olomolaiye, et al., 1987; Jarkas & Younes, 2012; Waris, et al., 2014). ARII values were extracted for diverse stakeholder groups, divided by role within a team, role within the discipline, and experience in DDI.

Equation 3-1: ARII

$$\mathbf{ARII} = \frac{\sum_{i=1}^{5} w_i n_i}{A * N} = \frac{5n_5 + 4n_4 + 3n_3 + 2n_2 + 1n_1}{5 * N}$$

w is the constant that shows the weighting of each response (1 – not important up to 5 – extremely important)

ni is the frequency of the responses (i = 1 to 5 based on the Likert scale)

N is the total number of the responses.

The ARII value ranges from 0 to 1, with 0 not inclusive. The higher the value of ARII, the more important the criteria. The comparison of ARII with the corresponding importance level is evaluated from the transformation matrix as proposed by Chen et al. (2010) and shown below.

- High (H): 0.8 < ARII < 1.0
- High-Medium (H-M): 0.6 < ARII < 0.8
- Medium (M): 0.4 < ARII < 0.6
- Medium-Low (M-L): 0.2 < ARII < 0.4
- Low (L): 0 < ARII < 0.2

Additional analysis was then followed with Spearman Rank Correlation Coefficient Test (rs) to identify agreement or disagreement among the diverse groups on ranking factors. This test is generally used to understand agreement on the relative importance of the identified factors (Jarkas & Younes, 2012). Spearman Rank Correlation Coefficient Test (rs) values range between -1 and +1, where a perfect negative correlation is rs = -1 and a perfect positive correlation is rs = +1. Values close to 0 indicate low or no correlation (AlSehaimi, et al., 2013).

3.4.5.1 Preliminary analysis and factor extraction

The response rate, based on the 136 returned questionnaires, reflects a confidence level of 95% with an 8% margin error. In addition, in their studies, Vishwakarma (2017) and Field (2009) indicate that a minimum sample of 100 responses is required for extracting valuable results. Therefore, this response rate meets the minimum sample size threshold for further analysis to follow. Initial analysis showed a KMO value of 0.599, and Bartlett's sphericity test was significant ($p<0.001$), revealing that the variables' multicollinearity is at an acceptable level. To better understand why the KMO value is less than 0.60, the anti-image correlations for the pairs of variables – the individual questions asked - in the anti-image matrix were considered and revealed some variables with values lower than 0.5, which is the satisfactory limit (Field, 2009). The variables with a value lower than 0.4 were initially removed, and the tests were performed again with 30 variables in total. The variables removed were Q18 and Q30 (Table 3-1). The KMO value was then 0.627 (where KMO>0.6 is considered acceptable), and Bartlett's sphericity test was significant ($p<0.001$) (Table 3-3). Therefore, it was considered that the EFA technique could be employed.

Table 3-3: Results for KMO and Bartlett’s Test indicating that EFA technique can be employed

KMO and Bartlett’s Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.627
Bartlett's Test of Sphericity	Approx. Chi-Square	1183.717
	df	435
	Sig.	.000

Commonalities (h2) represent the proportion of the variance in that variable that can be accounted for by all extracted factors. The preliminary analysis from the anti-image matrix revealed that the commonalities of all variables are between 0.507 and 0.813. These findings indicate the valuable contribution of all the 30 variables in DMP (Child, 2006) (Table 3-1).

Although the Kaiser's criterion, the scree plot (Figure 3-3), supports the choice of eight factors in total as being the most appropriate, the cumulative variances for eight factors do not meet the 60% threshold (Hinkin, 1998) (Table 3-4). Therefore, the authors used ten factors where cumulative variances are 63.160%, which is acceptable. The authors formulated the naming of the ten factors to reflect the strategic perspective of the DMPs while encapsulating the underlying concepts of the variables included in each factor.

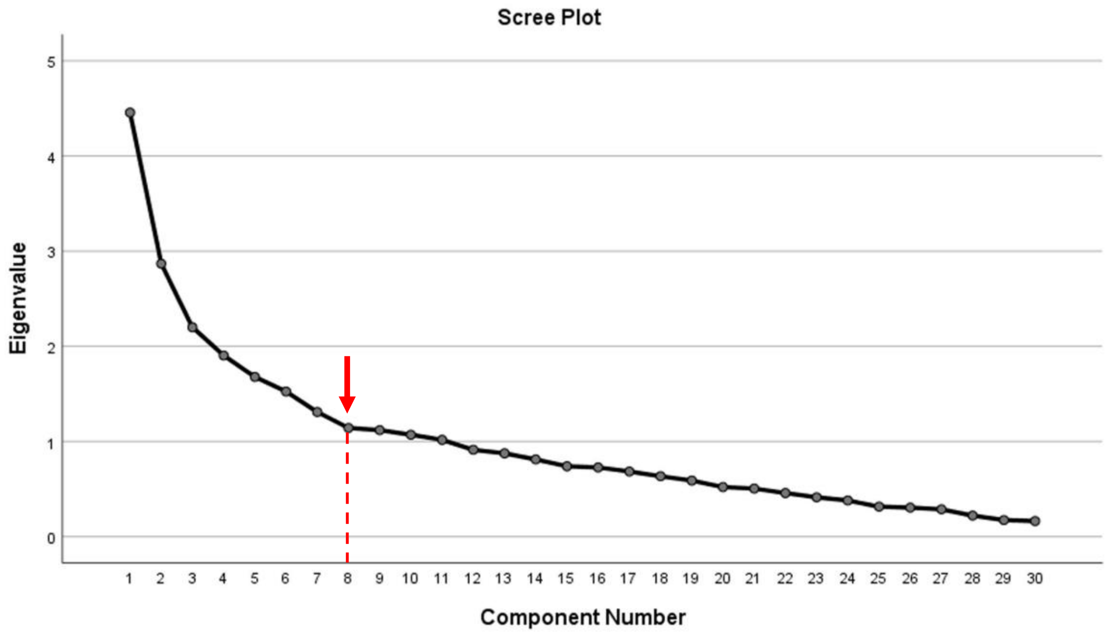


Figure 3-3: Scree plot indicating the choice of eight factors (component number) as being the most appropriate

Factor loadings show the correlation of the factors (factor dimension) with the variables (Costello & Osborne, 2005). The higher the factor loadings are, the higher their significance for the DMP (Ghosh & Jintanapakanont, 2004). The correlations of the variables are the result of the use of the orthogonal (Varimax) rotation and oblique rotations that allow the analysed variables to correlate and finally decide on the factor space (Field, 2009) (Table 3-4). The overall reliability analysis was performed together with Cronbach's coefficient alpha, having a value of 0.717, regarded as acceptable (Nunnally & Bernstein, 1994).

Table 3-4: Summary results of EFA: variables of decision-quality in DMP processes and their internal consistency

ID	Factor Dimension	Variable	Factor Loading	Variance explained percentage (%)	Cumulative percentage (%)
1	**Potential for dynamic operation**	Q34	0.872	11.025	11.025
		Q31	0.822		
		Q33	0.801		
		Q32	0.754		
2	**Control**	Q12	0.749	9.190	20.215
		Q13	0.676		
		Q16	0.662		
		Q1	0.624		
		Q6	0.554		
		Q7	0.491		
3	**Instinctiveness**	Q4	0.761	6.835	27.051
		Q5	0.680		
4	**Optimising**	Q27	0.847	5.815	32.866
		Q26	0.800		
5	**Recency of tools**	Q23	0.810	5.674	38.540
		Q24	0.632		
		Q25	0.527		

6	**Thoroughness**	Q19	0.795	5.384	43.924
		Q21	0.521		
		Q15	0.423		
7	**Principled**	Q8	0.853	5.187	49.111
		Q2	0.585		
8	**Social resistance**	Q20	0.697	4.770	53.881
		Q9	0.571		
		Q17	0.564		
		Q3	0.850		
9	**Hesitancy**	Q14	0.874	4.663	58.544
		Q11	0.439		
10	**Experience**	Q28	0.703	4.616	63.160
		Q10	0.688		

3.4.5.2 Questionnaire content validity

The content validity of the questionnaire survey was based on identified variables as described in Table 3-1, which is further assessed (Wiese, et al., 2015; Almanasreh, et al., 2019). Cronbach alpha values and the mean inter-item correlation for the items was calculated and reported in Table 3-5 (Piedmont, 2014). Cohen (1988) identified that if inter-item correlation is within 0.10 and 0.29, there is a weak correlation among the variables. If inter-item correlation lies within 0.30 and 0.49 a medium correlation, while an inter-item correlation between 0.50 and 1.00 shows a strong correlation. However, Cristobal et al. (2017) noted that items with a corrected item-total correlation value up to 0.20 are acceptable for exploratory study for inter-item and item-total correlation.

Table 3-5: Factors of decision-making quality in DMP processes' validity

ID	Factor Dimension	Cronbach's Alpha	Variable	Item-Total Correlation
1	**Potential for dynamic operation**	0.859	Q31	0.695
			Q32	0.688
			Q33	0.667
			Q34	0.777
2	**Control**	0.731	Q1	0.528
			Q6	0.411
			Q7	0.441
			Q12	0.536
			Q13	0.471
			Q16	0.410
3	**Instinctiveness**	0.541	Q4	0.380
			Q5	0.380
4	**Optimising**	0.692	Q26	0.529
			Q27	0.529
5	**Recency of tools**	0.528	Q23	0.362
			Q24	0.414
			Q25	0.263
6	**Thoroughness**	0.506	Q19	0.319
			Q21	0.367
			Q15	0.290
7	**Principled**	0.431	Q8	0.280
			Q2	0.280
8	**Social resistance**	0.318	Q9	0.239
			Q3	0.058
			Q17	0.088
			Q20	0.290
9	**Hesitancy**	0.335	Q11	0.201
			Q14	0.201
10	**Experience**	0.311	Q10	0.209
			Q28	0.209

3.4.5.3 Average relative importance index ranking from factor analysis

The average relative importance index (ARII) is calculated for each factor, and the findings are presented in Table 3-6 (Ghosh & Jintanapakanont, 2004). ARII values indicated that the Potential for dynamic operation factor has the highest ranking, with High importance, followed by the Recency of Tools with a score of 0.82463 and 0.82451, respectively. Nevertheless, as described earlier, a breakdown of the ARII values was further considered due to the diverse influencing factors, based on previous experiences, disciplines, and team roles.

Table 3-6: Average relative importance index (ARII) for each factor and overall ranking

ID	Factor Dimension	Variable	Relative importance	Average Relative importance index (ARII)	Importance Level	Overall Ranking
1	**Potential for dynamic operation**	Q31	81.912	0.82463	H	1
		Q32	84.853			
		Q33	81.765			
		Q34	81.324			
2	**Control**	Q1	81.471	0.74289	H-M	4
		Q6	75.735			
		Q7	75.294			
		Q12	76.618			
		Q13	76.029			
		Q16	60.588			
3	**Instinctiveness**	Q4	71.324	0.67279	H- M	6
		Q5	63.235			
4	**Optimising**	Q26	65.294	0.66985	H- M	7
		Q27	68.676			

5	**Recency of tools**	Q23	84.559	0.82451	H	2
		Q24	84.412			
		Q25	78.382			
6	**Thoroughness**	Q19	80.147	0.77304	H-M	3
		Q21	79.853			
		Q15	71.912			
7	**Principled**	Q8	64.265	0.64412	H- M	8
		Q2	64.559			
8	**Social resistance**	Q9	81.471	0.70772	H-M	5
		Q3	79.118			
		Q17	56.029			
		Q20	66.471			
9	**Hesitancy**	Q11	60.29412	0.60735	H -M	9
		Q14	61.17647			
10	**Experience**	Q10	60.29412	0.5125	M	10
		Q28	42.20588			

3.5 Results

3.5.1 The influence of previous experiences, disciplines, and team roles on decision-making processes

A comparison among the levels of experience, disciplines, and team roles was conducted to identify the influence of these parameters on DMPs. The comparison was performed, with ARII values calculated for each of these groups.

3.5.1.1 Previous experience

A comparison between levels of experience regarding the ARII values has been conducted, and findings are shown in Table 3-7. The results indicate that the *Instinctiveness* and *Social Resistance* factors ranked relatively high for the non-experienced users, in addition to the experienced users, in which *Instinctiveness* ranked almost closed to value 0 (~ 0.013). The experienced users in DDI considered *Control* as the most important factor, with an ARII value of ~0.79, while the non-experienced ones, the *Recency of Tools*, with a score of ~0.82.

Table 3-7: Summary results of ARII values and the factors' ranking among stakeholders with different levels of experience in DDI implementation

	Role of previous experience				**Overall**	
Sample Size	113		23		136	
Factor Dimension	Non- Experienced in DDI		Experienced in DDI		Overall	
	ARII	**Rank**	**ARII**	**Rank**	**ARII**	**Rank**
Potential for dynamic operation	**0.819**	**2**	**0.793**	**2**	**0.825**	**1**
Control	**0.733**	**4**	**0.797**	**1**	**0.743**	**4**
Instinctiveness	0.669	6	0.013	10	0.673	6
Optimising	0.654	7	0.63	7	0.67	7
Recency of tools	**0.825**	**1**	0.643	6	**0.825**	**2**
Thoroughness	**0.768**	**3**	**0.713**	**5**	**0.773**	**3**
Principled	0.635	8	0.43	9	0.644	8
Social resistance	**0.71**	**5**	0.53	8	**0.708**	**5**
Hesitancy	0.602	9	**0.778**	**3**	0.607	9
Experience	0.462	10	**0.717**	**4**	0.513	10

The questions and results of the relevant variables regarding the Control factor are shown in Figure 3-4 for experienced users. Results indicated that there is confidence in planning ahead and enjoyment of the DMPs due to the participants' previous experiences. Control is one factor to consider when dealing with non-experienced decision-makers.

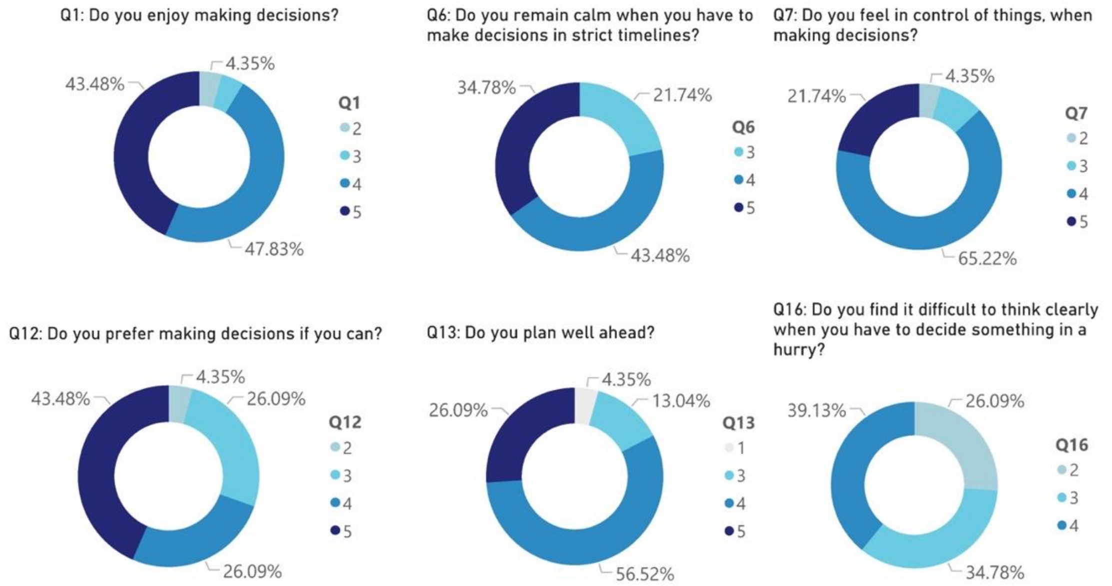


Figure 3-4: Variables of Control factor for experienced users only (inclusive of replies with ranking of 4 and 5). (1 = Very infrequently or never and 5 = Very frequently or always).

Another interesting aspect for the experienced decision-makers in the use of DDI was that the *Hesitancy* factor ranked high, as of high - medium importance, in the third position with a value of ~ 0.7782. In more detail, an increased number of participants rated as sometimes (3) their answers to questions Q11: "*Do you take the safe option if there is one?*" and Q14: "When making decisions do you find yourself favouring first one option then another?"(Table 3-1). The percentages of the responses with a value of 3 in these questions were 39.1% and 65.2%, respectively. In addition, the same number of people rated the Q11 variable as infrequently (2) and frequently (4) with 26.1%, while 8.7% replied as very frequently or always (5), indicating that experienced decision-makers in the use of DDI are more willing to follow a riskless scenario.

For both experienced and non-experienced DDI decision-makers, the *Potential for dynamic operation* factor ranked high. It was defined as high-medium importance, with an ARII value of ~0.7935 for experienced decision-makers. For the non-experienced decision-makers, the ARII value is ~0.819 and is highly important. Results including all types of decision-makers returned similar results, with more than 79.5% agreeing with the beneficial use of data and new technologies within the DMPs of design. More specifically, Q32 (Table 3-2) had the higher percentage of agreement by the participants, with 84.6% agreeing while only 1.47% strongly disagreeing that the use of such types of data and technologies would improve DMPs in the design industry.

Non-experienced users ranked *Recency of Tools* as the most important factor influencing decision-making. Details of the included variables are shown in Figure 3-5. Due to the lack of experience, participants feel confident with the use of the digital tools provided to them and 72.6% agrees that the design process is digitalised to optimise design, therefore, DMPs in the design of buildings and their surroundings. Interestingly, the views of the non-experienced decision-makers of DDI are positive and feel that the digitalised design process assists DMPs in terms of best outcome or effort and time spent. In addition, since the Potential for dynamic operation is highly important for the specific group of participants, it indicates how decision-makers involved with the design process have experienced improvements in the project outcomes.

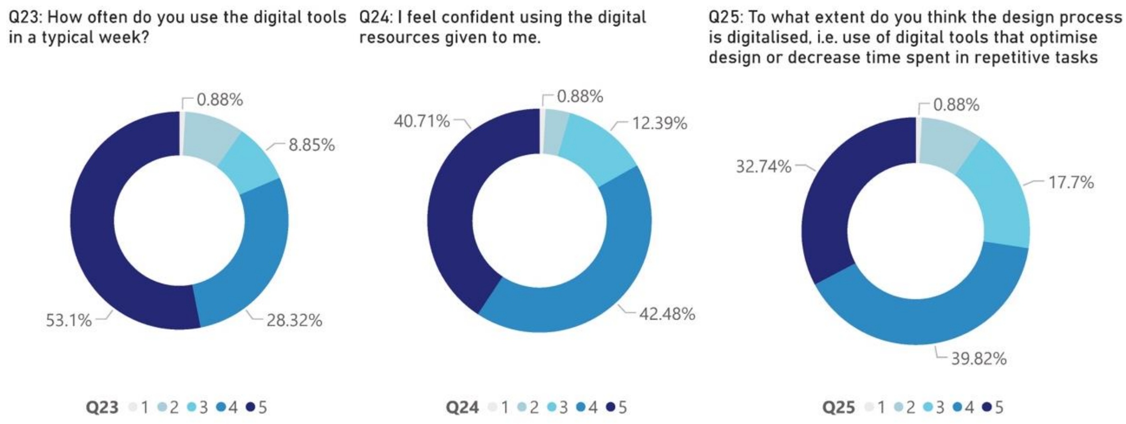


Figure 3-5: Variables of Recency of tools factor for non-experienced users of DDI only (inclusive of replies with ranking of 1,2 and 3). (For Q23: 1 = Very infrequently or never and 5 = Very frequently or always, For Q24 & Q25: 1 = Strongly disagree and 5 = Strongly agree).

Even though there is an overall positive feeling for DDI and its potential to improve DMPs, several participants raised some of the most controversial issues when dealing with real-life situations and the nature of the decision-makers as individuals. Participants referred to the practical implementation of digital approaches in the decision-making process, while others raised their concern on the human parameter involved in data analysis. Below are some supporting comments extracted from the responses. The statements include the views of an experienced and a non-experienced DDI user, respectively.

"*Completely digital design is very possible in the future. Will it lead to better decision-making regarding real-life implementation? Most probably, it will hardly affect it at all".* – Experienced DDI participant

"Tools may be useful to answer some questions. However, I am not sure about it being strategic. Decision making is deeply personal and bias in my experience. At the end of the day, data can be misrepresented to support what we would like to see and hence unreliable. Can we trust data presented?" – Non- experienced DDI participant

3.5.1.2 Discipline

Similarities in ARII values arise from people in disciplines that are heavily involved with the design process and its planning. Results from different disciplines within the building design and its surroundings are displayed in Table 3-8 below for total respondents among architects, engineers, and project managers.

Table 3-8: Summary results of ARII values and the factors' ranking among stakeholders with different roles within the company structure (discipline).

	Role of discipline						**Overall**	
Sample Size	13		27		62		136	
Factor Dimension	Project Manager Discipline (PM)		Engineers Discipline (ENG)		A&MP Discipline		Overall	
	ARII	**Rank**	**ARII**	**Rank**	**ARII**	**Rank**	**ARII**	**Rank**
Potential for dynamic operation	**0.817**	**1**	**0.85**	**1**	**0.816**	**2**	**0.825**	**1**
Control	**0.794**	**2**	**0.696**	**5**	**0.748**	**4**	**0.743**	**4**
Instinctiveness	**0.725**	**5**	0.67	6	0.678	6	0.673	6
Optimising	0.725	6	0.663	7	0.629	8	0.67	7
Recency of tools	**0.761**	**3**	**0.825**	**2**	**0.843**	**1**	**0.825**	**2**
Thoroughness	**0.744**	**4**	**0.753**	**3**	**0.771**	**3**	**0.773**	**3**
Principled	0.725	7	0.656	8	0.641	7	0.644	8
Social resistance	0.658	8	**0.704**	**4**	**0.728**	**5**	**0.708**	**5**
Hesitancy	0.65	9	0.637	9	0.603	9	0.607	9
Experience	0.525	10	0.463	10	0.497	10	0.513	10

The results indicate that the five factors ranking higher for the groups of architects and engineers are the same. However, with different rankings overall, while for the project managers, the four out of five categories remain the same. These are *Potential for dynamic operation, Control, Recency of tools, Thoroughness*, and *Social resistance*. Although it remains of high importance for the other two groups, the *Social resistance* factor is replaced in the PM category by the *Instinctiveness* factor. *Social resistance* factor is ranked in the eighth place and considered of high- medium importance for project managers, in contrast to the other disciplines, where *Social resistance* ranks in the five most important factors (Table 3-7).

The variables considered in the *Social resistance factor* indicate the need to belong to a wider team or a close collaboration (Figure 3-6). The overall low score in this factor may have resulted from the difference in decision-making processes compared to other disciplines. For example, when only one project manager role is assigned in a project rather than a team of people monitoring progress, or when there is a need for practical solutions for the sustainability of the project.

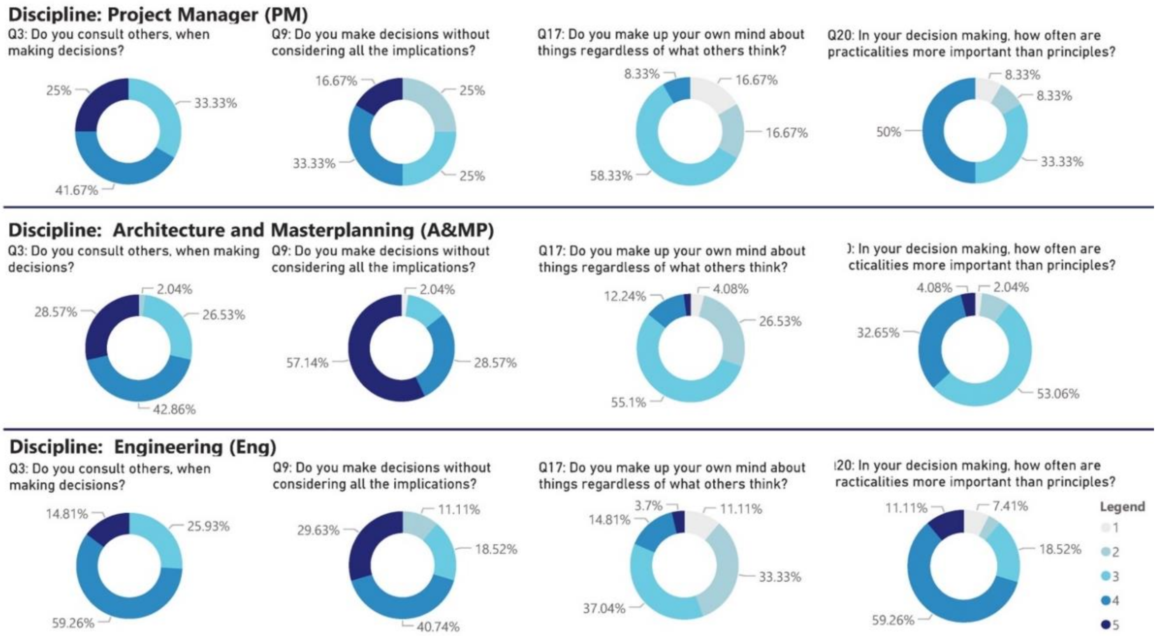


Figure 3-6: Variables of Social Resistance factor against Architecture, Engineering and Project Manager disciplines. (1 = Very infrequently or never and 5 = Very frequently or always).

3.5.1.3 Role within a team

The relative influence of design team members varies throughout the design process and based on the specific roles; the factors that influence their decisions also vary. Three key categories of the design-team members have been identified: Members of the team (MoT), Lead designer roles (LD), and Project manager roles (PM). The distinction of the project manager role as a separate from the member of the team is due to the different responsibilities associated with this role, team structure, or is not as heavily involved with the design as the rest of the disciplines.

ARII values have been calculated focusing on the key design-team categories and displayed in Table 3-9. As expected, MoT and LD showed similarities for most of the factors. For MoT, Recency of Tools scored higher than the rest of the factors, while for the LD, the Potential for dynamic operation factors holds first place. For both groups, the score is 0.84 (high importance).

Table 3-9: Summary results of ARII values and the factors' ranking among stakeholders with different roles within a team

	Role within a team						Overall	
Sample Size	21		38		61		136	
Factor Dimension	Project Manager (PM)		Lead designers (LD)		Member of team (MoT)		Overall	
	ARII	**Rank**	**ARII**	**Rank**	**ARII**	**Rank**	**ARII**	**Rank**
Potential for dynamic operation	**0.812**	**2**	**0.836**	**1**	**0.81**	**2**	**0.825**	**1**
Control	**0.813**	**1**	**0.754**	**4**	**0.715**	**4**	**0.743**	**4**
Instinctiveness	**0.743**	**5**	0.671	6	0.65	7	0.673	6
Optimising	0.733	6	0.613	8	0.666	6	0.67	7
Recency of tools	**0.81**	**3**	**0.809**	**2**	**0.835**	**1**	**0.825**	**2**
Thoroughness	**0.784**	**4**	**0.782**	**3**	**0.759**	**3**	**0.773**	**3**
Principled	0.676	8	0.637	7	0.64	8	0.644	8
Social resistance	0.683	7	**0.728**	**5**	**0.698**	**5**	**0.708**	**5**
Hesitancy	0.633	9	0.603	9	0.603	9	0.607	9
Experience	0.529	10	0.492	10	0.516	10	0.513	10

Both MoT and LD are responsible for the deliverables of the project, with the first ones having the responsibility to produce the material and the latter ones the quality to be delivered. The *Recency of Tools* factor includes three variables, most of them closely involved with the decision support methods related to the use of tools and techniques to achieve the end goal: the delivery. The *Potential for Dynamic Operation* factor is relevant to the requirements and insights for increased quality on the deliverables and the future of delivery. Decision-makers in leadership positions are responsible for client-facing communication, while in MoT roles is not a given.

Results for PM revealed that factors of *Control, Potential for Dynamic Operation,* and *Recency of Tools* are equally important and considered of high importance with an ARII value of 0.81 for all three factors. Significant differences can be observed in the individual questions posed to the participants for the *Control* factor, revealing the key differences in the DMPs for PM roles (Figure 3-7). Planning ahead, quality of delivery and future potential are considered the driving parameters for the specific role. Results imply that these differences are due to the nature of the PM as a role, which is to ensure submission of the deliverables and their quality while managing at the same time client interactions and future collaborations.

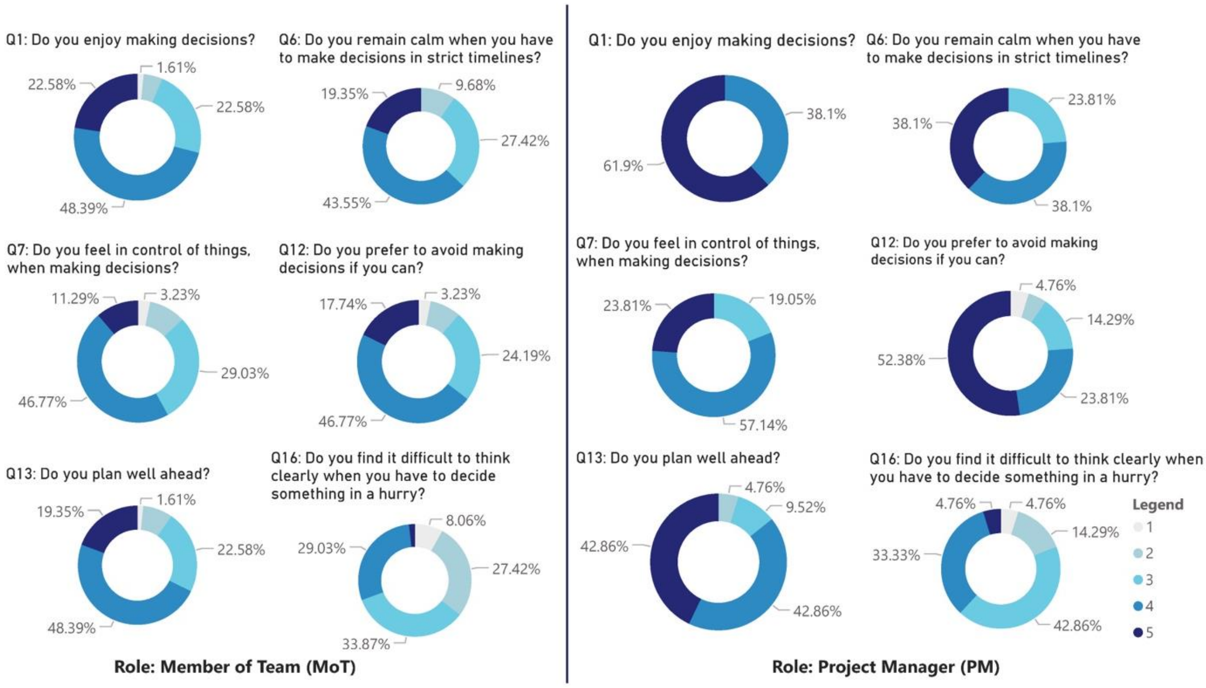


Figure 3-7: Variables of Control factor against MoT and PM roles. (1 = Very infrequently or never and 5 = Very frequently or always).

3.5.1.4 Highly important influencing factors for all decision-makers

The ARII values and factors ranking calculated for the three categories of the decision-makers are displayed in Table 3-7, Table 3-8 and Table 3-9. Results indicate that for all types, four factors are consistently ranked as high-important ones: *Potential for Dynamic Operation, Recency of tools, Thoroughness and Control*. The lowest score recorded among these factors was for *Control* when dealing with engineering disciplines, while the highest score was recorded for *Potential for Dynamic Operation* for the same group. The overall ARII ranking places *Potential for Dynamic Operation* as the most important factor, shared with the *Recency of Tools* factor. *Recency of Tools* factor is also highly important for all categories, ranking within the first five groups of factors, except for the group of experienced in the DDI process. The two factors ranking last were *Hesitancy* and *Experience*. *Social resistance* factor is part of the five most important factors considered by the respondents for five out of the nine overall categories. *Social resistance* ARII values illustrate that for some groups, this factor is of importance, while for others is not considered as important. These findings imply existence of common principles for all types of stakeholders within the buildings and urban design. However, the overall ranking of the identified factors varies significantly amongst roles, disciplines, and experiences, implying a constant need to understand the DMPs in each project better.

An implication is that based on a sample of 136 respondents, only 17% of the respondents had previous experience with DDI, based on weighting values of 4 (Piloting DDI) and 5 (Effectively using DDI). This finding indicates the lack of specialised skills within the industry and the need for training to enhance understanding of DDI processes rather than using new technologies blindly. More specifically, one of the respondents that had ranked the DDI experience with a weighting of 2 (Considering DDI) noted:

“Lack of uptake in DDI across the industry limits its benefit. Lack of understanding of software, process and resources limits current application and integration in design process.”

Another participant who ranked DDI experience with a weighting of 1 (No experience) also added a similar comment (partial extract to reflect the relevant observation):

"Digital Tools and engineering methodology are the future for construction design processes, but staff training is extremely important to ensure this vision is fully realised."

3.5.2 Level of agreement among the stakeholders' perceptions

The Spearman's Rank Correlation Coefficient was tested for the different views of stakeholders and overall ranking (Table 3-10). The stakeholder categories tested are: Project Manager roles (PM), Lead Designers (LD), Member of team (MoT), Project management discipline (PMD), Engineering Discipline (ENG), Architecture and Master planning discipline (A&MP), Experienced in DDI (EXP) and Non-experienced in DDI (Non-EXP). The test returns high coefficient values, ranging between 0.818 and 0.988, implying that there is a positive strong correlation amongst the diverse rankings against each category, indicating a high-level of agreement among the groups, signifying the consistency, validity, and reliability of these findings.

Table 3-10: Spearman's Rank Correlation Coefficients among Stakeholders within a team, discipline, and experience levels

	Role within a team					
	PM vs LD	PM vs MoT	LD vs MoT	PM vs Overall	LD vs Overall	MoT vs Overall
Spearmans Rank Correlation Coefficient (rs)	0.867	0.867	0.952	0.891	0.988	0.976

	Discipline					
	PMD vs ENG	PMD vs A&MP	ENG vs A&MP	PMD vs Overall	ENG vs Overall	A&MP vs Overall
Spearmans Rank Correlation Coefficient (rs)	0.818	0.855	0.964	0.891	0.988	0.976

	Previous experience		
	Non-Exp vs Exp	Non-Exp vs Overall	Exp vs Overall
Spearmans Rank Correlation Coefficient (rs)	0.176	0.988	0.224

The highest correlation is observed between ENG and A&MP disciplines. In addition, the test returned the same coefficient value for the PM roles against design roles (LD and MoT). These findings highlight that design driven roles follow similar decision-making processes, while they differ from the PM roles. The only low correlation value was observed between the Non-EXP vs EXP stakeholder groups, where rs = 0.176, indicating that they have different perceptions of DMPs in design, as also observed in the ARII analysis. Nevertheless, it should be noted that only 17% of the respondents had experience in DDI processes, therefore, further investigation is required for that specific group.

3.5.3 Data-Driven Innovation processes potential for implementation

Participants had the opportunity to select more than one possible answer and no ranking was required for their responses. in Q29. The highest frequencies for Q29 were recorded for the design sectors of: (a) Residential, (b) Master-planning and (c) Commercial, while the lower ones were recorded for (d) Aviation and (e) Landscape design (Figure 3-8). The rest of the sectors in the given list (Urban planning, Rail and Education) also had relatively high frequencies. Recommendations for additional sectors to be considered were made from 13 respondents overall, such as Healthcare, Hospitality, Automotive, and Institutional.

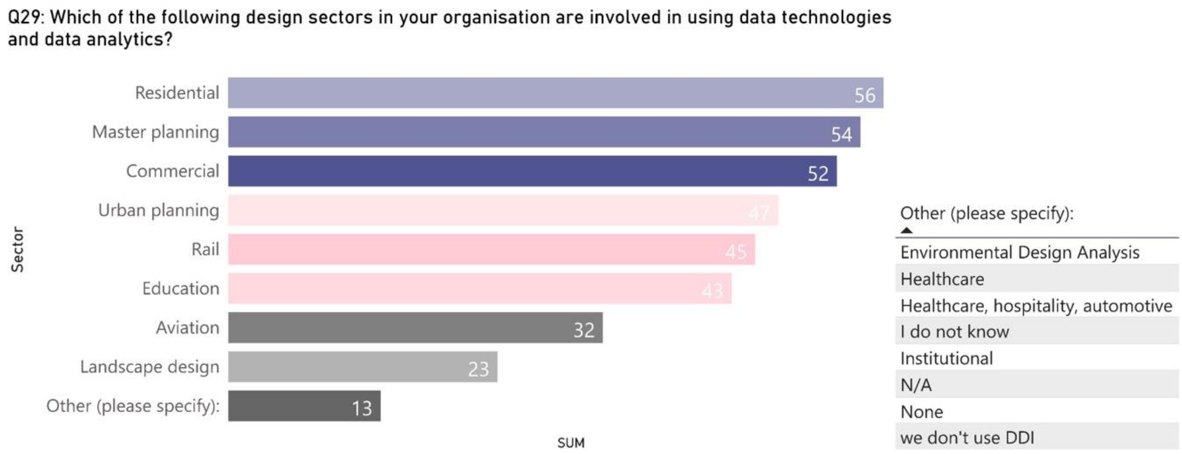


Figure 3-8: Frequencies of Q29 indicating the design sectors involved with DDI

Following that, two additional questions formed part of this study. Q35 notes the perception of people for only using digital design processes for DMPs, while Q36 investigates the feelings of the stakeholders for the combined approach of digital design process with data analytics approaches. Frequencies for Q35 and Q36 are shown in Figure 3-9 and Figure 3-10 respectively.

Regarding Q35, most of the participants responded that digital and conventional methods should both be employed, with a great number of participants stating a positive feeling of a complete application of digital design process for the built environment and its surroundings. A lower percentage of participants responded that the application of the digital design process might limit creative thinking, while several participants felt that relying on digital ways of designing is causing limitation of the design skills.

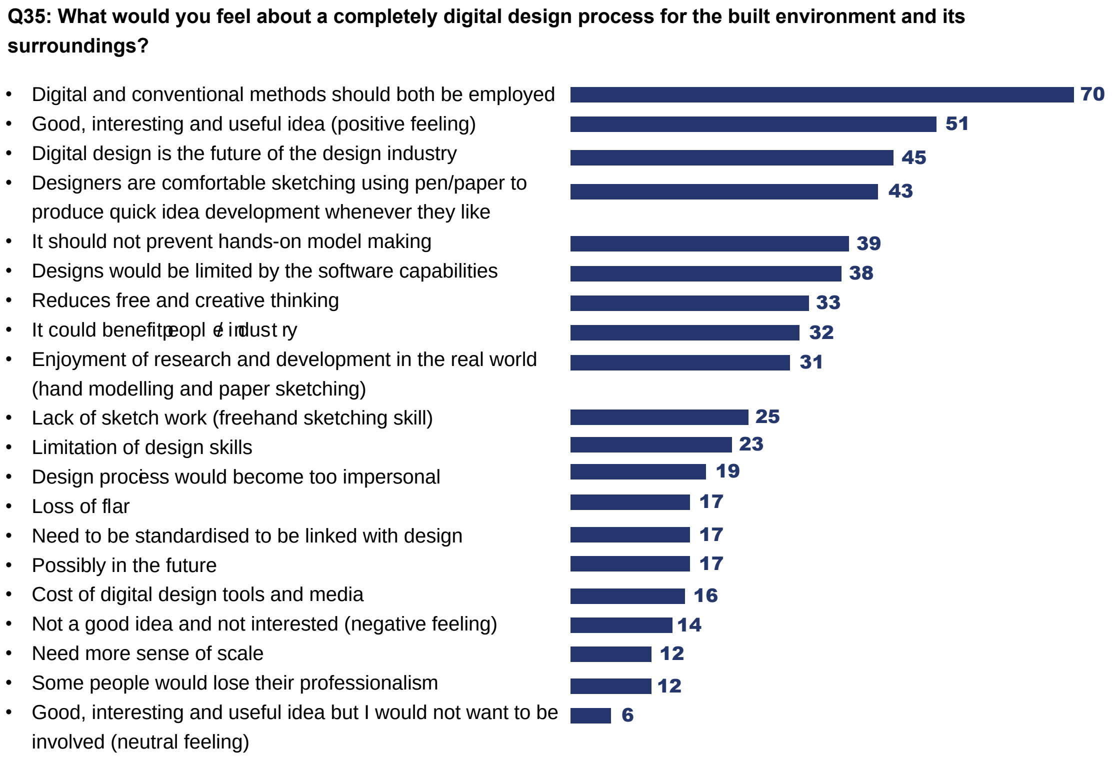


Figure 3-9: Frequencies of Q35 capturing participant responses for a completely digital design process.

Similar results were recorded for Q36 with majority of the respondents having a positive feeling overall. Significant number of participants felt that this could not be immediately applied and could be possible in the future, while others felt that this approach requires more sense of scale. For both variables, the perception of some of the participants was negative or they have stated that they are not interested in being a part in these applications. Participants also felt that this may result in a complete loss for specific professions. Another important parameter was raised which relates to the cost of these applications and the need for the processes to be standardised to be linked to the design, avoiding unnecessary complications.

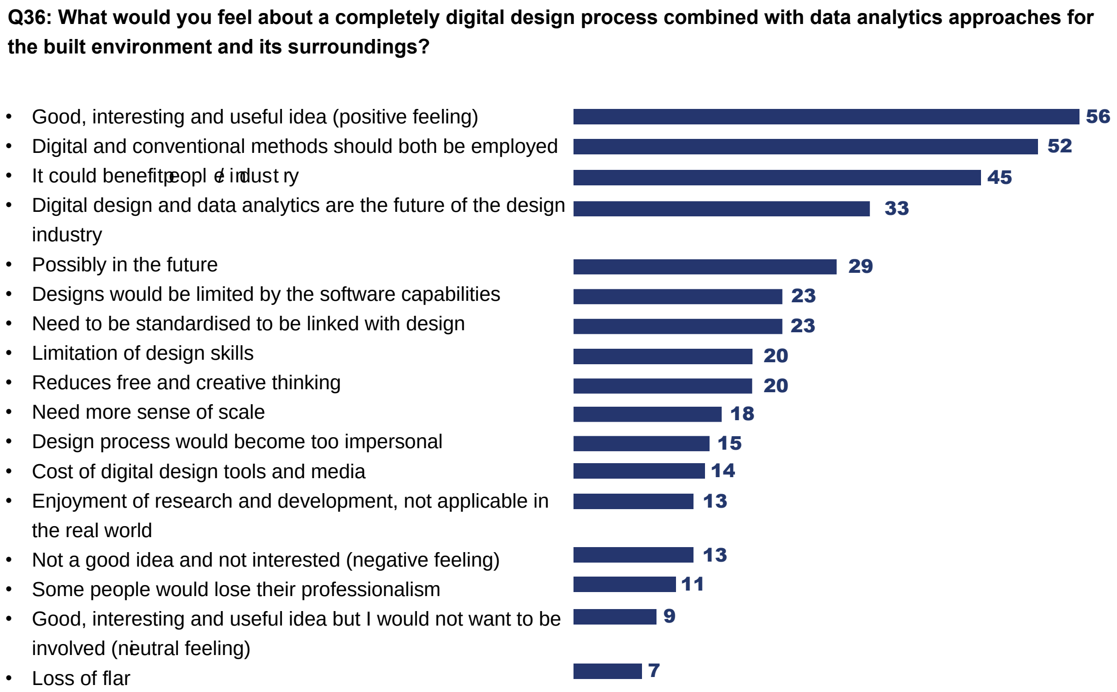


Figure 3-10: Frequencies of Q36 capturing participant responses for a completely digital design process combined with data analytics.

Some of the participants noted the need for a collaboration of the DDI processes and their tools with the designers, rather than digitising the design processes completely. More specifically, one of the respondents noted:

"*Digital design is ultimately part of a larger transition in terms of how we undertake projects using such tools alongside our foundation of knowledge, culture and values. Data-driven processes will become more embedded in our work in the near future and will be a matter for the designer or architect to decide how to utilise this information and which tools to use..."* - Participant Non-experienced in DDI

In addition, some of the participants felt that digital tools and their processes are just a way to move towards data-driven and evidence-based approaches, rather than allowing the tools to make the decision-making for the designers. Finally, one contributing comment towards this direction from a Non-experienced in DDI participant noted:

“*I believe that good designers will instinctively make good design decisions and that digital design tools and data analysis can be used to evidence these ideas and refine them to produce better outcomes”.* – Participant Non-experienced in DDI

3.6 Discussion

Qualitative information was obtained from the stakeholders via an online questionnaire survey. This information assisted in understanding and evaluating their beliefs and attitudes when a decision is to be made in the building and urban design. The factors identified are strongly correlated, suggesting that the variables are significantly associated with each other.

Decision-making in design relates to how people interpret and combine information to arrive at specific solutions (Bruch & Feinberg, 2017). The findings imply that the decision-maker's role within a team does not heavily influence decision-making processes in design. However, more significant differences were found in the decision-makers that form part of a different discipline. Results revealed that although DMPs present some differences between the disciplines of A&MP and Engineers, the most significant difference is observed between the Project Management and A&MP disciplines. This finding reflects the difference in the end-goals between the two disciplines and the perception of what successful design of a project entails. This can heavily influence the choices made along the design process. For example, the delivery of the project on time and budget, reflecting aspects of the role of the Project Management discipline versus the design outcome from an architectural perspective, reflecting the key goals of the A&MP discipline.

Comparison of the ARII between experienced (EXP) and non-experienced (Non-EXP) DDI users illustrates the differences in stakeholders' views. One of the key observations is the value for the *Instinctiveness* factor, where, although it is generally considered as of high-medium importance, for experienced DDI users, this factor has a score of ARII= 0.013. This finding indicates the lower level of reliance of the experienced users on their analysis outcomes and evidence-based methodologies. Experienced DDI users rely on instincts to decide, linking these findings to the concept of heuristics (Shah & Oppenheimer, 2008). In addition, although the impact of the *Hesitancy* and *Experience* factors is generally considered of having a low influence within the rest of the stakeholder categories, for EXP, these factors ranked in the third and fourth place, respectively. These results imply that EXP decision-makers are conservative against their decision-making process, while Non-EXP are open to exploring and altering their initial favourable option.

Social pressures have been recognised as one of the factors influencing decision-making processes (Pereira, et al., 2016). The results revealed that although the stakeholders generally value the *Social Resistance* and *Principled* factors, for EXP, the ARII values are low. These findings illustrate the differences in the way of thinking between EXP and Non-EXP, revealing that experienced DDI users are not heavily influenced by their decision environment. Instead, they rely heavily on information to make a design decision. Non-experienced users, though, value other people’s opinions. As the confidence of decision is one of the risks involved in the decision-making process (Finucane, et al., 2005), findings indicate the preference of working in a collaborative environment, where decisions can arise from workshops and discussions, minimising the associated risks.

Overall, findings revealed that the decision-making processes in buildings and urban design vary significantly among the decision-makers. Results indicated that their individual characteristics heavily influence the process, and subjectivity is essential in decision-making. Multiple other influencing parameters exist in the decision-making processes as revealed from literature, including organisational, environmental, and managerial factors (Rajagopalan, et al., 1997). However, these were not specifically examined in this study. Findings are limited to the decision-makers beliefs and attitudes when a decision in buildings and urban design is to be made.

3.7 Conclusion

This study aims to investigate, quantify, and rank the relative importance of the decision-making factors contributing to the design of buildings and urban projects. A survey was conducted to gain an insight into stakeholders' perceptions as to which are the influencing factors affecting decision-making processes in the design of buildings and places. Exploratory Factor Analysis, Average Relative Importance Index, and Spearman's Rank Correlation Coefficient analysis identified the key factors and their relative importance influencing DMPs in design. This study provides a new means to evaluate the performance of decision-making processes when these are undertaken by developing and applying a quantitative data-driven, evidence-based methodological framework. The current analysis built an applicable framework that evaluated the importance of 30 variables in total, quantifying the way designers make decisions and capturing the social aspects introduced by the technological advancements and increasing data availability and their use in the design process to be evaluated and optimized.

Four highly important and distinct factors were generated: *Potential for Dynamic Operation, Recency of tools, Thoroughness and Control.* This study revealed that DMPs vary significantly among the decision-makers, which is heavily influenced by their individual characteristics. Hence, there is an urgent need to identify how this process is undertaken each time, leading to optimised decisions. In addition, it has been revealed that although DDI processes are generally received as a positive addition to the DMPs from the individuals, it is not yet achievable due to key practical and cultural barriers identified. These include the skills needed to employ DDI processes and the diversity in experiences and roles within the decision-making process. The current analysis can help practitioners assign the most fitting roles to stakeholders within the design process. Additionally, this model can also be utilised to encourage organisations in the building design industry to improve their approaches and final decision outcomes.

Implications/ limitations of this study include that EFA conclusions are based on post hoc analysis, therefore being subject to possible errors. In addition, respondents possessed inadequate experience with DDI; hence variables may not have been properly evaluated. Another limitation is that the sample size includes respondents worldwide. However, DMPs in urban and building design may differ due to present regulations or lack of available resources (i.e., data and tools).

Future research could be performed to validate results on a case-study-based approach and collection of in-depth qualitative data to understand better aspects of DMPs that were not included in this research. In addition, a focused country-based sample size selection could reveal further insights.

REFERENCES

Alkarkhi, A. & Alqaraghuli, W., 2019. Chapter 9 - Factor Analysis. In: *Easy Statistics for Food Science with R.* Malaysia: Academic Press.

Almanasreh, E., Moles, R. & Chen, T., 2019. Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy,* 15(2), pp. 214-221.

Almeida, S., Resende, T. & Dieter Stobäus, C., 2016. Validity, Reliability and Convergent Analysis of Brazilian Version of Selection, Optimization and Compensation Questionnaire (QSOC). *Creative Education,* 7(15), pp. 2074-2087.

AlSehaimi, A., Koskela, L. & Tzortzopoulos, P., 2013. Need for Alternative Research Approaches in Construction Management: Case of Delay Studies. *Journal of Management in Engineering,* pp. 407-413.

Andrienko, G. & Andrienko, N., 2008. *Spatio-temporal aggregation for visual analysis of movements.* Columbus, OH, USA, IEEE, pp. 51-58.

Böhme, G. & Stehr, N., 1986. *The Knowledge Society. The Growing Impact of Scientific Knowledge on Social Relations.* Dordrecht: Springer.

Brodaty, H., Mothakunnel, A., de Vel-Palumbo, M., Reppermund, S., Kocha, N.A., Savage, G., Trollor, J.N., Crawford, J. & Sachdev, P.S., 2014. Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging. *Annals of epidemiology,* Volume 24, pp. 63-71.

Bruch, E. & Feinberg, F., 2017. Decision-Making Processes in Social Contexts.. *Annu Rev Sociol.,* Volume 43, pp. 207-227.

Carroll, A. B., 1987. In search of the moral manager. *Business Horizons,* 30(2), pp. 7-15.

Castelli, N., de Carvalho, A.F.P., Vitt, N., Taugerbeck, S., Randall, D., Tolmie, P., Stevens, G. & Wulf, V., 2020. On technology-assisted energy saving: challenges of digital plumbing in industrial settings. *Human–Computer Interaction,* 0(0), pp. 1-29.

Chan, D. & Kumaraswamy, M., 1997. A comparative study of causes of time overruns in Hong Kong construction projects. *International Journal of Project Management,* 15(1), pp. 55-63.

Chan, K. et al., 2016. Why protect nature? Rethinking values and the environment. *PNAS,* 113(6), pp. 1462-1465.

Chen, A., McGinnis, B. & Ullman, D., 1990. *Design history knowledge representation.* Chicago, American Society of Mechanical Engineers, p. 175–184.

Chen, E., Okudan, G. & Riley, D., 2010. Sustainable performance criteria for construction method selection in concrete building. *Automation in Construction,* 19(2), pp. 235-244.

Chen, S.-F., Wang, S. & Chen, C.-Y., 2012. A simulation study using EFA and CFA programs based the impact of missing data on test dimensionality. *Expert Systems with Applications,* 39(24), pp. 4026-4031.

Child, D., 2006. *The essentials of factor analysis.* New York: Continuum International Publishing.

Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences.* New York: Lawrence Erlbaum Associates.

Cooper, N., Bassett, D. & Falk, E., 2017. Coherent activity between brain regions that code for value is linked to the malleability of human behavior. *Scientific Reports,* Volume 7, Article: 43250.

Costello, A. & Osborne, J., 2005. Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment,* 10(7), pp. 1-9.

Cristobal, E., Flavián, C. & Guinalíu, M., 2017. Perceived e-service quality (PeSQ): Measurement validation and effects on consumer satisfaction and web site loyalty. *Managing Service Quality: An International Journal,* 17(3), pp. 317-340.

Davis, F., Bagozzi, R. & Warshaw, P., 1992. Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. *Journal of Applied Social Psychology,* 22(14).

De Bruin, W., Parker, A. & Fischhoff, B., 2007. Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology,* 92(5), pp. 938-956.

Delmonico, D., Jabbour, C.J.C., Pereira, S.C.F., de Sousa Jabbour, B.I., Renwick, D.W.S. & Thomé, A.M.T., 2018. Unveiling barriers to sustainable public procurement in emerging economies: Evidence from a leading sustainable supply chain initiative in Latin America. *Resources, Conservation and Recycling,* Volume 134, pp. 70-79.

Denton, H., 1997. Multidisciplinary team-based project work: planning factors. *Design Studies,* 18(2), pp. 155-170.

Dong, Y. & Peng, C.-Y., 2013. Principled missing data methods for researchers. *Springerplus,* 222(2).

Dutilh, G., Annis, J. & Brown, S., 2019. The Quality of Response Time Data Inference: A Blinded, Collaborative Assessment of the Validity of Cognitive Models. *Psychonomic Bulletin & Review,* Volume 26, p. 1051–1069.

Elgendy, N. & Elragal, A., 2016. Big Data Analytics in Support of the Decision Making Process. *Procedia Computer Science,* Volume 100, pp. 1071-1084.

Enquist, C., Jackson, S.T., Garfin, G.M., Davis, F.W., Gerber, L.R., Littell, J.A., Tank, J.L., Terando, A.J., Wall, T.U., Halpern, B., Stephenson, N.L. & Willi, M.A., 2017. Foundations of translational ecology. *Frontiers in Ecology and the Environment,* 15(10), pp. 541-550.

Ernest, A., 1982. Design in the Decision-Making Process. *Policy Sciences,* 14(3), pp. 279-92.

Ferreira, N., Lage, M., Doraiswamy, H., Vo, H., Wilson, L., Werner, H., Park, M. & Silva, C., 2015. Urbane: A 3D framework to support data driven decision making in urban development. In: *IEEE Conference on Visual Analytics Science and Technology (VAST).* Chicago, IL, USA: IEEE, pp. 97-104.

Field, A., 2009. *Discovering Statistics Using SPSS: (and Sex, Drugs and Rock'n'roll).* Los Angeles: SAGE Publication.

Finucane, M., Mertz, C., Slovic, P. & Schmidt, E., 2005. Task complexity and older adults' decision-making competence.. *Psychology and Aging,* 20(1), pp. 71-84.

Gal, Y. & Pfeffer, A., 2008. Networks of Influence Diagrams: A Formalism forRepresenting Agents' Beliefs and Decision-Making Processes. *Journal of Artificial Intelligence Research,* Volume 33, pp. 109-147.

Gandomi, A. & Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management,* 35(2), pp. 137-144.

Ghattas, J., Soffer, P. & Peleg, M., 2014. Improving business process decision making based on past experience. *Decision Support Systems,* Volume 59, pp. 93-107.

Ghosh, S. & Jintanapakanont, J., 2004. Identifying and assessing the critical risk factors in an underground rail project in Thailand: a factor analysis approach. *International Journal of Project Management,* 22(8), pp. 633-643.

Gigerenzer, G. & Goldstein, D., 2002. Models of ecological rationality: The recognition heuristic. *Psychological Review,* 109(1), pp. 75- 90.

Hall, D. & Davis, R., 2007. Engaging multiple perspectives: A value-based decision-making model. *Decision Support Systems,* 43(4), p. 1588–1604.

Hansen, C. & Andreasen, M., 2004. *A mapping of decision-making.* Dubrovnik, Design society, pp. 1409-1418.

Harris, R., 2012. *Introduction to decision making.* [Online] Available at: https://www.virtualsalt.com/crebook5.htm

Hassenzahl, M., Diefenbach, S. & Goritz, A., 2010. Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers,* 22(5), p. 353–362.

Hilbert, M., 2012. Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making.. *Psychological Bulletin,* 138(2), p. 211–237.

Hilbig, B. & Pohl, R., 2008. Recognition users of the recognition heuristic. *Experimental Psychology,* 55(6), pp. 394-401.

Hinkin, T., 1998. A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Method,* Volume 1, p. 104–121.

Hofmann, M., Westermann, J., Kowarik, I. & Van der Meer, E., 2012. Perceptions of parks and urban derelict land by landscape planners and residents. *Urban Forestry & Urban Greening,* 11(3), pp. 303-312.

IBM Corp., 2019. *IBM SPSS Statistics for Windows, Version 26.0.* NY: Armonk, NY: IBM Corp.

Janssen, M., van der Voort, H. & Wahyudi, A., 2017. Factors influencing big data decision-making quality. *Journal of Business Research,* Volume 70, pp. 338-345.

Jarkas, A. & Younes, J., 2012. Principle factors contributing to construction delays in the state of Qatar. *International Journal of Construction Project Management,* 6(1), pp. 39-62.

Karekla, M. & Michaelides, M., 2017. Validation and invariance testing of the Greek adaptation of the Acceptance and Action Questionnaire-II across clinical vs. nonclinical samples and sexes. *Journal of Contextual Behavioral Science,* 6(1), pp. 119 -124.

Kensing, F., Simonsen, J. & Bodker, K., 1998. MUST: A Method for Participatory Design. *Human–Computer Interaction,* 13(2), pp. 167-198.

Kim, J. & Ryu, H., 2014. A Design Thinking Rationality Framework: Framing and Solving Design Problems in Early Concept Generation. *Human–Computer Interaction,* 29(5.6), pp. 516-553.

Kirkman, G., Cornelius, P., Sachs, J. & Schwab, K., 2002. *The Global Information Technology Report 2001–2002: Readiness for the Networked World,* New York & Oxford: Oxford University Press.

Kirsi, A., 2011. Project stakeholder analysis as an environmental interpretation process. *International Journal of Project Management,* 29(2), pp. 165-183.

Kuddus, M., Tynan, E. & McBryde, E., 2020. Urbanization: a problem for the rich and the poor?. *Public Health Reviews,* Volume 41, Article 1.

Lahdelma, R. & Salminen, P., 2000. SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making. *Operations Research,* 49(3), pp. 325-468.

Lamberton, D., 1994. *Knowledge Societies by Nico Stehr.* London: Sage Publications.

Little, R. & Rubin, D., 1987. *Statistical Analysis with Missing Data.* New York: John Wiley & Sons.

Lytras, M. & Sicilia, M., 2005. The Knowledge Society: a manifesto for knowledge and learning. *International Journal of Knowledge and Learning,* 1(1-2), pp. 1-11.

Madu, C. & Georgantzas, N., 1991. Strategies thurst of manufacturing decisions: a conceptual framework. *IIE Transactions,* 23(2), pp. 138- 148.

Malone, T. W., 1981. *What Makes Things Fun to Learn? A Study of Intrinsically Motivating Computer Games.* Palo Alto, California,, Palo Alto Research Center, Cognitive and Instructional Sciences Group, pp. 162-169.

Nunnally, J. & Bernstein, I., 1994. The Assessment of Reliability. *Psychometric Theory,* Volume 3, pp. 248-292.

OECD, 2015. Data-driven innovation: Big data for growth and well-being. In: *OECD digital economy papers.* Paris: OECD, p. 456.

Oller, D., 2014. Exploratory factor analysis as a tool for investigating complex relationships: when numbers are preferred over descriptions and opinions. *SAGE Research Methods Cases.*

Olomolaiye, P., Wahab, K. & Price, A., 1987. Problems influencing craftsmen's productivity in Nigeria. *Building and Environment,* 22(4), p. 317–323.

Papadakis, V., Lioukas, S. & Chambers, D., 1998. Strategic Decision-Making Processes: The Role of Management and Context. *Strategic Management Journal,* Volume 19, pp. 115 -147.

Pereira, G., Prada, R. & Santos, P., 2016. Integrating social power into the decision-making of cognitive agents. *Artificial Intelligence,* Volume 241, pp. 1-44.

Pettigrew, A., 1990. Longitudinal Field Research on Change: Theory and Practice. *Organization Science,* 1(3), pp. 213-337.

Piedmont, R., 2014. Inter-item Correlations. In: *Encyclopedia of Quality of Life and Well-Being Research.* Dordrecht: Springer.

Ponto, J., 2015. Understanding and Evaluating Survey Research. *Journal of the Advanced Practitioner in Oncology,* 6(2), pp. 168-171.

Prasad, D., Pradhan, R.P., Gaurav, K., Kaur, I., Dash, S. & Nayak, S., 2018. Analysing the critical success factors for implementation of sustainable supply chain management: an Indian case study. *Decision,* Volume 45, p. 3–25.

Qui, L., Lindberg, S. & Nielsen, A., 2013. Is biodiversity attractive?—On-site perception of recreational and biodiversity values in urban green space. *Landscape and Urban Planning,* Volume 119, pp. 136-146.

Quinn, P. & Cockburn, A., 2020. Loss Aversion and Preferences in Interaction. *Human–Computer Interaction,* 35(2), pp. 143-190.

Rajagopalan, B., Lall, U. & Cane, M., 1997. Anomalous ENSO Occurrences: An Alternate View. *Journal of Climate,* 10(9), p. 2351–2357.

Sachdev, S. & Verma, H., 2004. Relative importance of service quality dimensions: A multisectoral study. *Journal of services research,* 4(1), pp. 0-116.

Santagata, R. & Yeh, C., 2016. The role of perception, interpretation, and decision making in the development of beginning teachers' competence. *ZDM Mathematics Education,* Volume 48, p. 153–165.

Shah, A. & Oppenheimer, D., 2008. Heuristics made easy: An effort-reduction framework.. *Psychological Bulletin,* 134(2), pp. 207-222.

Shash, A., 1993. Factors considered in tendering decisions by top UK contractors. *Construction Management and Economics,* 11(2), pp. 111-118.

Simonet, A., Fedak, G. & Ripeanu, M., 2015. Active Data: A programming model to manage data life cycle across heterogeneous systems and infrastructures. *Future Generation Computer Systems,* Volume 53, pp. 25-42.

Snyder, J., Baumer, E.P.S., Voida, S., Adams, P., Halpern, M., Choudhury, T. & Gay, G., 2014. Making Things Visible: Opportunities and Tensions in Visual Approaches for Design Research and Practice. *Human–Computer Interaction,* 29(5-6), pp. 451-486.

Stanitsas, M. & Kirytopoulos, K., 2021. Investigating the significance of sustainability indicators for promoting sustainable construction project management. *International Journal of Construction Management,* 0(0), pp. 1-26.

Stanitsas, M. & Kirytopoulos, K., 2022. Underlying factors for successful project management to construct sustainable built assets. *Built Environment Project and Asset Management,* 12(2), pp. 129-146.

Stauffer, R. & Ullman, D., 1991. Fundamental processes of mechanical designers based on empirical data. *Journal of Engineering Design,* Volume 2, p. 113–126.

Tempesta, T. & Vecchiato, D., 2015. Public preferences for electricity contracts including renewable energy: A marketing analysis with choice experiments. *Energy,* Volume 88, pp. 168-179.

Ullman, D. & Paasch, R., 1994. *Issues critical to the development of design history, design rationale and design intent systems.* Minneapolis, The American Society of Mechanical Engineers, p. 249–258.

United Nations, 2019. *World urbanization prospects,* New York: United Nations.

Venkatesh, V., Morris, M., Davis, G. & Davis, F., 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly,* 27(3), pp. 425-478.

Vishwakarma, G., 2017. Sample Size and Power Calculation. In: *Nursing Research in 21st Century: Research Methodology.* New Delhi, India: CBS Publishers & Distributors, pp. 1-21.

Wang, H., Xu, Z., Fujita, H. & Liu, S., 2016. Towards felicitous decision making: An overview on challenges and trends of Big Data. *Information Sciences,* Volume 367–368, pp. 747-765.

Waris, M., Shahir Liew, M., Khamidi, M. & Idrus, A., 2014. Criteria for the selection of sustainable onsite construction equipment. *International Journal of Sustainable Built Environment,* 3(1), pp. 96-110.

West, R., Toplak, M. & Stanovich, K., 2008. Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology,* 100(4), pp. 930-941.

Wiese, C., Shuffler, M. & Salas, E., 2015. Teamwork and Team Performance Measurement. In: *International Encyclopedia of the Social & Behavioral Sciences (Second Edition).* Oxford: Elsevier.

Yi, X., Liu, F., Liu, J. & Jin, H., 2014. Building a network highway for big data: architecture and challenges. *IEEE Network,* 28(4), pp. 5-13.

Zillner, S., 2021. Innovation in Times of Big Data and AI: Introducing the Data-Driven Innovation (DDI) Framework. In: *Curry E., Metzger A., Zillner S., Pazzaglia JC., García Robles A. (eds) The Elements of Big Data Value.* Munich, Germany: Springer, p. 289–310.

4 Challenges and applications of Big Data Approaches in the context of Urban Informatics

4.1 Abstract

Big Data can provide valuable insights and enhance the decision-making process in urban design. Urban Informatics approaches have the capacity to utilise Big Data Approaches for retrofitting and renovating urban design. Nevertheless, their practical application in the urban planning industry is increasingly complex, and often, limited. The aim of this chapter is to synthesize and present a state-of-the-art structured analysis of Big Data Approaches in Urban Informatics, identifying the challenges that arise through their use. To address this, a two-step approach is adopted, being firstly a literature review conducted focussing on themes relating to the different forms of Big Data challenges and analytical methods employed. Secondly, an online survey questionnaire is used to investigate stakeholder perceptions of the major challenges and potential of Big Data Approaches in urban research and practice. A critical commentary is provided concerning the educational, technical, and behavioural barriers that limit their potential in urban design, planning and management. The study concludes by identifying the required skillset for practitioners seeking effective implementation of Big Data strategies in Urban informatics.

4.2 Introduction

Big Data (BD) is recognised as comprising a new generation of tools and techniques designed to extract value from large volumes and varieties of data, enabling capturing, analysis, and knowledge-discovery in high-velocities (Esteves & Curto, 2013). Big Data studies have highlighted several “Vs” to define the key characteristics (Brock & Khan, 2017; Comuzzi & Patel, 2016). These are described initially as Volume, Velocity, and Variety but often extended to include a fourth V represented by Veracity (Laney, 2001). Since then, a range of diverse studies has been adding to these Vs, referring to Value, Validity, Variability/Volatility, Virtual, Visualization/Visibility, adding to the complexity of the Big Data definition (Demchenko, 2014; Saxena, 2016; Drus & Hassan, 2017). For this study, Big Data will be defined as the data generated by activities, including transactional, operational, design and planning, social and others, or post-processed designed data.

Urban informatics, in combination with Big Data Approaches (BDAs), stimulated the interest of decision-makers in practice and academics, seeking solutions in planning and governance-related issues (e.g., (Hu & Han, 2019; Zhao, et al., 2018)). Multiple terms exist for the definition of the urban informatics field. According to Foth et al. (2011, p. 4), "*Urban informatics is the study, design, and practice of urban experiences across different urban contexts that are created by new opportunities of real-time, ubiquitous technology and the augmentation that mediates the physical and digital layers of people networks and urban infrastructures*". Additionally, one of the most prominent terms for urban informatics is 'urban computing', which focuses on combining and aligning both computer science and human-computer interaction design and research. Urban informatics, as defined in this study, relates to those urban systems utilising novel tools and technologies to collect information related to the physical and digital layers of humans and urban spaces. Therefore, urban informatics research approaches refer to a theory-driven perspective in the context of data-driven real-life scenarios. BDAs refer to the combination of diverse datasets and related technologies, able to extract insights concerning complex systems through novel organisational and analytical capabilities (Pollard, et al., 2018). Although their research potential in addressing urban design and planning issues is promising, technical and knowledge discovery challenges slow down their practical implementation and application in academia.

This paper synthesises and presents a state-of-the-art structured analysis to review areas of Big Data Approaches BDAs' application in Urban Informatics, identifying the challenges that arise through their use and identifying key challenges in their use. A two-step approach is followed, firstly through an investigation of previous literature, using thematic analysis that identifies and reports the different types of Big Data challenges and the analytical methods employed. Secondly, an online survey questionnaire was employed to investigate stakeholders' perceptions in relation to significant challenges and the potential of BDAs implementation in research and practice. This study contributes to the existing literature in two ways: by providing a review of the challenges that limit the potential of BDAs in the urban design and planning context and by providing the skillset requirements for practitioners that seek effective implementation of BDAs in Urban informatics. Key recipients of the findings will be urban planners, decision-makers, and academics interested in delivering urban environments aligned to the end-user needs, utilising novel technologies.

4.3 Types and key challenges in urban Big Data processing

The use of BD generates a range of technological and methodological challenges associated mainly with the types of data utilised. Although there are several ways of grouping this plethora of data, this study adopts the categorisation of data types described by Thakuriah et al. (2017) and Pan et al. (2016), reflecting the user communities (UC) linked to the data generation rather than solely their technical characteristics (Table 4-1- Types). Therefore, it connects the data types directly to urban design and planning research and applications. An overview of the data types is presented in Table 4-1, along with examples of their practical use and the users they refer to.

Table 4-1: Description of main types of Urban Big Data, highlighting examples of data and information, sources, user communities (UC) and updating speed. Extended and adapted from Thakuriah et al. (2017, p. 7)and Pan et al. (2016). Updating rates are extended from Paganin, et al. (2018).

Type	Examples	Sources	User Community type	Updating rate
Sensor systems	Sensor systems for managing the built environment: water consumption, traffic flow count, energy, and buildings' performance, mobile phone, monitoring camera, Internet of Things	Connected Systems of sensors (e.g. WSN) and other IoT Networks: (mobile and fixed) device networks and sensor networks in urban areas (e.g., public utility control units, environmental parameters detection networks, smart grids, etc.), Public utilities sensor systems, Building Management Systems (BMS), Smart Grids, Surveillance System, Geographic Information Systems (GIS), Satellite Earth observation service (Earth Observation Satellites System & Land Observations System)	**UC1:** Public and private urban operations and management organizations, independent ICT developers, researchers in the engineering sciences	High
User-generated content (UGC)	Participatory sensing system, social media, personal tracking devices, such as mobile phones and smart watches, network use, global positioning system (GPS)	Participatory sensing systems, social media, access, and logins; Global Positioning System (GPS), Global Navigation satellite systems (GNSS), online social network, mobile application, Blogging & web 2.0	**UC2:** Private businesses, customer/client focused public organizations, independent developers, researchers in data sciences and urban social sciences.	High

Administrative data	Registration, transactions, tax information, revenues, payment, population data, traffic flows, land use, housing, employment, medical information, education	Administrative portals, regional and municipal portals of the Government (open and confidential microdata on population)	**UC3:** Open data innovators, civic hackers, researchers, government data agencies, urban social scientists involved in economic and social policy research, public health, medical researchers.	Low
Private sector transaction data	Transaction data, customer data, product purchase and service agreements	Records, memory cards and company documents; workforce management system; data on public services and financial institutions, product purchase register and terms of service agreements	**UC4:** Private businesses, public agencies, independent developers, researchers in data sciences and urban social sciences	Low
Humanities data	Texts, images, media repositories	Cultural heritage data (manuscripts, texts, artefacts), Open governmental data, research data, digital humanities	**UC5:** Urban design community, historical, art, architecture, and digital humanities organizations; community organizations, data scientists and developers	High
Hybrid data	Post-processing data, metadata	Research data made available by others	**UC6:** Urban social scientists involved in economic and social policy research; public health; medical researchers	High

The first BD category noted is sensor systems. The advancement of network technologies (e.g., internet, wireless communication, mobile, locative media, etc.) enabled the installation of sensor systems in urban infrastructure, resulting in vast amounts of data being generated. Such data are characterised by both data in motion and at rest, as they can detect activities and changes throughout a specified period (Grommé, 2016). New insights on patterns related to demand and usage of urban infrastructure systems can be extracted utilising such systems, bringing value to a range of stakeholders, from private to public organisations, researchers, and practitioners involved in the engineering and design sciences. A key challenge for sensor data is the storage capacity and data management practices adopted as sensor systems capture significant volumes of data in real-time and various types, incorporating substantial amounts of sensitive data, such as location and movement data.

User-generated content (UGC) is the second BD type analysed. Sensor systems have been evolved and can now be found recording personal information from individuals, resulting from their participation in social media or personal tracking devices, such as mobile phones and smart watches (Resch, et al., 2020). Such data are generated proactively when users voluntarily generate their data by their social or civic activities. Like the sensor data, large volumes of information are created every second, resulting in challenges around their storage and management. However, such data also raise questions regarding biases and lack of generalisability due to the expression of personal views or the limited representation of the whole population, restricted to those with access to such technologies. As an example, groups of people who do not own smartphones or smart watches, running the risk of perpetuating discriminations of access (e.g., considering the needs of only privileged groups).

Administrative data is recognised as another type of BD in literature (Thakuriah, et al., 2017; Pan, et al., 2016). Detailed information on citizens' everyday activities is being collected by various governmental bodies, such as registration, transactions, tax information, and others related to urban policy evaluation. Such data are often beneficial in their use as they act as cheap sources of information, are gathered less intrusively than other types of data (e.g., UGC), and are easy to comprehend, while they include larger population samples with greater participation, compared to traditional survey data (Gowans, et al., 2012). One of the most significant advantages of administrative data is that their accessibility is constantly increasing, as they have been incorporated as part of "Open data" initiatives, with the aim to reflect transparent and innovative governments and create accessible "knowledge" and citizenship creativity for all (Open Data Institute, 2018; Ubaldi & OECD, 2013). Over the last decade, open data initiatives focused on creating and maintaining open data portals with the aim of assisting a variety of actors to locate the data they need to create insights. More than 2,600 open data portals exist worldwide, creating ecosystems of information (OpenDataSoft, 2018). Open data efforts initially were focused on providing raw data, which then was followed by open file format standards, such as CSVs in place of PDFs. Since then, open standards for data have been implemented, allowing users to access and share better quality data (Open Data Institute, 2018).

Nevertheless, not all data within this category are a part of these initiatives as sensitive information is associated in some respects. Therefore, restrictions on their accessibility may apply. However, such detailed information is of great interest to many urban researchers and practitioners, as it can allow them to create linkages that will support an in-depth exploration of spatial and temporal variations. Some of the key challenges associated with administrative data involve data security requirements, quality, and privacy legislation, as they may be linked to closed government cultures (Huijboom & van den Broek, 2011).

Similar to the administrative data category, private sector transaction data is a type of BD that originates from businesses that collect their own data as part of their daily interactions with their clients. Such data are collected and monitored from privately owned sensor systems, which continuously track their clients' activities and use patterns (Thakuriah, et al., 2017). Companies utilise such information to improve their services and business processes. This data is of great interest to urban researchers as they can improve understanding of behavioural implications or tackle contemporary challenges cities face, such as fuel poverty. Nevertheless, these datasets are not always accessible.

Humanities data as a type of BD refers to cities and human activities that produce vast amounts of data at an unparallel pace (Shi & Abdel-Aty, 2015; Reddy, et al., 2020). Nevertheless, activities within a city may not relate only to undertaking acts, such as walking or shopping, but can also represent creativity and cultural identities found in various sources of unstructured information, such as texts, images, media repositories, and others. This information describes the social, cultural, and physical dynamics of a city, which are of interest to a range of cultural, heritage, and digital organisations, such as galleries, museums, and other institutions (Thakuriah, et al., 2017). There are several challenges associated with these types of data, mainly due to the increased variety of information which is not structured and is being created not in an organised and temporal manner. Therefore, although such information can shed light on a range of qualitative attributes within quantitative urban modelling approaches, their processing involves diverse technical and methodological difficulties and limitations, such as specialised analytical software knowledge or interpretation capabilities.

The final type of BD is hybrid data, which results from post-processing techniques, including a range of linked information and data types required to support specific outcomes. Such connections can be proven computationally challenging, but one of its key limitations lies in acquiring such data to be shared by their creators. In addition, as post-processed data, they involve many decisions made throughout the process; hence their quality can be questioned.

Big Data is closely linked to their promise of providing novel insights that will change how we understand and organise societies. One of the processing challenges of Big Data is how these data should be captured, analysed, and interpreted to extract value from the results (Sivarajah, et al., 2017). Data management, storage, and analytics are only a part of data-driven operations and planning. Much of the effort should be concentrated on the critical interpretation of the results and their impact on society.

4.4 Urban Informatics: background and trends

Urban informatics is "*a disciplinary domain situated at the intersection of notions, trends, and considerations for place, technology, and people in urban environments*" (Foth, et al., 2011, p. 2). Urban planning involves multiple disciplines, and their interests shift due to changing perspectives in planning policies, architecture, social and environmental conditions, and governance (Sanchez, 2020). Professionals deal with great challenges that evolve through time, while researchers argue that urban planning is an "action-oriented" field, indicating the necessity of data-driven insights on the urban dynamics (Loh, 2017). The increasing interest of the research community demonstrates the uses of BD in an urban context (Kontokosta, et al., 2018; Thakuriah, et al., 2017). These studies focus on using novel sources of data and computing methods to progress operational decisions in city management. Much of the literature, however, fails to recognise that planning practice and theory has struggled with how best to integrate novel technologies and quantitative methods into the decision process (Batty, 2014; Krizek, et al., 2009; Kontokosta, 2018). Awareness of the need for new models and types of data (e.g., behavioural information) represents a radical change from domain-specific models and current processes (Kontokosta, 2018).

4.5 Materials and methods

4.5.1 Analysis and diagnosis: Thematic analysis

This study explores areas of BDAs' application in the Urban informatics context following a thematic analysis (Braun & Clarke, 2012). A review was conducted of those studies focusing on the urban planning research themes and the role of novel technologies data collection and analysis. An identification of relevant literature was undertaken through the search of key variables within the categories "title, abstract, and keywords" online database (Scopus) in December 2020. Only articles published in English were included and the keywords used included combination of terms "*spatial analysis*", "*urban planning*" and "*big data*". The thematic analysis is presented and discussed below in the five steps as shown in the analytical framework (Figure 4-1).

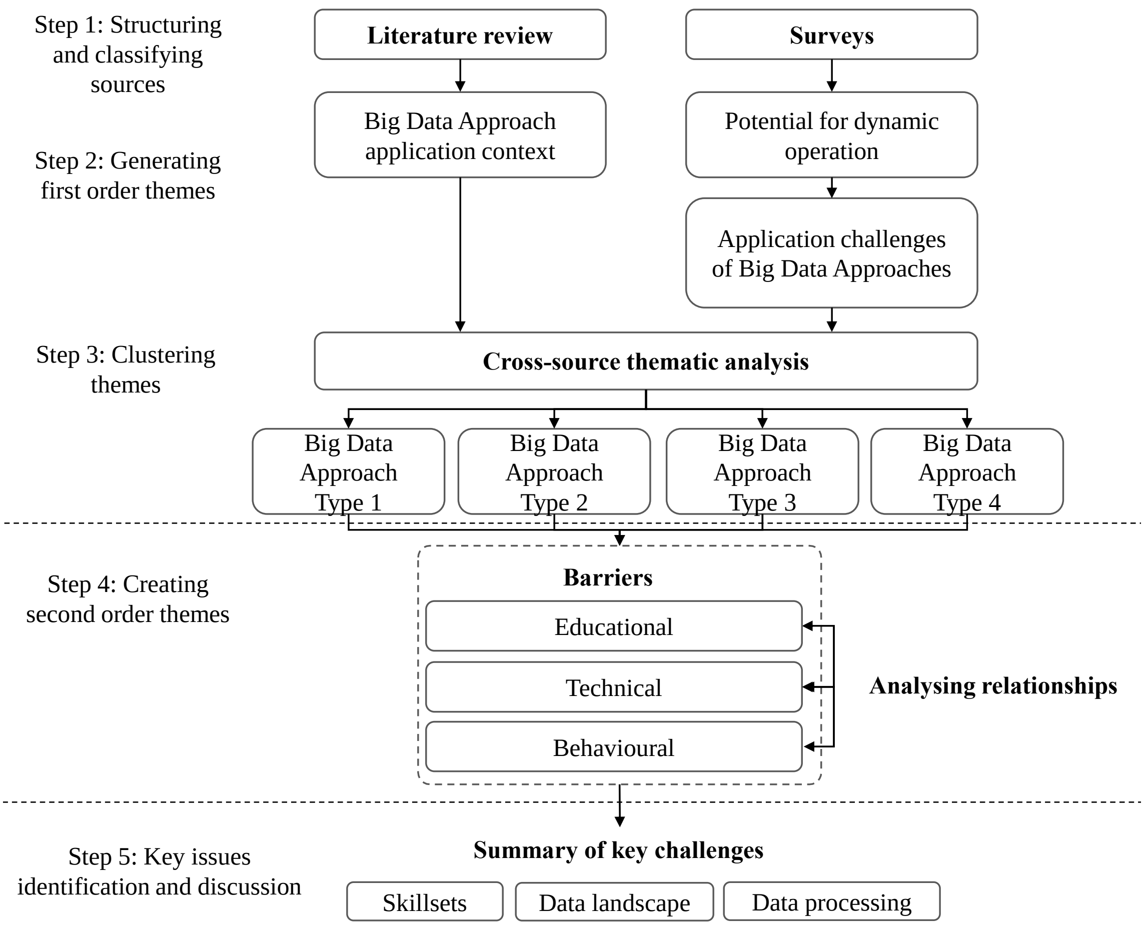


Figure 4-1: Descriptive analytical framework.

4.5.2 Empirical data collection via questionnaires

Following Chapter 3, focused on identifying the factors influencing decision-making processes (DMPs), a survey was conducted with the aim of exploring stakeholders' perceptions related to DMPs for the design of the built environment and its surroundings. The questionnaire survey method was selected as suitable due to its ethical advantages, as survey is a less intrusive research method without exposing the individuals to invasive techniques (Fox, et al., 2000). The questionnaire was formed in four distinct parts; Part 1: relating to the respondents' background; Part 2: including evaluating questions concerning decision-making quality; Part 3: looking at recency of tools to improve quality and Part 4: exploring views of novel technologies and their potential for dynamic operation.

Respondent backgrounds consist of different levels of roles, hierarchy within companies' structure and levels of experience in the use of data-driven innovation techniques (DDI). The responses were recorded using a five-point Likert scale, except for Parts 3 and 4, which included a multiple-choice scale (Shah & Oppenheimer, 2008; Gigerenzer & Goldstein, 2002; Hilbig & Pohl, 2008). The targeted sample of participants was identified via LinkedIn accounts and consultancy organisations with direct approach. The period conducted was a four-month period, starting December 2019. Some 184 responses were obtained in total, of which 136 ones had usable responses, reflecting a confidence level of 95% of the total population approached with 8% margin error.

For the purposes of this study, key questions from the questionnaire were extracted, based on relevance of levels of experience, tools and techniques and the area of research potential (Table 4-2), aiming to investigate stakeholders' perceptions against major challenges and potential of the Big Data implementation in practice.

Table 4-2: Questions used to investigate stakeholders' perceptions of the major challenges and potential of the Big Data implementation in practice.

Variable No.	**Description**
Q0**1**	To what extent do you have experience with big data (BD) and data-driven innovation (DDI)?
Q0**2**	To what extent do you think data-driven decision-making is part of the organisation's culture?
Q0**3**	To what extent do you think the use of data in the design process would benefit your organisation?
Q0**4**	To what extent do you think new technologies, data tools, modelling and visualisation can improve decision-making processes resulting in better designs?
Q0**5**	To what extent do you think monitoring, integrating sensors, citizen science, public engagement and inclusion can improve decision-making processes resulting in better designs?
Q0**6**	To what extend do you think infrastructure, data accessibility and usability for stakeholders can improve decision-making processes resulting in better designs?
Q0**7**	What would you feel about a completely digital design process for the built environment and its surroundings?
Q0**8**	What would you feel about a completely digital design process combined with data analytics approaches for the built environment and its surroundings?

The questions were focused on understanding stakeholder perceptions regarding the potential for dynamic operation. Potential for dynamic operation refers to the dynamic dimension of a design model to change and adapt over time. The questions reviewed the diverse challenges of BDAs and DDI, and their relation to the design process, investigating perceptions of using digitalised decision-making processes in design and a combination of digital processes and data analytics.

On conclusion of the questionnaire, participants had the option to leave additional comments. Their responses have been used, via direct quotation of participants (Ernest, 1982), to provide detailed understanding on the major challenges and potential of BDAs' implementation in research and practice. Figure 4-2 shows how the data from literature has provided deep insights into the state-of-the-art areas of BDAs' application in Urban Informatics, whilst the questionnaire data provided greater understanding from practitioners, used to further identify the challenges that arise through their use.

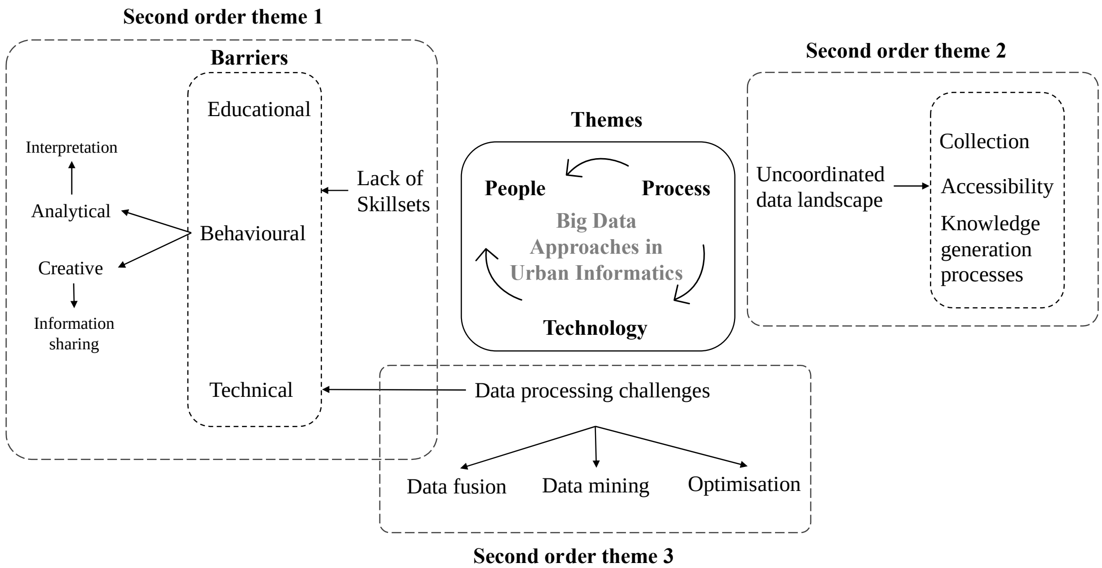


Figure 4-2: Themes generated from data from the literature and questionnaire.

4.6 Results

4.6.1 BDAs for Urban Informatics

The analysis of urban systems is shaped by various economic, social, behavioural, and physical principles (Thakuriah, et al., 2017). Some existing urban models aim to provide an understanding of the complex interactions in urban dynamics, while others aim to improve urban planning and program evaluation. Data-driven research approaches in urban informatics aim to create new methodological approaches using Big Data, while also providing an understanding of urban systems via the use of these unstructured streams of data (Thakuriah, et al., 2017). The urban informatics applications using Big Data can be structured in four categories encapsulating the underlying concepts (content) of urban informatics. The categorisation occurred from the identified research themes covered in the selected literature and can be organised as follows: BDA 1: Spatial and building analysis: form, structure, and performance, BDA 2: Urban Planning and management, BDA 3: High-volume individual behavioural data collection and analysis and BDA 4: Novel approaches with Big Data. The thematic analysis provided insights into the types of BDAs, as well as the technologies used to reveal insights from their application. The findings are described in the following sections.

4.6.1.1 BDA 1: Spatial and building analysis: from a data-scarce to a data rich environment

This first BDA concerns the spatial and building analysis as derived through the literature review process to better understand a design project. In practice, the stages of a design project are defined to help organise the process of briefing, design and construction of buildings and urban projects. These are divided in eight stages and can be found in the Royal Institute of British Architects Plan of Work (RIBA PoW) (RIBA, 2020). Digital tools and new sources of information have been combined to transform several aspects of a traditional design project workflow, which has resulted into a paradigm shift, rather than the sole execution of tasks (RIBA, 2020). Therefore, since 2013, the RIBA PoW has been updated in 2020, incorporating the changes that occurred in the industry the last five years (RIBA, 2013; IBM Corp., 2019).

In this new version, a new section referred to as “BIM Overlay” is included, looking at the growing complexity of data requirements in design and building the foundations for the challenges of using new types of deliverables. The Building Information Modelling (BIM) concept covers building geometries, spatial relationships, energy performance and many other components enabling a greater understanding over the whole life cycle of construction (Pärn, et al., 2017). “BIM Overlay” focuses on the idea of keeping the projects live by embedding data in evidence-based design processes or to serve management purposes.

A rapid growth has been observed in the development of software suitable for spatial data analysis in the period since the late 1980s when few such tools existed (Haining, 1989). Initially the focus of such tools was directed to the integration of spatial statistic methods in the Geographical Information Systems (GIS) environment, and the types of techniques that should be included in such a framework. Nowadays, a collection of spatial data analysis software is readily available, ranging from specialised programs to customized scripts and extensions, as well as open-source software environments such as coding languages, i.e. R (R Core Team, 2017), Java (Arnold, et al., 2005) and Python (Van Rossum & Drake Jr, 1995). In addition, many new platform technologies and Artificial Intelligence (AI) systems are now employed in diverse urban domains (Moore, 2017), while new concepts such as Digital Twins (Batty, 2018) (a general term of the coupling system of infrastructure and deployed Internet of Things (IoT) systems (Parott & Warshaw, 2017)) and BIM are evolving the digital interfaces with infrastructure assets, providing opportunities in the realm of monitoring and prediction.

New tools and data sources offer a range of opportunities for urban design and spatial analysis, from monitoring to real-time management applications for all types of infrastructure. There are many examples of BDAs with data collected in real-time from ticketing systems, vehicle tracking devices, closed-circuit television (CCTV) and others, handling the data via spatial data models (Jenkins, et al., 2016; Li, et al., 2014). However, not all models are suitable to handle large data sets in an efficient way. Over the last decade, novel ways of capturing geospatial information have emerged, such as data collected from geo-sensor networks. The new paradigm has turned from a “data-scarce” to a “data-rich” environment, leading to an increased availability of spatial information (Miller & Goodchild, 2015). Sensor systems are installed in urban infrastructure (transportation, health, water systems, buildings etc.) and they can detect activity and changes in a range of urban phenomena, such as building structures, movement, and activities.

4.6.1.2 BDA 2: Urban Planning and Management: managing the increasing complexity of cities

The second BDA concerns the urban planning and management decision-making. Urban management is an ambiguous concept, as many perceive this as a coupled set of policies, plans and practices maintaining public services, while others think of it as public administration (Kearns & Paddison, 2000; Davey, 1993). Bačlija (2011) defined it as a city administration with the goal of achieving social and economic development. Urban management presents great similarities with urban planning concept, as both have a high degree of complexity and uncertainty upon space, although each of those has its own focus. The latter involves a range of different stakeholders with the aim of improving the urban environment, while urban management tackles urban issues and sets out the bigger planning vision (Hall & Tewdwr-Jones, 2010).

Traditional approaches to research and design of urban systems utilise mathematical models of human spatial interaction. For example, measuring flows of travellers or services between points specified in urban areas (Erlander, 1980; Wilson, 1971) and studying urban structure or the interactions between land-use and mobility (Burgess, 1925; Alonso, 1960; Fujita, 1988). These models are constantly contributing towards planning, policy, and operational decisions the past decades. Therefore, new types of models are increasing in number in which emergent forms of data are being used to manage the increasing complexity of cities and the plethora of available data. For example, machine learning techniques applied to Global Positioning System (GPS) and social media data to predict pedestrian traffic and housing market dynamics (Aschwanden, et al., 2019; Antenucci, et al., 2014; Rae, 2014). Recent studies also employed similar techniques to define quality of urban environments, such as streets and plazas, using image recognition techniques or prediction models at a building scale, such as weekly waste generation or emergency services (Ye, et al., 2019; Kontokosta, et al., 2018).

Agent-based Models (ABM) are another type of complex systems tool used in large-scale modelling practices in the urban planning and management context. Such models use data sources coming from specialised surveys or administrative data, such as GPS trajectories or social network data, which are then studied and used as outcomes to inform individual agent actions and their interactions with the environment. ABMs usually focus on human mobility patterns and individual attributive information, and they typically employ research and statistical techniques (Zellner & Reeves, 2012; Tilahun & Levinson, 2013; Evans & Hugh, 2004).

4.6.1.3 BDA 3: High-volume individual behavioural data collection and analysis

Improvements in sensor systems and embedded systems in daily life allow for personalised data to be collected, contributing towards the research of quality of life and wellbeing. Data capture is now implemented in everyday devices, such as individual smartphones, these being able to acquire environmental geospatial information with a great accuracy. Mobile health and assistive technologies or other wearable and sensor-based physical health recommendation systems have also presented numerous opportunities for researchers to better understand and define the relationship between the built environment and activities in space (Consolvo, et al., 2006; Lin, et al., 2011).

Strategies such as focus groups and workshops were primarily used to generate ideas and solutions to problems. Spatial data were usually treated in traditional research as static data, ignoring their temporal and other unique attributes (Hao, et al., 2015). Individual attributive data were mainly collected by questionnaires or observational data which either lack of dynamic attributes such as preference and satisfaction or are in relatively small sample (Wu, et al., 2017).

Advances in new technologies have introduced new models of citizen's participation into problem solving and planning, allowing idea generation and data analysis from open data portals (e.g., APIs). New models and emerged behavioural patterns have stimulated research into a range of social issues, from active travel and wellbeing to participatory sensing systems. For example, a recent study with data found in social media portals from which researchers are extracting sentiments and opinions evolving through time and space, enabling the monitoring of public opinion (Golbeck & Hansen, 2013). This has allowed the identification of needs and reactions to policy changes, which was not previously possible. These approaches use "natural language processing" algorithms to extract key words and convert qualitative data to quantitative information of millions of data points. However, it should be noted that these types of information are characterised by a strong temporal aspect where users "post" information on a specific time of a day.

4.6.1.4 BDA 4: Novel approaches with Big Data

BDAs are increasingly utilised for urban research and design applications. New forms of data and data-driven modelling techniques allow for urban processes and behaviours to be studied in an efficient and detailed manner. Spatial urban structures, from a micro scale to the macro scale, are constantly reinvented by the end-user's needs and the various interactions among private stakeholders, citizens, and governments. For example, the coupling of distributed energy resources with Information Communication Technologies (ICTs), leading to a new phenomenon of bi-directional energy flow systems, in which the energy consumers become "prosumers" who both produce and consume energy (Luo, et al., 2014). Novel service models have also emerged, altering traditional engagement practices of the various stakeholder groups, and creating new emerging market models facilitated via community-based cloud services (Engin, et al., 2020).

Such innovations bring great advantages and opportunities; however, they also carry a risk of disadvantaging specific segments of a society. Inclusion and wider participation of all parts of society in these informational innovations is necessary, as human-centred urban big data are promising to maximise the shared intelligence of a community. Citizens need to be empowered and provided with the necessary requirements to create a change for themselves, building towards to what Helbing and Pournaras (2015) have defined as "*digital democracies*".

4.6.2 BDAs challenges in their practical application: a view from practice

The questionnaire survey provided useful insights into BDAs challenges in their practical implementation. All respondent stakeholder groups had previous experience in the fields of project management, construction, and design. To obtain a diverse sample and gain representative conclusions, the stakeholders identified were located worldwide. The categories and the professional background of the respondents are displayed in Figure 4-3.

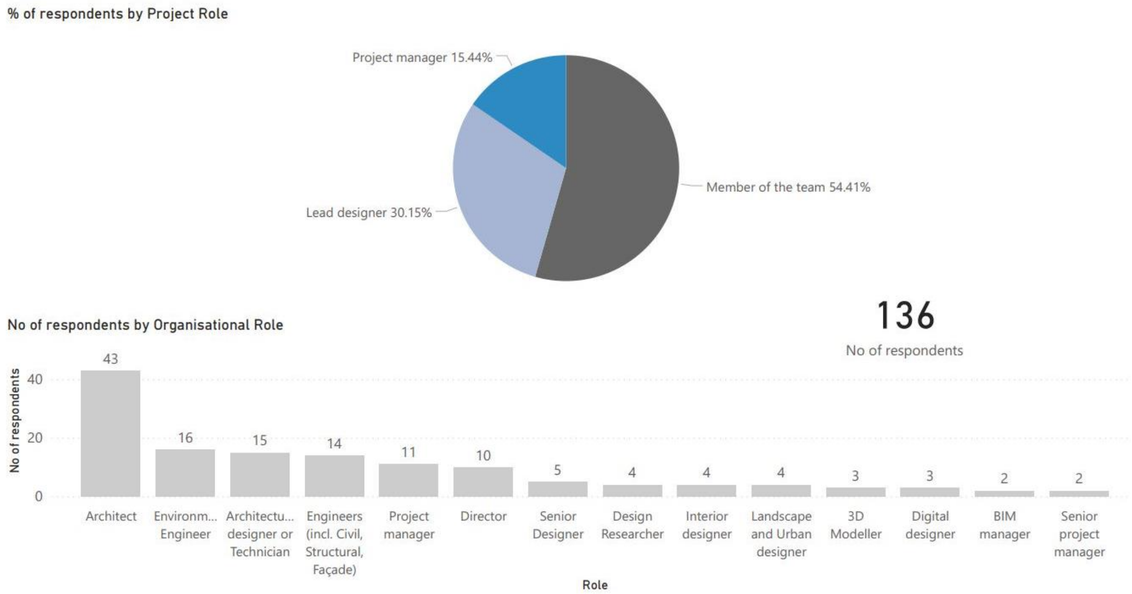


Figure 4-3: Respondent's professional background.

A comparison between levels of experience was investigated, with respondents asked to rate their experience against DDI implementations. Some 44.85% of the respondents had no experience in DDI implementation, while 23.53% had only recently started considering their use. Only 7.53% of the respondents ranked their experience with a '5', where they had been using DDI in their roles, while an extra 9.56% were currently piloting DDI (ranking response of 4).

In general, respondents were positive in their experience of DDI implementation, with 40.44% of the respondents noting that data-driven decision-making was already part of the organisations' culture, while only 16.91% ranked their responses as strongly disagree and disagree. Nevertheless, respondents expressed positive feelings with regard to the use of data, tools and modelling approaches, monitoring, and the collection and accessibility of new streams of data to improve decision-making; 31.62% of the respondents strongly agreed that the use of data can benefit design process within their organisations, while the greatest percentage of 46.32% who ranked their response with '5' – Strongly agree, was in responses to Q04, considering the data processing workflow lifecycle. Negative views in these indicators were limited, with percentages varying between 3.68% to 7.36%, including respondents who ranked their view as either strongly disagree or disagree.

Participants had the opportunity to select more than one possible answer, while ranking in terms of importance was not included. Most of the respondents (51.4%) for Q07 noted that both digital and conventional methods should be employed, while 36.7% of the participants expressed a positive feeling overall and only 10.2% indicated that digitalised design process is not a good idea, and they were not interested in participating. Although the future potential of digital design was recognised by 33% of the participants, 31.6% of the respondents referred on how designers are comfortable sketching using pen/paper to produce quick idea development whenever they like, while 27.9% reflected to the outcome of the process, stating that designs would be limited by the software capabilities.

Responses were similar for Q08, with 41.1% of the participants expressing overall positive feeling in a combined approach with data analytics for design processes, while 9.5% stated their negative views. Although respondents recognised that digital design and data analytics are the future of the design industry, 21.3% of the participants implied that this may be possible to be established in the future. Some of the challenges noted by their selection centred around limitations encountered by software capabilities and design skills, the need for standardisation and sense of scale in such approaches and finally in the cost of their implementation in the process. More specifically, 16.9% of the participants in Q08 noted that there was a need to standardise these approaches to be able to be linked directly to the design process, indicating inability to execute the increased complexity found in DDI and BDAs without specialised skills. More specifically, one of the participants, who ranked its DDI experience as weighting of 2 (Considering DDI), stated:

"*Digitisation and inclusion of 'big data' needs to be weighed against the additional complexity it adds. Often, big data can provide interesting insights that are not actually that useful."*

In addition, it should be noted that in many cases, barriers in utilising BDAs and DDI in design are linked to behavioural aspects, which are not limited only to designers and their willingness to explore other methods of design than traditional ones. Such barriers are present on the clients' side, which can be driven by several other parameters, such as cost of implementation, timeframes or even way of thinking. One of the participants referred to such limitations, highlighting that such information may not receive the appropriate level of interaction, stating:

"*Passion and emotion in design - and not least big "P"- and small "p" politics of clients, whether public or private, remains unpredictable in response to logical design decision processes and evidence-based big data expertise...isn't it?* "

Another implication of the survey results is that only 17% of the respondents had previous experience with DDI, based on weighting values of 4 (Piloting DDI) and 5 (Effectively using DDI). This fact indicates the need for training to enhance understanding of DDI processes rather than using new technologies blindly. One of the respondents that had previously ranked the DDI experience as weighting of 2 (Considering DDI), noted:

"*Lack of uptake in DDI across the industry limits its benefit. Lack of understanding of software, process and resources limits current application and integration in design process."*

This was also raised by another participant who ranked DDI experience as weighting of 1 (No experience), who also added a similar comment (partial extract to reflect the relevant observation):

"*Digital Tools and engineering methodology are the future for construction design processes, but staff training is extremely important to ensure this vision is fully realised."*

4.7 Discussion

A review of relevant literature has identified four BDAs that hold the potential to define the future of planning and design of cities. New streams of data are altering the way cities are defined and operate, changing the focus for policy makers, citizens, and private stakeholders, from the long term to the short term. The expansion from building information mapping and gathering to the macro scale is underway with embedded data in all stages of design, transforming the profession for architects, urban designers, and planners. Although in terms of new technologies and their implementation in the design process, the framework is currently limited to the BIM guidance (RIBA, 2020), it is a radical change compared to previous editions, and it introduces a first step towards dealing with the new challenges that designers face.

The use of BDAs in research and practical applications is not a separate part of technology, instead it is the constant act of seeking new types of information and models to tackle challenges evolving from the everchanging needs of end-users and cities. Theory-driven models and BDAs are necessary in understanding human and natural processes in cities while new urban models, trends and functions need to reflect developments in data-rich environments, which differ from the way urban designers and architects had engaged in the past.

Uptake of BDAs face multiple challenges, limiting their practical implementation in organisational contexts. Koronios et al. (2014) identified the key components determining failure or success of the BDAs application, as being "people, technology and process". Technology and process refer to the scale of investment required in an organisational context, from tools to data management techniques, while people refer to the individuals' abilities to carry out and successfully incorporate BDAs in decision-making process in design and management of urban space. The following sections discuss the three key challenges identified, limiting their practical implementation, namely: Lack of appropriate skillsets; Uncoordinated data landscape: collection and accessibility and Data processing challenges.

4.7.1.1 Lack of appropriate skillsets

One of the biggest challenges faced by the construction industry is the lack of widespread expertise; the right people to implement such strategies. More specifically, the recruitment of individuals who are not solely trained as data scientists or information technology (IT) professionals, but rather urban designers with a range of additional skills (Manthey, et al., 2012). The abilities of designers and planners are no longer limited by the cost of data collection or computing powers, but by the knowledge on data analytics and visualisation techniques. A new data rich environment forms the urban data infrastructure now, from data on systems to behaviours, however, new training methods have not been yet introduced in traditional education (Few, 2009; Cuzzocrea, et al., 2011).

In addition, effective data visualisation can help in communicating key outcomes of the analysis and reveal patterns to decision-makers and public, increasing the efficiency and sustainability of the urban planning, design, and management. Traditionally, there have been several techniques employed to visualise data, such as tables, histograms, charts, and graphs. However, in many cases, these have succeeded in communicating results only to a specific audience of professionals. In addition, the data types are evolving, becoming larger and more complex which pose the challenge to data visualisation. Historically, urban planners and designers are using the language of visualisation via maps and drawings, however, new visualisation techniques have emerged in recent years, such as real-time visualisation tools and platforms, many of them occurring via the use of application programming interfaces (APIs).

New emergencies, such as climate change or the phenomenon of urbanisation, are causing pressure on the maintenance of existing infrastructures, while new ways of living are changing the ways people engage with their environments. Technological innovations reshape the future of urban areas, contributing to the overarching complexity, however, promising to help understand the cities and their needs better. Participants in the questionnaire expressed the need to standardise BDAs and link them directly to the design process, indicating inability to execute the increased complexity found in DDI and BDAs without specialised skills. These findings imply that complexity versus value is not taken into consideration, and in cases such types of approaches often provide information which cannot be directly applied in decision-making in design, due to lack of being relative. This indicates the need to better implement BDAs into the design process, but also up-skill designers to better interpret their outcomes and extract the value from information.

Skills requirements are thus not limited to the technical capabilities of the researchers and practitioners, but should be extended to research, analytical and creative skillsets, which will turn information processing to power. Existing educational limitations, with lack of training programs in BDAs, are restricting the skills to be developed in individuals to produce employees with such expertise.

Another implication is the lack of understanding of decision-making processes and their performance, as these can be deeply personal and subjective. Therefore, concern was raised by the participants regarding the outcomes' veracity, as data can be misrepresented to support what stakeholders would like to see and hence, they can become unreliable. These results highlight educational and contextual barriers which arise by lack of designers' skills in visualising and interpreting the results in a clean and transparent way, conveying information in a manner that is not intended to be misleading.

One of the key constraints identified in BDAs application in research and practice is the lack of appropriate skills from research and professional communities. A summary of the required skills for each individual BDA is provided in Table 4-3, as a result of the literature review. The user communities have been mapped against each set of data types to better understand the diversity of domain knowledge needed for each BDA type (Thakuriah, et al., 2017).

Table 4-3: Summary of skills requirements for each BDA type in urban design and planning. User community's column further developed after Thakuriah et al. [14], p. 18.

BDA type	**Data Types**	**User Community**	**Tools (skill requirements)**
BDA 1	Sensor systems, Administrative data	UC1 and UC3	• Statistical techniques (e.g., aggregate analysis) • Spatial models (e.g., network, topological) • Open-source databases • GIS • Domain-specific knowledge • Programming (e.g., Python or R) • Cloud-based computing • BIM • Visualisation techniques • Data collection techniques (e.g., sensors) • Storytelling/ data communication skills
BDA 2	Sensor systems, Administrative data, Hybrid data	UC1, UC3, UC6	• Statistical techniques (e.g., aggregate analysis) • Spatial models (e.g., network, topological) • Open-source databases • GIS • ABM • Domain-specific knowledge • Machine learning algorithms • Programming (e.g., Python or R) • Visualisation techniques • Storytelling/ data communication skills

BDA 3	UGC, Private Sector Data, Hybrid data Humanities Data	UC2, UC4, UC5, UC6	• Domain-specific knowledge • GIS • ABM • Machine learning algorithms • Fieldwork • Parallel processing • Supercomputers • Programming (e.g., Python or R) • Natural language processing • Visualisation techniques • Data collection techniques (e.g., wearables, sensors) • Storytelling/ data communication skills
BDA 4	Humanities Data, UCG, Private Sector Data	UC2, UC4, UC5	• Domain-specific knowledge • Machine learning algorithms • Parallel processing • Supercomputers • Programming (e.g., Python or R) • Natural language processing • Storytelling/ data communication skills

Those who have domain knowledge and experience in managing the challenges of urban systems hold a critical place to the results that the BDAs will generate (Athey, 2017; Pollard, et al., 2018). Implementation of BDAs in practice requires a change of operations in the industry. Although a growing fraction of construction and design companies have adopted many of these, the promise for less errors in their processes, bring greater pressure for improved performance of their staff, hence resistance to adopt technological changes. Organisations that are concerned about their cultural impact continue their traditional ways of working, rather than investing in training for their existing workforce. However, a new generation of scientists is required who will perform a "hybrid" role. Such a role will require scientists to have the abilities to apply BDAs in novel contexts, while holding domain-specific knowledge, to carefully define the research questions to be explored. Nevertheless, awareness of the technical nature of BDAs is also required from the side of decision-makers, such as policy makers, to ensure their applicability. In addition, as data capturing is growing, to maximise the shared intelligence of a community, sharing and retaining such information of post-processed data will require an understanding of their potential usage and their user communities.

4.7.1.2 Uncoordinated data landscape: collection and accessibility

Data collection efforts and techniques differ based on data type and scale. Some data types are collected via conventional surveys or large-scale collection techniques, such as census data, while others via the use of available emergent technologies and techniques, such as sensor systems or crowd-sourced data. BDAs have been criticised for their applications, as data sources vary in many ways and are often lacking quality, based on the way they have been collected (Feng, et al., 2021; Labrinidis & Jagadish, 2012). Data are being collected via a range of technologies and system architectures, which have at their core four technological developments: Data logging and sensor platforms; Real-time data recording and analysis; New analytic frameworks and Data storage innovations. The overall data landscape is uncoordinated, while lacks data infrastructure, such as knowledge generation processes and policies. The quality of insights derived from advanced analytical processes is closely correlated with data collection and data interpretation choices. Nevertheless, creation and adoption of these standards, as well as maintenance and enrichment of data infrastructures, require a variety of technical, economic and policy expertise, limiting the number of people and organisations that can participate in their creation. Finally, as a lot of open data is derived from personal data, such as national censuses, the data ecosystem needs to be built on ethical considerations regarding its collection and management, to increase levels of trust and ensure equity to those engaging with the data.

4.7.1.3 Data processing challenges

The third and final challenge identified is related to data processing. Quantitative urban research relied on traditional statistics or optimisation methods, which are still relevant, however, some of these may fail due to the high-volume of datasets. Approaches including data mining, machine learning, network analysis, pattern recognition and visualisation techniques are better suited to BDAs. More specifically, data processing techniques involve curation stages, such as cleaning, editing, normalising and feature extraction and selection, which require a great understanding and knowledge of both internal and external validation techniques but also of the context in which data have been collected and are suitable for (Goodchild, 2013). Data processing and analysis techniques can be grouped into three categories, namely: Data fusion, Data mining and Optimisation (Manyika, et al., 2011).

Fusion techniques are used to consolidate data produced by multiple sources, while mining techniques are used to reveal patterns in large datasets. These could be for example relationships between nodes in a transportation network or in walking patterns of pedestrians. Such techniques are of high importance to approaches that are trying to extract complex behaviours but require deep knowledge of information processing. For example, feature matching inherent in fusion techniques when attempting to provide a more complete spatio-temporal representation of individual activities. Finally, optimisation techniques are used to reorganise complex systems and processes to enhance their performance.

All these techniques draw on a range of fields including statistics, computer science, applied mathematics and economics, adding to the challenge of making BDAs accessible to the profession of designers and urban planners. Furthermore, analysis and processing steps inherit biases and limitations, which require a thorough documentation before they can support urban design and management decisions (Longley, et al., 2018). However, in practical applications, even if these assumptions are well-documented and communicated to the design team, the lack of knowledge, from the design team perspective, in data processing and analysis techniques, discourages the critical review and evaluation of the decisions made during the analysis. Hence, these are embedded into the design process as inherited.

Additionally, some datasets, for example social media data (e.g., Twitter), do not have a standardised format and are unstructured. Therefore, significant processing efforts are required to extract useful information. In addition, another key constraint of these data source types is that only a limited percentage of data is geotagged. Thus, the data can be mainly used for calibration purposes (Gaggioli, et al., 2003). Therefore, reducing the risk of lacking statistical significance due to the inaccuracy in the population sample.

Real-time processing also poses the challenge of storage. Therefore, new approaches have emerged, involving in-memory processing. Data storage costs though have dropped significantly and the remote storage of data in data centres has led to “cloud” computing emergence, allowing direct access and analysis of distributed data. This aggregates several disparate workloads in large clusters, dealing effectively with issues of scalability in big data. Nevertheless, it has introduced multiple challenges on how to run and execute such jobs to meet workload goals effectively. At the same time, it allows dealing with the system failures in a structured manner, which occur more often as such systems now operate in larger clusters, requiring complex machine learning tasks.

4.8 Conclusion

This study aims to synthesise and present a state-of-the-art structured analysis of BDAs’ application in Urban Informatics, identifying the challenges that arise through their use. To attain the aim, the authors employed a two-step approach. Firstly, through an investigation of previous literature, using thematic analysis that focuses on identifying and reporting the different types of Big Data challenges and the analytical methods employed. Secondly, an online survey questionnaire was employed to investigate stakeholders’ perceptions in relation to significant challenges and the potential of BDAs implementation in research and practice. This paper highlighted educational, contextual, and behavioural barriers which are limiting the potential of BDAs in the urban design, planning and management context. BDAs offer a new paradigm in research, design, construct, operation, and management of urban space, promising to deliver new insights on societal and environmental complexities. Urban systems encompass a variety of other complex systems, such as environment and transportation, which cannot be expected to be directed towards a traditional, non-digital, approach. These challenges need to be addressed for the BDAs outlined in this study to inform decision-making and to create better environments for its users.

Four BDAs have been identified and their potential to enhance urban research and its practical applications is discussed. BDAs can help to address contemporary urban challenges as they can assist decision-makers in setting the wider urban planning vision or successfully regenerating areas to achieve improved well-being for its users. Efficient implementation of BDAs though, relies on academics, professionals, and policy makers, capable of overcoming challenges associated with data collection, processing, and visualisation techniques. Its stakeholders face challenges in their implementation on a project or an organisation, with the main constraint being the lack of skills in its stakeholders. Lack of expertise has been identified in all the steps of Big Data collection and analysis lifecycle indicating, not only the need for technical skills to execute BDAs, but also their abilities to interpret and embed the results into design and management challenges found in urban systems.

This research directly contributes to existing knowledge by identifying key skillsets required in each of the identified approaches, to help industry and educational organisations in creating curricula better addressed to the future industry standards. Future studies should be focused in addressing the lack of controllability and evaluation of decision-making processes in urban design, while working towards the creation of frameworks to help in standardisation of BDAs, allowing them to be better embedded in design processes. The recipients of the findings will be the urban planners, decision-makers and academics who are interested in delivering urban environments aligned to the end-users' needs, utilising novel technologies.

REFERENCES

Alonso, W., 1960. A Theory of the Urban Land Market. *Regional Science,* 6(1), pp. 149-157.

Antenucci, D., Cafarella, M., Levenstein, M.C., Ré, C. & Shapiro, M.D., 2014. *Using Social Media to Measure Labor Market Flows,* Michigan: Report of the University of Michigan node of the NSF-Census Research Network (NCRN) supported by the National Science Foundation under Grant No. SES 1131500.

Arnold, K., Gosling, J. & Holmes, D., 2005. *The Java programming language.* Second edition ed. Boston, MA United States: Addison-Wesley Longman Publishing Co., Inc.

Aschwanden, G., Wijnands, J.S., Thompson, J., Nice, K.A., Zhao, H. & Stevenson, M., 2019. Learning to walk: Modeling transportation mode choice distribution through neural networks. *Environment and Planning B: Urban Analytics and City Science,* 48(1), pp. 186-199.

Athey, S., 2017. Beyond prediction: Using big data for policy problems. *Science,* 355(6234), pp. 483-485.

Bačlija, I., 2011. Urban management in a European context. *Urbani Izziv,* 22(2), pp. 137-146.

Batty, M., 2014. Can It Happen Again? Planning Support, Lee's Requiem and the Rise of the Smart Cities Movement. *Environment and Planning B: Planning and Design ,* 41(3), pp. 388-391.

Batty, M., 2018. Digital twins. *Environment and Planning B: Urban Analytics and City Science,* 45(5), pp. 817-820.

Braun, V. & Clarke, V., 2012. Thematic analysis.. In: *H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), APA handbook of research methods in psychology, Vol. 2. Research designs: Quantitative, qualitative, neuropsychological, and biological.* s.l.:American Psychological Association, pp. 57-71.

Burgess, E., 1925. The Growth of the City: an Introduction to a Research Project. In: *R. E. Park, E. W. Burgess and R. D. Mackenzie (eds). The City..* Chicago: University of Chicago Press, pp. 47- 62.

Consolvo, S., Everitt, K., Smith, I. & Landay, J., 2006. *Design requirements for technologies that encourage physical activity.* Chicago, USA, Association for Computing Machinery, p. 457–466.

Cuzzocrea, A., Song, I.-Y. & Davis, K., 2011. *Analytics over large-scale multidimensional data: the big data revolution!.* Glasgow, UK, Association for Computing Machinery.

Davey, K., 1993. *Elements of urban management,* Washington DC: The World Bank.

Engin, Z. et al., 2020. Data-driven urban management: Mapping the landscape. *Journal of Urban Management,* 9(2), pp. 140-150.

Erlander, S., 1980. *Optimal Spatial Interaction and the Gravity Model.* Verlag Berlin and Heidelberg GmbH & Co. KG: Springer.

Ernest, A., 1982. Design in the Decision-Making Process. *Policy Sciences,* 14(3), pp. 279-92.

Evans, T. & Hugh, K., 2004. Multi-scale analysis of a household level agent-based model of landcover change. *Journal of Environmental Management,* 72(1-2), pp. 52-72.

Feng, Y., Duives, D., Daamen, W. & Hoogendoorn, S., 2021. Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment,* Volume 187, Article 107329.

Few, S., 2009. *Now you see it: Simple visualization techniques for quantitative analysis.* Berkeley, USA: Analytics Press.

Foth, M., Choi, J. H.-J. & Satchell, C., 2011. *Urban Informatics.* New York, NY, USA, Association for Computing Machinery, pp. 1-8.

Fox, N., Mathers, N. & Hunn, A., 2000. Surveys and Questionnaires. In: W. M. H. B. Wilson A, ed. *Research Approaches in Primary Care.* London, UK: Radcliffe Medical Press/Trent Focus.

Fujita, M., 1988. A monopolistic competition model of spatial agglomeration: Differentiated product approach. *Regional Science and Urban Economics,* 18(1), pp. 87-124.

Gaggioli, A., Bassi, M. & DelleFave, A., 2003. Quality of experience in virtual environment.. In: *G. Riva, F. Davide, & W.A. IJsselsteijn (Eds.), Being there: Concepts, effects and measurement of user presence in synthetic environments.* Amsterdam: Ios Pres.

Gigerenzer, G. & Goldstein, D., 2002. Models of ecological rationality: The recognition heuristic. *Psychological Review,* 109(1), pp. 75- 90.

Golbeck, J. & Hansen, D., 2013. A method for computing political preference among Twitter followers. *Social Networks,* Volume 36, pp. 177-184.

Goodchild, M., 2013. The quality of big (geo)data. *Dialogues in Human Geography,* 3(3), pp. 280-284.

Gowans, H., Elliot, M., Dibben, C. & Lightfoot, D., 2012. *Accessing and sharing administrative data and the need for data security,* Swindon, UK: Administrative Data Liaison Service.

Grommé, F., 2016. Provocation: Technology, resistance and surveillance in public space. *Environment and Planning D: Society and Space,* 34(6), pp. 1007-1024.

Haining, R., 1989. Geography and spatial statistics : current positions, future developments. In: *Macmillan B (ed) Remodelling Geography.* Oxford: Basil Blackwell, p. 191–203.

Hall, P. & Tewdwr-Jones, M., 2010. Urban and regional planning. *Routledge,* Volume 5th ed., p. 3.

Hao, J., Zhu, J. & Zhong, R., 2015. The rise of big data on urban studies and planning practices in China: Review and open research issues. *Journal of Urban Management,* Volume 4, pp. 92-124.

Helbing, D. & Pournaras, E., 2015. Society: Build digital democracy. *Nature,* Volume 527, pp. 33-34.

Hilbig, B. & Pohl, R., 2008. Recognition users of the recognition heuristic. *Experimental Psychology,* 55(6), pp. 394-401.

Huijboom, N. & van den Broek, T., 2011. Open data: an international comparison of strategies. *European Journal of ePractice,* 12(1), pp. 1-13.

Hu, Y. & Han, Y., 2019. Identification of Urban Functional Areas Based on POI Data: A Case Study of the Guangzhou Economic and Technological Development Zone. *Sustainability,* 11(5), p. 1385.

IBM Corp., 2019. *IBM SPSS Statistics for Windows, Version 26.0.* NY: Armonk, NY: IBM Corp.

Jenkins, A., Croitoru, A., Crooks, A. & Stefanidis, A., 2016. Crowdsourcing a Collective Sense of Place. *Public Library of Science,* 11(4), Article e0152932.

Kearns, A. & Paddison, R., 2000. New challenges for urban governance. *Urban Studies,* 37(5-6), pp. 845-850.

Kontokosta, C., 2018. Urban Informatics in the Science and Practice of Planning. *Journal of Planning Education and Research,* 41(4), pp. 382-395.

Kontokosta, C., Hong, B., Johnson, N. & Starobin, D., 2018. Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities. *Computers, Environment and Urban Systems,* Volume 70, pp. 151-162.

Koronios, A., Gao, J. & Selle, S., 2014. *Big Data Project Success - a Meta Analysis.* Chengdu, China, AIS.

Krizek, K., Forysth, A. & Schively Slotterback, C., 2009. Is There a Role for Evidence-Based Practice in Urban Planning and Policy?. *Planning Theory & Practice,* Volume 10, pp. 459-478.

Labrinidis, A. & Jagadish, H., 2012. Challenges and opportunities with big data. *Proceedings of the VLDB Endowment,* 5(1), pp. 2032-2033.

Lin, Y., Jessurun, J., de Vries, B. & Timmermans, H., 2011. *Motivate: Towards context aware recommendation mobile system for healthy living.* Dublin, Ireland, IEEE, p. 250–253.

Li, Q., Shao, C.-F. & Zhao, Y., 2014. A robust system for real-time pedestrian detection and tracking. *Journal of Central South University,* 21(4), pp. 1643-1653.

Loh, C., 2017. Learning from Practice, Learning for Practice in Local Land Use Planning Research. In: T. W. Sanchez, ed. *Planning Knowledge and Research.* London, UK: Routledge, p. 24–34.

Longley, P., Cheshire, J. & Singleton, A., 2018. *Consumer Data Research.* London: UCL Press.

Luo, Y., Itaya, S., Nakamura, S. & Davis, P., 2014. *Autonomous cooperative energy trading between prosumers for microgrid systems.* Edmonton, Canada, IEEE, pp. 693-696.

Manthey, K., Kobelius, J. & Krishnan, K., 2012. *Big Data for Enterprise: Technology , Strategy , Adoption and Outlook Practical advice, recommendations and predictions on the expanding application of big data in enterprise.* San Francisco, Data Driven Business.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Hung, B.A., 2011. *Big data: The next frontier for innovation, competition and productivity.,* New York: McKinsey & Company: McKinsey Global Institute.

Miller, H. & Goodchild, M., 2015. Data-driven geography. *Geojournal,* 80(4), pp. 449-461.

Moore, S., 2017. *Opportunities for conversational AI in government.* [Online] Available at: https://www.gartner.com/smarterwithgartner/opportunities-for-conversational-ai-in-government/
[Accessed 27 03 2022].

Open Data Institute, 2018. *Open standards for data.* [Online] Available at: https://standards.theodi.org/

OpenDataSoft, 2018. *A comprehensive list of 2600+ open data portals around the world.* [Online]

Available at: https://www.opendatasoft.com/a-comprehensive-list-of-all-open-data-portals-around-the-world/

Pärn, E., Edwards, D. & Sing, M., 2017. The building information modelling trajectory in facilities management: A review. *Automation in Construction,* Volume 75, pp. 45-55.

Paganin, G., Talamo, C. & Atta, N., 2018. Knowledge management and resilience of urban and territorial systems.. *TECHNE - Journal of Technology for Architecture and Environment,* Volume 15, pp. 124-133.

Pan, Y., Tian, Y., Liu, X., Gu, D. & Hua, G., 2016. Urban Big Data and the Development of City Intelligence. *Engineering,* 2(2), pp. 171-178.

Parott, A. & Warshaw, L., 2017. *Industry 4.0 and the digital twin: Manufacturing meets its match,* USA: Delloitte.

Pollard, J., Spencer, T. & Jude, S., 2018. Big Data Approaches for coastal flood risk assessment and emergency response. *WIREs Climate Change,* 9(5), Article e543.

R Core Team, 2017. *R: A Language and Environment for Statistical Computing,* Vienna, Austria: R Foundation for Statistical Computing.

Rae, A., 2014. Online Housing Search and the Geography of Submarkets. *Housing Studies,* 30(3), pp. 453-472.

Reddy, G., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G. & Baker, T., 2020. Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access,* Volume 8, pp. 54776-54788.

Resch, B., Puetz, I., Bluemke, M., Kyriakou, K. & Miksch, J., 2020. An interdisciplinary mixed-methods approach to analyzing urban spaces: The case of urban walkability and bikeability. *International Journal of Environmental Research and Public Health,* 17(19), p. 6994.

RIBA, 2013. *The RIBA Plan of Work,* London: RIBA.

RIBA, 2020. *RIBA Plan of work,* London: RIBA.

Sanchez, T. W., 2020. The Most Frequently Cited Topics in Urban Planning Scholarship. *Urban Science,* Volume 4, p. 4.

Shah, A. & Oppenheimer, D., 2008. Heuristics made easy: An effort-reduction framework.. *Psychological Bulletin,* 134(2), pp. 207-222.

Shi, Q. & Abdel-Aty, M., 2015. Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies,* Volume 58, pp. 380-394.

Sivarajah, U., Kamal, M., Irani, Z. & Weerakkody, V., 2017. Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research,* Volume 70, p. 263–286.

Thakuriah, P., Tilahun, N. & Zellner, M., 2017. Big Data and Urban Informatics: Innovations and Challenges to Urban Planning and Knowledge Discovery. In: *Thakuriah P., Tilahun N., Zellner M. (eds) Seeing Cities Through Big Data.* Switzerland:Springer Geography, Springer, Cham., pp. 11-45.

Tilahun, N. & Levinson, D., 2013. An Agent-Based Model of Origin Destination Estimation (ABODE). *The Journal of Transport and Land Use,* 6(1), pp. 73-88.

Ubaldi, B. & OECD, 2013. *Towards Empirical Analysis of Open Government Data Initiatives,* OECD iLibrary: Open Government Data.

Van Rossum, G. & Drake Jr, F., 1995. *Python reference manual.* Amsterdam: Centrum voor Wiskunde en Informatica Amsterdam.

Wilson, A., 1971. A family of spatial interaction models, and associated developments. *Environment and Planning,* 3(1), pp. 1-32.

Wu, H., Liu, L., Yu, Y. & Peng, Z., 2017. Evaluation and Planning of Urban Green Space Distribution Based on Mobile Phone Data and Two-Step Floating Catchment Area Method. *Sustainability 2018,* Volume 10, p. 214.

Ye, Y., Zeng, W., Shen, Q., Zhang, X. & Lu, Y., 2019. The visual quality of streets: A human-centred continuous measurement based on machine learning algorithms and street view images. *Environment and Planning B: Urban Analytics and City Science,* 46(8), pp. 1439-1457.

Zellner, M. & Reeves, H., 2012. Examining the contradiction in ‘sustainable urban growth’: an example of groundwater sustainability. *Journal of Environmental Planning and Management,* 55(5), pp. 545-562.

Zhao, Y., Zhang, G., Lin, T., Liu, X., Liu, J., Lin, M., Ye, H. & Kong, L., 2018. Towards Sustainable Urban Communities: A Composite Spatial Accessibility Assessment for Residential Suitability Based on Network Big Data. *Sustainability,* 10(12), p. 4767.

5 Investigating pedestrian behaviour in urban environments: a Wi-Fi Tracking and Machine Learning approach

5.1 Abstract

Urban geometry plays a critical role in determining paths for pedestrian flow in urban areas. To improve urban planning processes and enhance the quality of life for end-users in urban spaces, decision-makers within the urban design and planning industry require a better understanding of the factors influencing pedestrian movement. This study presents a novel means to assess pedestrian routing in urban environments. A methodology framework is provided to classify pedestrian behaviours and spatial configuration interactions and utilise machine learning approaches applied to location data derived from Wi-Fi tracking techniques. The approaches developed can be used for observations in large-scale contexts, where traditional methods currently prove inadequate. Application of the framework in a high pedestrian traffic-dense retail urban area in London reveals clear and consistent relationships amongst spatial visibility, individuals' motivation, and knowledge of the area. Key behaviours established in the study area are grouped into two activity categories: (i) Utilitarian walking (with motivation - expert and novice striders) and (ii) Leisure walking (no motivation - expert and novice strollers). The approach offers an insightful means to understand pedestrian routing in urban contexts and informs wider wayfinding, walkability, and transportation knowledge.

5.2 Introduction

Interest in urban areas has increased in recent years due to rapid urbanisation, creating new demands to understand better the morphology of city open spaces (Imants, et al., 2021; Chang, et al., 2020). Urban planners and designers seek information to aid an understanding of multiple human needs, e.g., the provision of open spaces for recreation and improved ecosystem services via microclimate regulation and biodiversity provision (Reid, 2005; Andersson, et al., 2007).

Revealing both urban resident and visitor behavioural patterns is fundamental in designing and quantifying urban planning decisions (Gonzalez, et al., 2008; Pettit, et al., 2016). Movements within street networks aid planners in implementing road development, public transport, and placement of amenities, and create designs promoting physical activity, healthier communities, and well-being (Järv, et al., 2012; O'Sullivan, et al., 2000; Brown, et al., 2014). Nevertheless, although diversity in walking activities is generally recognised in wayfinding and walkability studies (Choi, 2012; Cornell, et al., 2003; Bitgood, 2010), there is not yet systematic knowledge about how to best categorise pedestrian walking behaviours, leaving a gap in knowledge. Partitioning the walking activities to reflect the different purposes and how the built environment influences provide insights into how urban spaces should be designed to address the needs of their inhabitants.

Increased urban data availability has renewed the interest in urban mobility for better understanding pedestrian activity and the effects of diverse physical factors on behaviour. Nevertheless, studies exploring human experience in urban spaces are mainly based on traditional data sources and analysis methods, such as observational data and statistical techniques (Hillier, et al., 1993; Krizek, et al., 2009; Gunn, et al., 2017). Such approaches highlight the potential of data-driven methodologies in supporting more informed design decisions, and enrich urban design and planning theory (e.g., (Whyte, 1980; Gehl, 2011)). However, it is unclear how novel sources of information and approaches could be realised to provide insights on pedestrian movement behaviours in urban spaces, leaving a second gap.

This study presents a novel means to assess pedestrian routing in urban environments. A framework methodology is proposed to classify behaviours and spatial configuration interactions, utilising machine learning (ML) algorithms and location data derived from Wi-Fi tracking techniques. The study provides a comprehensive review of the parameters influencing pedestrian movement and classification and identifies critical gaps in the current data collection and analysis methods employed in pedestrian movement behaviour, identifying opportunities to bridge these gaps.

5.3 Pedestrian behaviour and spatial production

5.3.1 Pedestrian movement behaviour in Urban areas

A range of parameters influences the human experience in urban spaces (Choi, 2014; Bozovic, et al., 2020). It is characterised by movements made and the duration spent in an area, forming together a dimension of the experience of movement. The successive experiences a user gains while moving through an urban environment create a harmonious and progressive space characterised as a specified and continuous structure with different qualitative attributes, e.g., satisfaction, comfort, and safety (Lynch, 1960). As the individuals choose diverse routes for differing purposes, e.g., commuting or shopping (Ki & Lee, 2021), a selection of points collectively representing their route can be analysed to understand better when specific areas are visible at a particular location and how this may influence movement patterns (Batty, et al., 1998).

Pedestrian movement and experiences gained within a space are influenced by the underlying motivation of the pedestrian to reach a specific place at a certain time (Dridi, 2015). Key destinations may form the person’s end goal, but the route followed constitutes their experience. Literature on wayfinding identifies two key types of behaviours within complex built environments. These are goal-oriented and non-goal-directed or exploratory behaviours (Gibson, 1988). The first refers to pedestrians’ motivation in moving towards specific points within a space, e.g., residential buildings or transit terminals (Wang, et al., 2014). The second is stimulated by visually attractive objects encountered along the path to the goal, e.g., window displays or street performances (Dridi, 2015; Forsyth, et al., 2008).

Numerous studies have attempted to provide theories reflecting pedestrians’ navigational behaviour in cases where there is no specific destination. Such studies base conclusions on simulations or survey data interpretation, assuming movements are directed along the lines of sight. In these cases, locales, where more lines of vision meet, are more likely to be chosen by individuals (Batty, et al., 1998; De Arruda Campos, 1997; Hillier, et al., 1993). Spatial visibility can be defined as the visible locations in a spatial layout (Turner, et al., 2001). Pedestrians with limited knowledge of an area (visitors) might seek areas of higher space visibility for better orientation. Previous research indicates that speed may vary based on the type of goal, permitting classification of pedestrian types against goal-oriented activities. Thus, commuters are slightly slower than workers, while shoppers are slightly faster than leisure trippers (Zhang, et al., 2020).

Research suggests that success in reaching a destination reflects the complexity of the spatial configuration (Montello, 2005; Carlson, et al., 2010). Where a journey is exploratory, information collected is not ordered but affects perception. The design of the built environment can therefore encourage and inhibit both individual behaviours and the act of wayfinding (Kuliga, et al., 2019). Two key factors that affect pedestrian movement are the spatial characteristics of a given setting, e.g., visibility, layout, or diversity, and the wayfinding support system, such as signage and information boards (Weisman, 1981; Montello, 2005; Li & Klippel, 2012).

Whyte (1980) observed how people seek more than mere physiological comfort and how pedestrians will consequently undergo a certain degree of physical discomfort to satisfy psychological needs. Gehl (2011) simplified outdoor activities in urban spaces into three categories: necessary, optional, and social. When the quality of the urban environment is good, optional activities increase in frequency. As those activities rise, the number of social activities also increases (Gehl & Gemzoe, 1996). Other researchers have divided the activities into utilitarian and leisure (Ki & Lee, 2021), while others have followed similar categorisations (Table 5-1).

Table 5-1: Summary of activities categorisation in research

Walking activities categories	Source	Title	Date
• Utilitarian walking • Leisure walking	Ki, D. and Lee, S. (Ki & Lee, 2021)	Analyzing the effects of Green View Index of neighborhood streets on walking time using Google Street View and deep learning	2021
• Walking for transport • Walking for recreation	Zhang X., Melbourne S., Sarkar C., Chiaradia A., Webster C. (Zhang, et al., 2020)	Effects of green space on walking: Does size, shape, and density matter?	2020
• Essential trips for commuting • Optional trips for recreational activities	Lee, J.M. (Lee, 2020)	Exploring Walking Behavior in the Streets of New York City Using Hourly Pedestrian Count Data	2020
• Stationery activities • Passer-by activities • Cultural and social activities	Istrate et al. (Istrate, A.-L.; Bosák, V.; Nováček, A.; Slach, O., 2020)	How Attractive for Walking Are the Main Streets of a Shrinking City	2020
• Optional • Necessary • Social (Resultant activities)	Jan Gehl & Birgitte Svarre (Gehl & Svarre, 2013)	How to Study Public Life	2013

Activity types define pedestrian motivations and needs (Malleson, et al., 2018). According to Transport For London (TFL) research, there are four different and distinct types of journeys, each with specific travel characteristics, thus: Novice strider, Expert strider, Novice stroller, and Expert stroller (Davies, 2007). Within this categorisation, knowledge of the area is incorporated for all activities. Previous research has indicated that visual cues create images for a given observer, combining to generate individual perceptions of surrounding environments (Lynch, 1960). This "stored" information provides a mental map utilised opportunistically to facilitate movement. Nevertheless, these theories categorise pedestrian movement and activities by simplifying complex aspects of walking and excluding specific groups of people. These include people with long-term impairments (e.g., blind, or neuro-divergent people) or walking constraints (e.g., walking with children, or carrying something heavy) (Deluka-Tibljaš, et al., 2022).

Environmental parameters are also recognised as key criteria influencing walking activities within public spaces (Brum-Bastos, et al., 2018; Gehl, 2010). Aspects of prevailing weather also impact, in different magnitudes, human behaviours, with the more significant factors identified as solar radiation and wind speed (De Montigny, et al., 2012). However, although the literature suggests strong correlations exist between walking patterns and seasonality, the effect of seasonality from the perspective of time within the year or a single day is not yet well examined.

5.3.2 The use of novel tools and techniques to explore pedestrian movement behaviour in Urban areas

Along with the shift from the 1960's car-oriented street design to today's user-experience focus, the enhancement of pedestrian experience is gaining importance in traffic engineering and urban planning. However, in practice, pedestrian volumes are analysed, primarily focusing on traffic safety, footfall in retail units, and the impacts of new developments and social value (Zhang & Fricker, 2021; Klein, et al., 2016; Pikora, et al., 2002). There is considerable research on the spatial orientation of humans within urban environments and on the prediction of individual movements (Gehl, 2011; Gehl & Svarre, 2013; Zacharias, 2001; Mehta, 2009; Clifton, et al., 2007). However, research on urban space recognition and the role of sensorial experience on movement patterns remains limited. Existing studies have illustrated the necessity of contemporary data collection and analysis methods while highlighting a lack of novel techniques employed in specific research areas and illustrating the current limitations concerning the types of pedestrian behaviour that can be studied with traditional approaches (Feng, et al., 2021). For example, recording concurrent crowd movements in public spaces via observational data is difficult, introducing biased information collected via experimental setups.

Evidence-based decision-making using observational data improves planning policies and urban design processes (Krizek, et al., 2009), highlighting the potential of such approaches in supporting more informed design decisions. New data streams and methods, such as open-source data or sensor systems, can assist designers and urban planners extract information related to pedestrian activity and develop models elaborating the diverse drivers and trade-offs influencing human behaviour. Multiple methods have been introduced assessing the street environment from the pedestrian's view to help understand user needs. These methods currently involve surveys, observational data collection, and structured assessments related to the availability of services, aesthetic appearance, and cleanliness of the space (Jones, et al., 2008).

There, therefore, exists a gap between human mobility behaviour and the influence of the built environment. Research studies have been primarily based on observational data and simulation (Batty, et al., 1998; De Arruda Campos, 1997; Hillier, et al., 1993; Kürkçüoğlu & Akin, 2013; Askarizad & Safari, 2020). The complex question of societal behaviour and spatial production was developed in architecture and urban planning research in the early 80s, leading to the investigation of various spatial models, known as space syntax research. However, such approaches remain until today limited to statically measuring space (e.g., geometry or topology) (Hillier & Hanson, 1984; Till, 2007; Capitanio, 2019). Whyte's work (1980) was seminal, one of the first attempts to quantify human activity in open spaces using data-driven approaches. However, the manual gathering of observational data remains time-consuming and labour-intensive and limits the scope of research. Although there are opportunities to capture unbiased behavioural data, this can result in relatively small sample sizes, not representative of the population, and further lacking temporal information (Feng, et al., 2021).

Therefore, the introduction of new sources and types of data presents opportunities to understand end-user needs better, capturing ‘panoptic’ data that is not easy to observe in the real world and addressing problems at both the city and neighbourhood scale. At the same time, such information reduces traditional approach limitations, such as cost and scale implications from traditional data collection techniques. For example, the employment of large-scale monitoring via smartphones and sensor networks enables researchers to study crowds in large settings (Wirz, et al., 2013). Such approaches though were limited due to a lack of information and validity (Duives, et al., 2019). New streams of data and models of citizens’ participation in problem-solving (Golbeck & Hansen, 2013) have stimulated research into a range of social challenges, e.g., health and mobility (Consolvo, et al., 2006). Key to this is the mobile phone, now being an integral part of human life, which has been used widely in urban planning and transportation sector applications (Shi & Abdel-Aty, 2015; Moreira & Ferreira, 2016; Martín, et al., 2019). The device itself is transformed into a complex gadget that includes multimedia technologies that can reveal user preferences regarding commercialism, daily routines, and cultural choices (Lee, K-S., 2011). Devices also have Wi-Fi and Bluetooth radio communication and connections of these to triangulated base station nodes can also be used for precise geolocation.

Consequently, the rapid development of the internet of things (IoT) and information and communication technology (ICT) is altering how people orient themselves in space (Kötteritzsch & Weyers, 2016; Lin, et al., 2011; Consolvo, et al., 2006). These tools also contribute to changing the research methods employed in the building design sector (Boniface, et al., 2015; Resch, et al., 2020; Szczepanek, 2020; Liao, et al., 2016). The increasing availability of data sources offers opportunities for researchers to renew the concepts and methods currently used in urban space design. A wide variety of new methodologies often referred to as “Big Data Approaches” (BDAs), have become apparent, including machine learning, network analysis, and visualisation techniques. These methodologies are used to better capture and analyse a range of complex urban space problems (Aschwanden, et al., 2019; Kontokosta, et al., 2018; Díaz-Álvarez, et al., 2018). Additionally, new types of approaches, such as Digital Twins (Batty, 2018) and Artificial Intelligence (AI) systems (Moore, 2017), are employed in diverse urban domains, evolving monitoring and prediction approaches, leading to new paradigms of connected data systems (Aschwanden, et al., 2019; Van Dijk, 2018; Ye, et al., 2019; Moore, 2017).

BDAs have been employed to manage increasing urban data complexity in the city, with ML techniques used, for example, in transportation and environmental studies (Aschwanden, et al., 2019; Kontokosta, et al., 2018). BDAs refer to the combination of diverse datasets and related technologies to extract insights from complex systems via novel organisational and analytical capabilities (Pollard, et al., 2018). ML techniques are a series of computational algorithms that imitate human intelligence while studying self-improvement methods to obtain better performance and knowledge (Wang, et al., 2009; El Naqa & Murphy, 2015). Unsupervised ML, and more specifically, clustering algorithms have been used in retail to reveal customer behaviours, while other researchers have used similar principles to cluster transit information based on temporal and spatial characteristics (Mauri, C., 2003; Chang & Chen, 2009; Ma, et al., 2013). Pedestrians are characterised by several key features, e.g., personal choice to move. However, their movement into space is greatly expanded by variations of pedestrian profiles, such as age, speed of movement, previous experience, time of day, occupation density, etc.

Although these models promise to provide answers to urban challenges, it is yet unclear to what extent and under which circumstances they can provide insights in the study of movement behaviours in the context of space recognition and individual preferences. An imbalance is evident in urban space recognition and pedestrian movement literature regarding the use of novel sources of information and models. Two gaps are identified regarding the collection, analysis, and existing research in current pedestrian behaviour studies. The first one relates to the lack of systematic knowledge in pedestrian behaviours categorisation. The second is the limited use of novel sources of information and models for studying pedestrian behaviour and spatial production.

This study presents a novel means to assess pedestrian routing in urban environments. A framework methodology is provided to classify pedestrian behaviours and spatial configuration interactions and utilise ML approaches applied to location data derived from Wi-Fi tracking techniques.

5.4 Study Area

A high-street in London, UK, was selected as a case study site to assess pedestrian walking patterns (Figure 5-1). More specifically, Oxford Circus in London is a road junction and one of the busiest pedestrian crossings in the city, connecting two of the most prominent retail streets, Oxford Street and Regent Street, located in London's West End. Oxford Street is a key transportation corridor, used extensively by taxis and cyclists and providing east-west routes for bus services and tube stations. The Oxford Street district includes residential and retail areas with commercial/office use. Regent Street is a significant shopping street containing flagship retail stores. Regent Street is c.1.3km long, while Oxford Street is 1.9km. Their intersection was transformed in 2009 from a segregated junction with barriers with a limited overflow of pedestrians to an open diagonal crossing allowing pedestrians to follow their desired route. This change reflects a shift in street design towards the concept of integration and space sharing to improve the quality of environments, further enhanced by removal of street furniture, and with as many shared "single" surfaces as possible (Mercieca, et al., 2011).

The study area was chosen based on the availability of datasets, the mix of uses, and street networks connecting to wider residential areas. Therefore, the study area comprises an urban context with significantly different conditions and potentials, while it presents environments similar to those found in many other cities and urban areas.

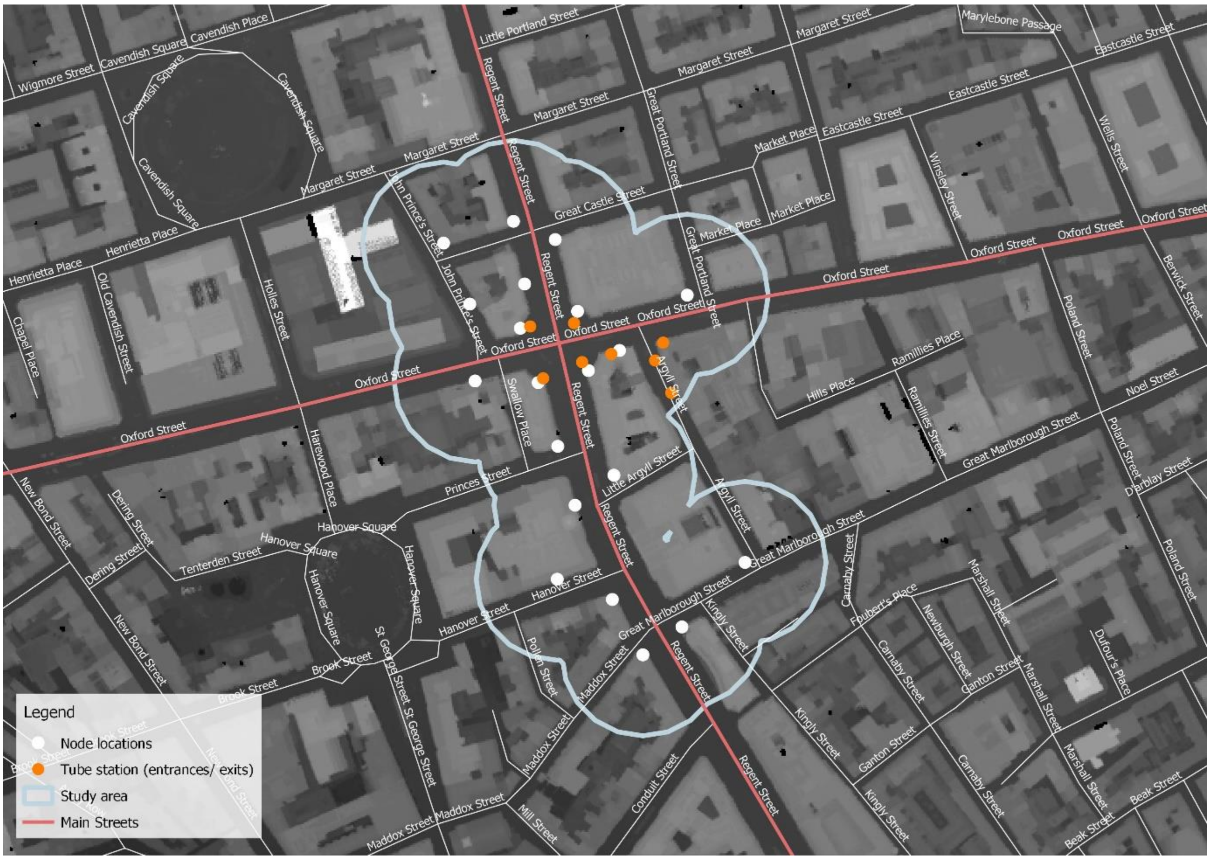


Figure 5-1: Study area and the distribution of Wi-Fi nodes

5.5 Materials and methods

5.5.1 Data preparation

This study utilises multiple data sources, allowing several spatial attributes to be explored. These data include pedestrian movement, urban geometry, and weather information. All datasets selected, except from the Wi-Fi location data, are publicly available, described in Table 5-2. Data information reflects the category of information captured in each dataset. Geometry type and date describe the shape of the captured information and the timeframe.

Table 5-2: Types of data collected for this study with source

Data information	Geometry type	Date	Source
Pedestrian Data	Point	2017	Wi-Fi tracking
Important buildings & Infrastructure	Point	2018	Ordance Survey: (https://osmaps.ordnancesurvey.co.uk/) Open street map: (https://www.geofabrik.de/geofabrik/) POIS: (https://osmaps.ordnancesurvey.co.uk/) Registered energy performance certificates (EPC) non-domestic: (www.gov.uk)
Park areas	Polygon		
Transportation access (bus stops/ tube entrances) location	Point		
Street geometry	Polygon		
Amenities	Point		
Hourly temperature, humidity & weather events	Point (temporal)	2017	Weather Underground/ Private weather stations

Pedestrian movement data, obtained by Wi-Fi tracking, was provided by The Crown Estate (The Crown Estate, 2021), with data deriving from an earlier study aiming to inform a base year model used to simulate existing pedestrian flow conditions and predictive impact of changes in demand or spatial layouts (Angelelli, et al., 2018). Data was collected by capturing signals from Wi-Fi devices across 19 Open Mesh nodes (OM2P-HS) attached to floodlights on building cornices (Figure 5-1– node location). The nodes were installed at 3 to 5 m height to be clear from obstructions that could affect signal reception. Data was transmitted in real-time via the 3G/4G mobile network for storage in a cloud-based server. The accuracy of the data has been tested in that same study, based on a comparison of the Wi-Fi obtained data and closed-circuit television (CCTV) image data.

The mobile data used covers the period of August to October 2017 (Figure 5-2), in total 22 days, incorporating 3,240,361 unique mobile users. The August period presents the most extensive continuous data sample and is the only period including weekend data collection. Data were pre-processed by the technology provider (Accuware Inc, 2017), eliminating all privacy-related information. The outputs were a multi-field .csv file per day, capturing unique Media Access Control (MAC) addresses, signal strengths, X Y Z coordinates location, and a timestamp. MAC addresses are unique identifiers assigned to a network interface controller for use as a unique network address, common in technologies, such as Ethernet, Wi-Fi, or Bluetooth. The provider used signal strength for triangulation purposes to derive location, indicating nodes in greater proximity to the recording devices. Data were captured on a frequency varying from 1-60 seconds, depending on the type of handset devices, manufacturer, or activity level (Accuware Inc, 2017).

All the data handling, storage, processing, and presentation observe the data security and privacy requirements specified in General Data Protection Regulation (GDPR) on handling personal data and protecting privacy (European Parliament and of the Council, 2016). Personal information was truncated via the system and converted to a non-personal form, thus permitting information collection without consent (Fuxjaeger & Ruehrup, 2018).

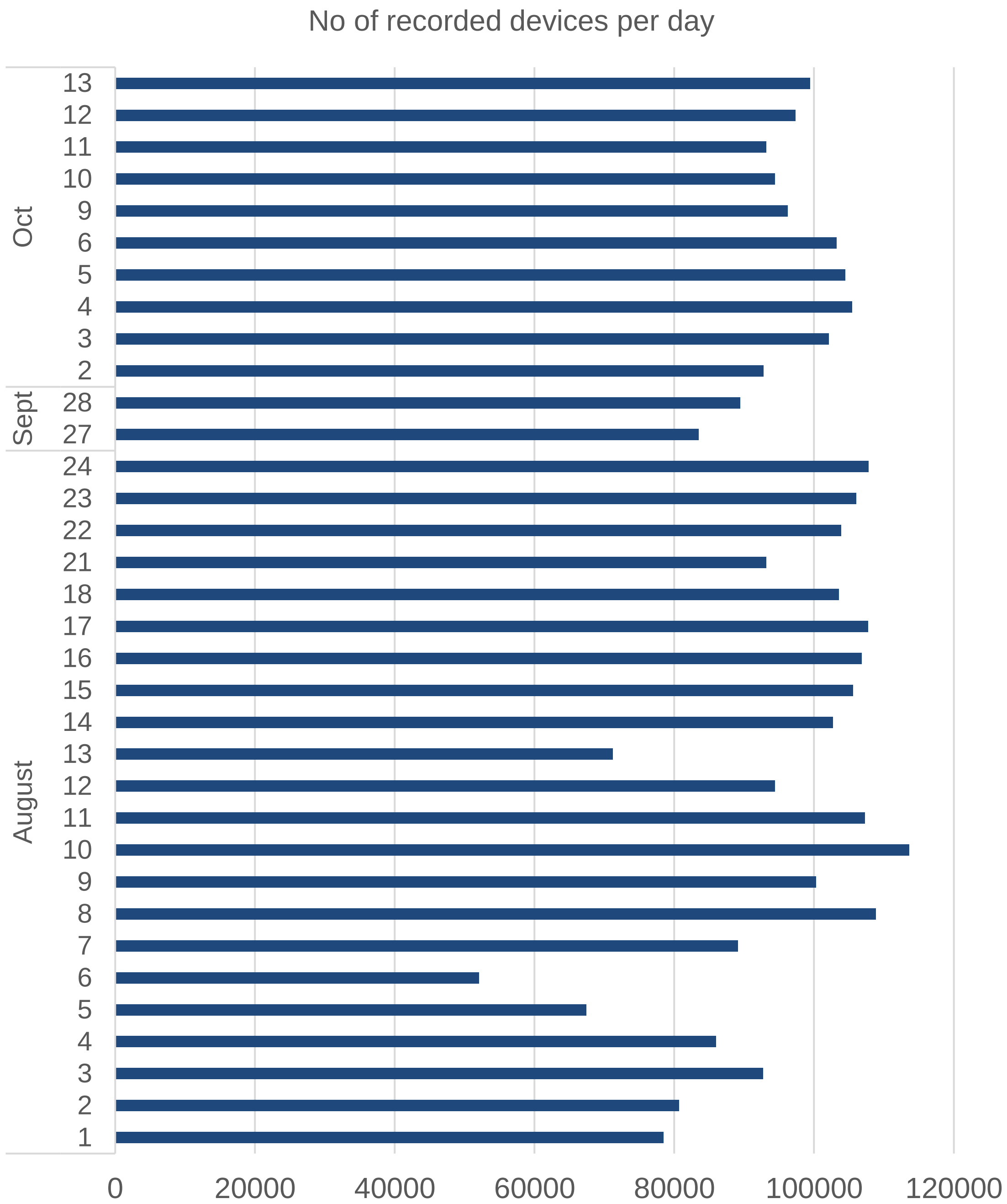


Figure 5-2: Sample size: Number of devices recorded for the 22 days (all data per day breakdown)

5.5.2 Methodological framework

Location data were assessed using a data analysis methodology framework developed based on the data collected (Figure 5-3). Data pre-processing and preparation is an integral step in ML algorithms as the quality of the data and any useful information that can be derived from the analysis directly affect the model's ability to learn. The chosen ML type is K-means clustering and is further discussed in the 5.5.2.2 section. The analysis was performed on a daily resolution, reflecting the pedestrian movement theory and literature findings. The analysis was performed in three distinct steps, as described below:

i. **Data pre-processing:** Data pre-processing utilising bespoke algorithms and space syntax methodologies were used to extract valuable information and to enrich existing datasets. Walking pedestrian characteristics, spatial visibility, and weather information were mapped against individual points recorded via Wi-Fi tracking technique, described in detail in the following sections.

ii. **Data preparation for K-means clustering analysis and Model Development:**

   a) Data cleaning and normalisation using outlier removal (Interquartile Range Method) (Vinutha, et al., 2018) and Min-Max scaler method (Han, et al., 2012).

   b) Multiple factor analysis to remove multi-dimensionality of the data and to select key variables for the model.

iii. **Analysis & Results:** Cluster analysis (unsupervised machine learning) to extract key behavioural patterns and identify classes of homogeneous profiles.

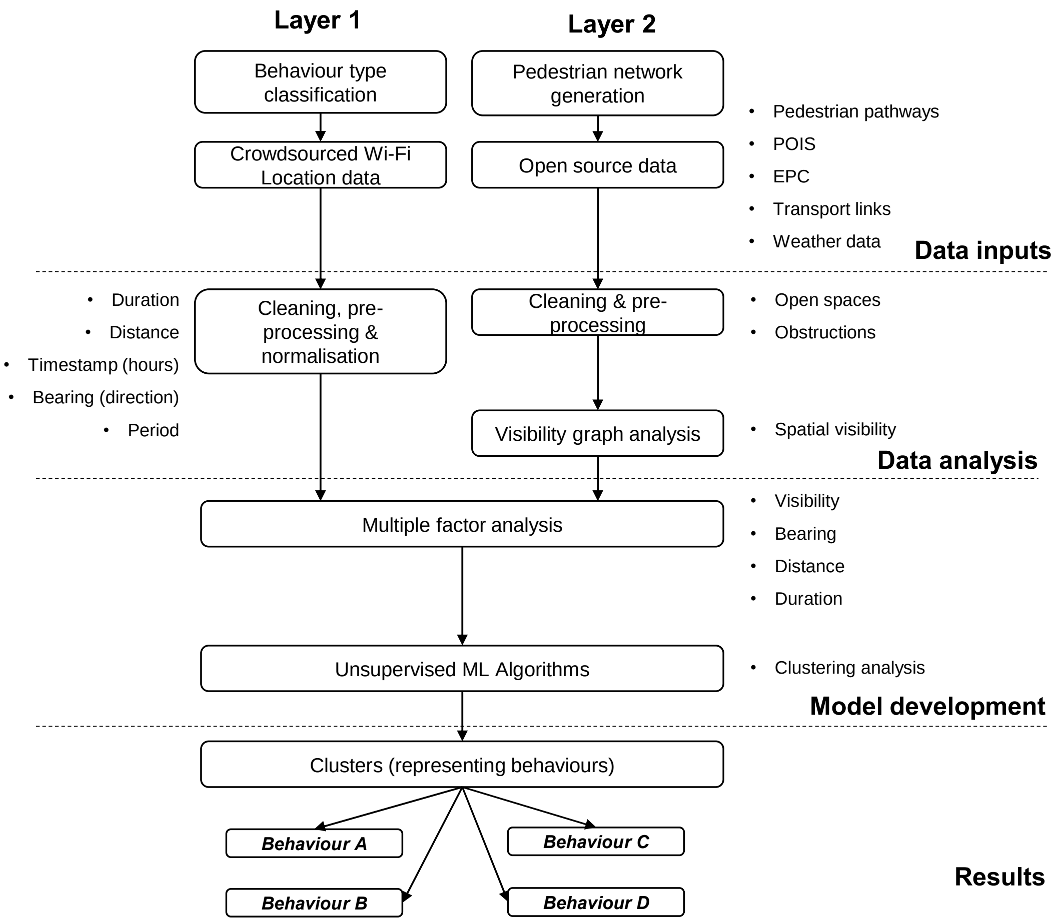


Figure 5-3: Overview of proposed methodology framework

Finally, based on the similarities returned by the clustering analysis, a data mapping exercise against these categories was undertaken to investigate three key hypotheses: (i) recorded speed and purpose present a clear and consistent relationship, (ii) visibility as a driver for movement and (iii) knowledge of the area based on number of unique recorded devices, representing repeated visits. The level of experience within the area was calculated by identifying and counting instances with only one visit recorded through the number of days of the recorded dataset across all the days, with the assumption that this implied non-regular use of the area.

5.5.2.1 Data Pre-processing

The variables extracted are (i) Duration in seconds, (ii) Distance in metres, (iii) Speed in m/s, (iv) Bearing in degrees, and (v) Day period (Table 5-3). A bespoke algorithm written in Python (Van Rossum & Drake Jr, 1995) was used to extract the variables calculated on a point-by-point basis. This method reflected accurate movement patterns from one recorded location to the other, as each location is tracked on an XYZ point. The point-by-point method follows the pattern of subtracting the n+1 point from the n point to extract the absolute values, where n is the first recorded location of the device within the study area. To remove false recordings, a threshold for outlier distances was set at 5,000m, removing all points with such recorded distances. The bearing was calculated to indicate direction. The transformation to geographical degrees considers the geographical north at 0/360°, the Python code thus calculating a bearing as N=0/ 360; E=90; S=180; W=270. Day period classification was undertaken as follows, based on August 1st sun cycle in London, U.K. which for consistency it was used for all the study days:

- 'First light': 05:24 to 08:40
- 'Morning': 08:40 to 12:10
- 'Lunchtime': 12:10 to 14:00
- 'Afternoon': 14:00 to 20:47
- 'Last light': 20:47 to 21:28
- 'Nighttime': 21:28 to 05:24

A key limitation with using such big datasets is that the processing of information required polynomial run time. The researchers had to utilise the High-Performance Computer (HPC) facility of Cranfield University to overcome this issue (Cranfield University, 2017). Sixteen CPU cores were needed, with a three-hour simulation time required per input .csv file.

Further information was acquired to understand better pedestrian movement concerning spatial attributes, using the space syntax methodology and, more specifically, the visibility graph analysis (VGA), serving as a constant for spatial visibility (Turner, et al., 2001). Space syntax theories were employed using VGA assessments via the open-source software DepthmapX_net_035 (Varoudis, 2017). An area of 2km diameter, with the circle's centre located in the middle of the Oxford Circus, was adopted as the distance threshold. This was adopted to prevent result distortion due to the small scale. According to Ahrné et al. (2009), minimum distance thresholds range from 300m to 1km radius; hence, the 2km diameter scale was chosen by the authors (Ahrné, et al., 2009). Following the VGA, a Geospatial Information Systems (GIS) platform and the function "*Extract Values to Points"* were used to map the values against each point. This step enabled spatial visibility (VGA) to be included in each recorded location from the Wi-Fi data, serving as an additional analysis parameter.

Finally, weather data were exported from a private weather station located in the area for each date. Weather information was recorded every five minutes, and the most appropriate fields were selected to reflect the microclimate conditions in the study area. The parameters comprised humidity, wind speed, precipitation, solar radiation, and temperature. All weather information was mapped against the location dataset, using Python's bisect method (array bisection algorithm). The complete set of variable inputs compiled is displayed in Table 5-3.

Table 5-3: Variable set used

Variable name	Data type	Description	Year collected	Source
ID	Categorical	MAC address	2017	Wi-Fi tracking
end_lon	float	X coordinate	2017	Wi-Fi tracking
end_lat	float	Y coordinate	2017	Wi-Fi tracking
Period	Categorical	Name of the assigned period	2017	Wi-Fi tracking
end_time	Categorical	Date and time	2017	Wi-Fi tracking
bearing_segment	float	Bearing in degrees	2017	Wi-Fi tracking
duration_segment	Integer	Time spent in seconds	2017	Wi-Fi tracking
distance_segment	float	Distance travelled in metres	2017	Wi-Fi tracking
speed_segment	float	Walking speed in m/s	2017	Wi-Fi tracking
rvalue_1	float	Spatial visibility	2018 (Building polygons used for VGA simulation)	https://osmaps.ordnancesurvey.co.uk/ https://www.openstreetmap.org, https://www.geofabrik.de/geofabrik/)
Humidity_%	float	Humidity in percentages	2017	Weather Underground. Weather Station ID: ILONDON636
Speed_mph	float	Wind speed in mph	2017	Weather Underground. Weather Station ID: ILONDON636
Precip. Rate _in	float	Precipitation rate in inches	2017	Weather Underground. Weather Station ID: ILONDON636
Solar_w/m^2	float	Solar radiation in w.m2	2017	Weather Underground. Weather Station ID: ILONDON636
Temperature_C	float	Temperature in Celsius	2017	Weather Underground. Weather Station ID: ILONDON636
hours	float	Hour of the day	2017	Wi-Fi tracking

5.5.2.2 Data preparation for K-means analysis and Model Development

Cluster analysis was utilised to reveal walking behaviours and identify key groups in the case study area. ML pattern-mining techniques are generally used to identify unknown patterns within normalised datasets (Abu-Bakar, et al., 2021). Cluster analysis labels observations (data points) within assigned groups or ‘clusters’, extracting key patterns and identifying classes of homogeneous profiles. K-means algorithms partition data into clusters by minimising the within-clusters sum-of-squares (Yuan & Yang, 2019) (Equation 5-1).

Equation 5-1: K-means within-clusters sum-of-squares

$$\boldsymbol{J} = \sum_{\boldsymbol{k}=\boldsymbol{1}}^{\boldsymbol{k}} \sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{n}} \left\| (\boldsymbol{x_i} - \boldsymbol{\mu_k})^{\boldsymbol{2}} \right\|$$

where J is the main function of sum of the squared error, k is the number of clusters, n is the number of observations, xi is observation i and μκ is the centroid formed for xi’s cluster. The mean of the recorded data is constantly updated, and each observation is placed within the cluster having the nearest centre until no more observations can be assigned (Forgy, 1965).

This method was chosen due to the nature of the input observations and the manner by which this method exclusively segregates clusters, so as each point belongs to one group only, where each partition is represented by one cluster only and k ≤ n (Han, et al., 2012; Zhu, et al., 2010). The following steps were performed to ensure data suitability for the unsupervised ML model (Celebi, et al., 2013; Zhang & Leung, 2003; Namratha Reddy & Supreethi, 2017), namely: (1) Input variables limited to numerical only, (2) Noise & outlier removal, (3) Data normalisation, (4) Reduction of the number of variables, (5) Collinearity, (6) Determining the optimal number of clusters. Each is discussed below.

(1) Input variables limited to numerical only

K-means uses distance-based measurements to determine the similarity between data points; therefore, numerical variables are the only input processed. Additionally, undefined (NaN) values were removed following the first .csv output from the raw data analysis and the weather mapping values. NaN values cannot be considered 0, as the 0 value is meaningful in this type of analysis. Nevertheless, the NaN values for the weather mapping exercise were less than 10%, an acceptable percentage when dealing with missing values (Bennett, D.A., 2001). Listwise deletion (LDel) was used to remove all the NaN values (Peng, et al., 2006).

(2) Noise & outlier removal

K-means is sensitive to outliers and 'noisy' data (Jin X., Han J., 2011). If data is not pre-processed to remove noise and outliers, K-means can return false results, driven by the most substantial information set. Outlier values were identified and removed using the interquartile range method for each variable in Python (Vinutha, et al., 2018; Thorndike, 1953).

3) Data normalisation

For the ML algorithm to consider all attributes as equal, they must all have the same scale; hence the Min-Max Scaler method was used, implemented via Python, and the help of scikit-learn package (Han, et al., 2012). This method was chosen as it transforms each value in the columns proportionally, within the bounded intervals, achieving a linear transformation on the original data (Abu-Bakar, et al., 2021). The Min-Max Scaler method is considered ideal for revealing patterns by highlighting any peaks or falls in a consistent manner (Abu-Bakar, et al., 2021).

Each variable was transformed by scaling to the range 0-1 (Equation 5-2).

Equation 5-2: Min-Max Scaler

$$z = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Where z is the normalised value, xi is the original value range, min(x) is the minimum range attribute and max(x) is the maximum attribute range. This step ensured that different scales would not skew the results and contribute equally to model fitting.

(4) Reduction number of variables

As the number of variables increases, a distance-based similarity measure converges to a constant value between any given points. The more variables, the more challenging to find strict differences between instances. One of the most popular approaches to dimensionality reduction is principal component analysis (PCA). This method seeks to reduce the number of variables while preserving the input data's most important structure and relationships. It has also been shown to produce improved results when the dimensionality of datasets is high by comparison with other methods (Reddy, et al., 2020).

The explained variance ratio metric was used for the first two components to define the number of variables. The explained variance ratio is the percentage of variance attributed by each selected component. Ideally, the number of components chosen for model inclusion is decided by adding the explained variance ratio of each component until a total of around 0.8 or 80% is attained to avoid overfitting (Joliffe & Morgan, 1992). The variance should not be less than 60% (Hair, et al., 2014). The variance explained is 35%, indicating the data is not useful and may need further measures. If the variance is less than 60%, there are most likely chances of more variables that can be apparent. The explained variance ratio returned with all the variables was 45%. Therefore, the additional parameters, as indicated in Table 5-3, were removed, resulting in the selection of only the variables of (1) visibility, (2) bearing, (3) distance, (4) duration (Table 5-3); the explained variance ratio was then recalculated as 76%.

(5) Collinearity problem

Correlated variables are not useful for ML segmentation algorithms, representing the same characteristic of a segment (e.g., noise). Speed was tested with PCA analysis as a suitable variable but was removed from the final list of input variables due to a high level of correlation with duration and distance (Reddy, et al., 2020).

(6) Determining the optimal number of clusters

Clustering algorithms rely on random initialisation of the cluster centre. To overcome this issue, the Elbow Method (EM) (Thorndike, 1953) and Silhouette Analysis (SA) (Rousseeuw, 1987) were undertaken to reveal the ideal number of clusters, where a randomised seeding technique guarantees the optimal solution is obtained. The EM is one of the most popular methods used to determine the optimal value of k, using two calculation metrics: distortion and inertia (Han, et al., 2012). Distortion is the average of the squared distances from the cluster centers of the respective clusters, and typically the Euclidean distance metric is used. Inertia is the sum of squared distances of samples to their closest cluster center. K-means clusters data separate samples to n groups of equal variances, minimizing the inertia or within-cluster sum-of-squares (WCSS) criterion. Therefore, the smaller the inertia, the denser the cluster (the closer the points are). The Silhouette Score ranges from -1 to 1 indicating how close or distant the clusters are from each other and how dense the clusters are.

5.6 Analysis and results

The analysis from the EM and SA method application identified the number of clusters within the data. The optimum number of clusters was returned as k=4 (Figure 5-4), indicating four different groups of behaviours. This test was performed initially using the daily resolution; however, it was performed again in both period and hourly resolutions, all returning the same results. Examples of the returned values are shown in the graphs below for the 5th of August 2017 (Figure 5-4). The analysis was performed in Python with the help of the scikit-learn package (Scikit-learn 0.19.1 documentation, 2018).

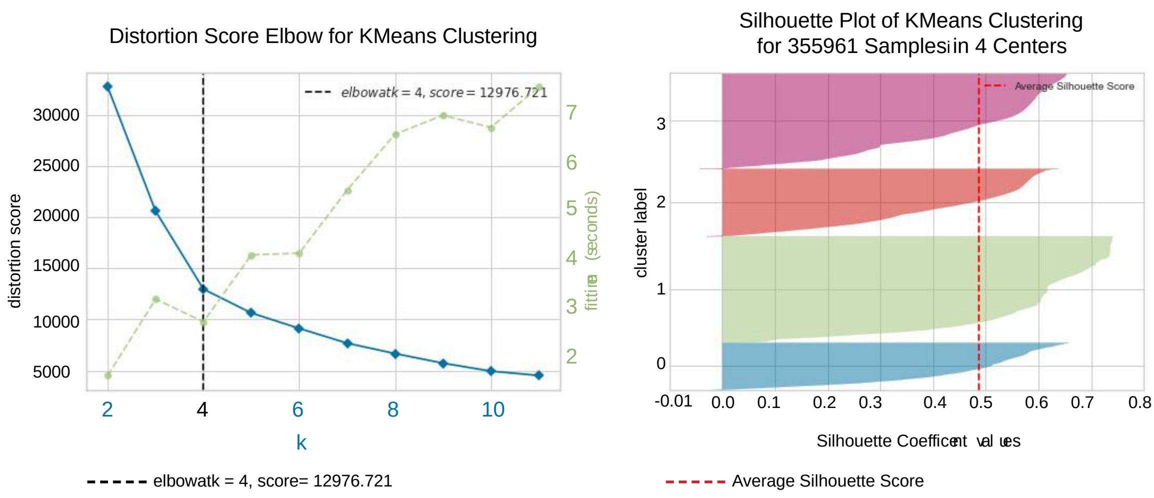


Figure 5-4: Elbow (left) and Silhouette Analysis (right) results example on a typical day in August (5th). Results indicate number of clusters k=4.

Cluster analysis was performed with k=4 (maximum iteration = 100 and random state = none) as indicated by the EM and SA, for each day of the overall dataset, due to the difference in patterns and behaviours occurring at different times of the year and weekdays. The input data were classified using the clustering algorithm and organised into classes sharing similar attributes (Figure 5-5).

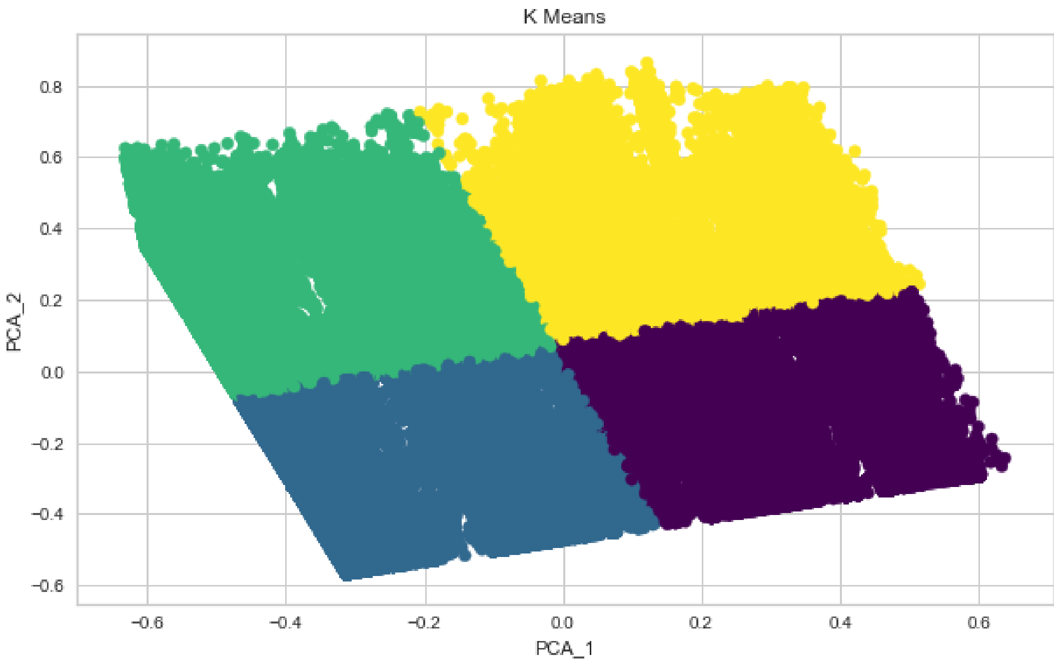


Figure 5-5: Cluster results example (5th of August 2017). Each colour represents a different cluster

The clustering steps were repeated for key days to validate the previous analysis. Results revealed that the same number of clusters exist, via the use of EM, followed by acceptable values in SA (Figure 5-6).

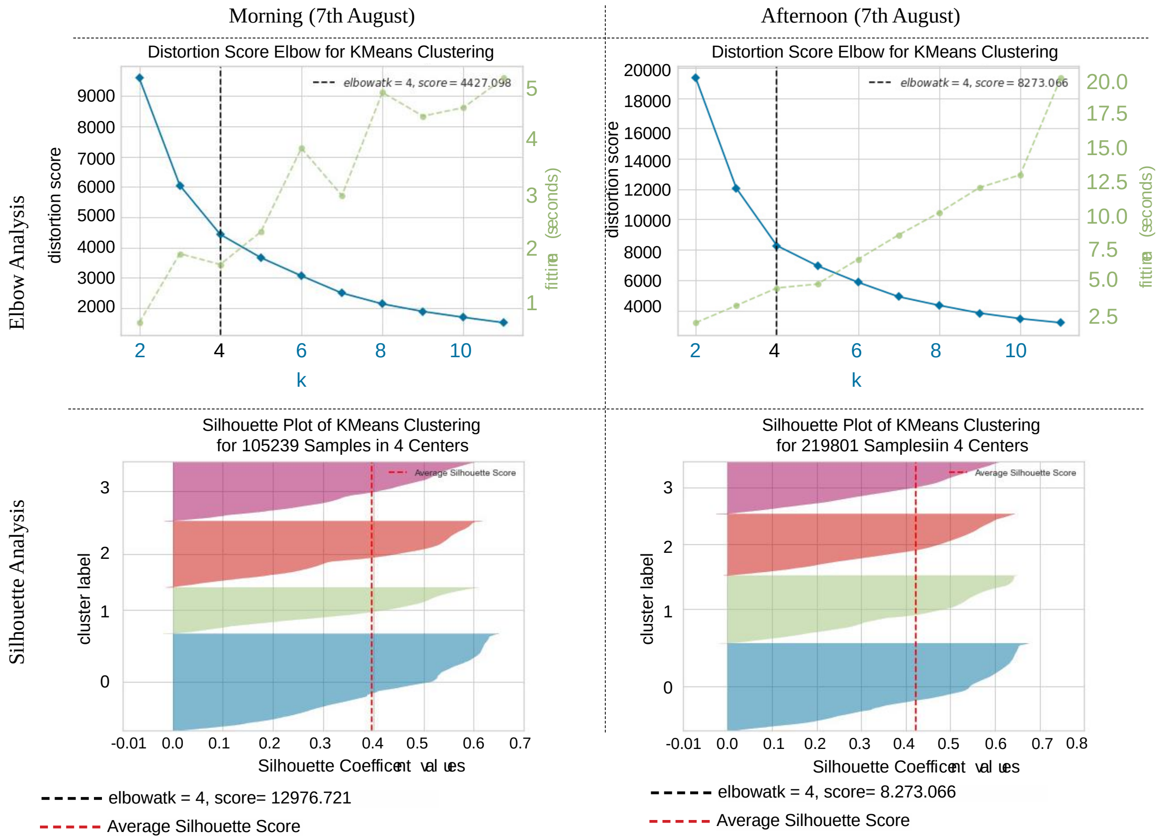


Figure 5-6: Analysis results in period resolution in a typical weekday (EM and SA methods)

An additional metric was selected to validate further the number of optimum clusters, the Calinski-Harabasz coefficient (CH), and analysis for this was undertaken for both resolutions (daily & period) (Figure 5-6). The CH, also known as the variance ratio criterion, is a measure based on the internal dispersion of clusters and the dispersion between clusters. The CH criterion was chosen as a validation method as it is fast to compute, considered appropriate due to the significant amounts of data in this study, and is established as one of the best-performing methods for estimating numbers of clusters (Milligan & Cooper, 1985). In addition, this metric is suitable when dealing with compact clusters (convex) (Calinski & Harabasz, 1974). The optimum number of clusters is the number that maximises the CH value (Calinski & Harabasz, 1974) without overfitting the model. As there is not a recognised threshold, the optimum number of clusters is chosen when a peak or an elbow is on the line plot of CH indices. The analysis was performed in Python with the help of the scikit-learn package, returning the optimum number of clusters as k= 4 (Figure 5-7).

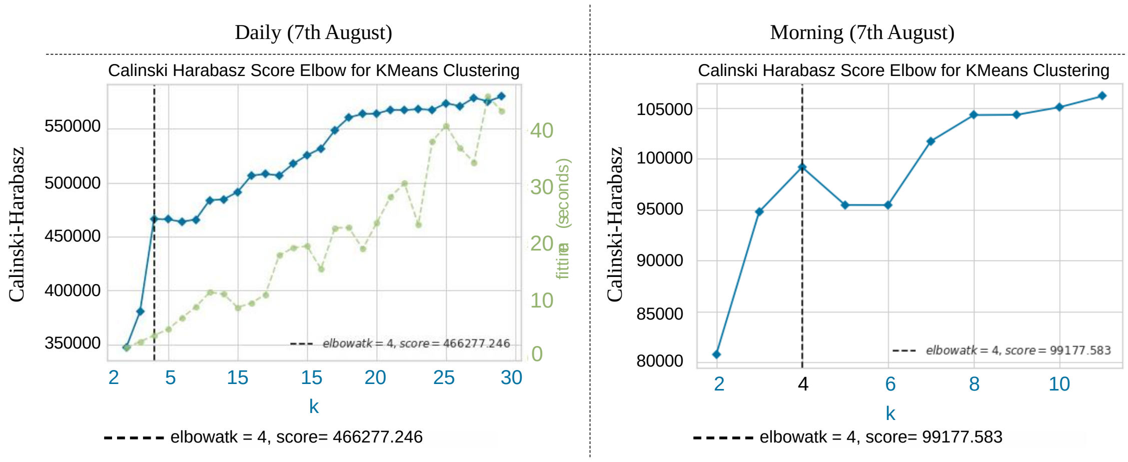


Figure 5-7 Analysis results examples in daily & period (morning) resolution (Calinski-Harabasz coefficient)

Following cluster analysis, feature importance extraction was performed in Python to understand clustering results drivers better. Each feature receives a score indicating that the higher the value, the more important or relevant it is towards the output variable. The results revealed that the main driver order for the clustering results was as follows: (i) spatial visibility (rvalue_1), (ii) route direction (bearing_segment), (iii) distance travelled (distance_segment), and (iv) time spent in point (duration_segment). Figure 5-8 shows ranking importance displayed in the correct order in the graph’s legend, named “variable”.

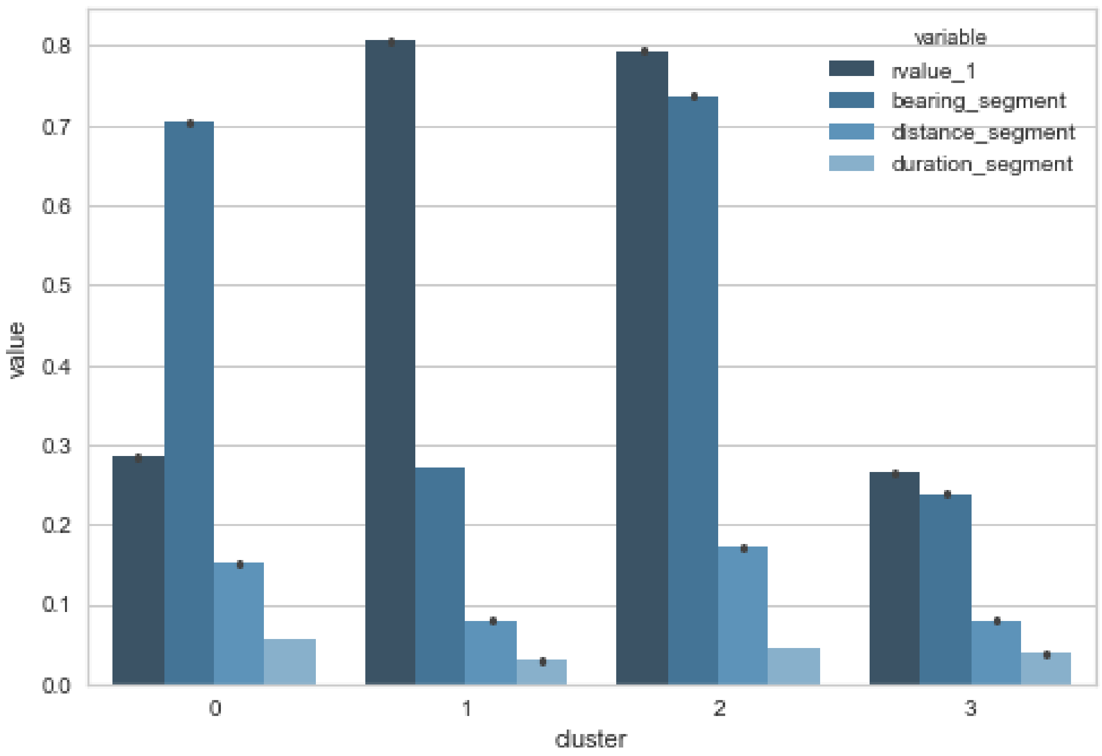


Figure 5-8 Point-based feature extraction per cluster and importance ranking of variables (example 7th August)

Clustering results revealed four distinct clusters for all the individual days assessed. Extraction of the key characteristics for each cluster was undertaken, and a summary of results was identified (Table 5-4). A descriptive summary of the results included count, mean, standard deviation, minimum and maximum values, plus lower and upper percentiles and the 50th percentile (median). Counting revealed cluster results to be balanced, with the sample sizes similar in all categories (Amin, et al., 2016). Results illustrate that Cluster 0 and 2 users generally travel longer distances, with the mean distance being 14.06 and 12.16m respectively between recorded points, whereas Clusters 1 and 3 users travelled 5.78 and 7.11m. Clusters 0 and 3 spent time in high visibility areas, with mean rvalue_1 recording being 13,784.69 and 13,988.06, respectively. The shortest time spent at each point segment was observed in Cluster 1, with a mean value of 31.15 seconds, while the highest was in Cluster 0, with a mean value of 66.74 seconds.

Table 5-4: Descriptive statistics of the clustering results (August 2nd to 18th)

Cluster 0	bearing_segment	duration_segment	distance_segment	speed_segment	rvalue_1	cluster
count	1,932,972.00	1,932,972.00	1,932,972.00	1,932,972.00	1,932,972.00	1,932,972.00
mean	235.55	66.74	14.09	0.40	13784.69	0.00
std	54.93	146.15	19.41	0.52	6762.04	0.00
min	76.09	5.00	0.00	0.00	136.15	0.00
25%ile	198.43	5.00	1.00	0.06	7,993.95	0.00
50%ile	216.03	15.00	4.00	0.18	14,710.07	0.00
75%ile	286.80	55.00	21.00	0.52	18,345.94	0.00
max	341.57	1,370.00	87.00	2.69	33,442.28	0.00

Cluster 1	bearing_segment	duration_segment	distance_segment	speed_segment	rvalue_1	cluster
count	2,547,052.00	2,547,052.00	2,547,052.00	2,547,052.00	2,547,052.00	2,547,052.00
mean	92.31	31.15	5.78	0.30	3,8045.32	1.00
std	59.28	92.78	11.42	0.44	6,843.82	0.00
min	0.00	5.00	0.00	0.00	22,783.48	1.00
25%ile	39.81	5.00	0.00	0.01	34,572.00	1.00
50%ile	90.00	10.00	1.00	0.15	36,654.67	1.00
75%ile	139.50	20.00	5.00	0.36	45,839.99	1.00
max	339.06	1,370.00	87.00	2.69	48,533.89	1.00

Cluster 2	bearing_segment	duration_segment	distance_segment	speed_segment	rvalue_1	cluster
count	2,105,173.00	2,105,173.00	2,105,173.00	2,105,173.00	2,105,173.00	2,105,173.00
mean	225.38	43.49	12.16	0.40	35,218.27	2.00
std	73.08	112.71	18.80	0.49	10,114.79	0.00
min	0.00	5.00	0.00	0.00	509.05	2.00
25%ile	183.37	5.00	1.00	0.09	29,216.61	2.00
50%ile	215.91	10.00	3.00	0.20	36,372.39	2.00
75%ile	287.65	30.00	15.00	0.51	44,098.32	2.00
max	341.57	1,370.00	87.00	2.69	48,533.89	2.00

Cluster 3	bearing_segment	duration_segment	distance_segment	speed_segment	rvalue_1	cluster
count	2,726,997.00	2,726,997.00	2,726,997.00	2,726,997.00	2,726,997.00	2,726,997.00
mean	79.88	46.18	7.11	0.24	13,988.06	3.00
std	42.45	118.99	14.71	0.44	8,600.58	0.00
min	0.00	5.00	0.00	0.00	136.15	3.00
25%ile	35.54	5.00	0.00	0.00	7,403.70	3.00
50%ile	90.00	10.00	1.00	0.05	14,307.09	3.00
75%ile	98.39	25.00	5.00	0.25	18,145.74	3.00
max	192.80	1,370.00	87.00	2.69	48,212.39	3.00

5.6.1 Hypothesis 1: Recorded speed and purpose

Cluster 0 and 2 have the highest recorded walking speeds in the August period, while Cluster 1 and 3 have the lowest (Figure 5-9– showing walking speed heatmaps), indicating that at this time, pedestrians' goals for Cluster 0 and 2 users were better defined than those in the other two clusters. The lowest speed was recorded for Cluster 3, with an average value of 0,24 m/s in August and 0,23 in October, while Cluster 2 has the highest speeds recorded in August with a mean value of 0,42 m/s. However, as this research explores walking patterns on a point-by-point basis rather than on a journey purpose basis, these types of classifications can only serve as baseline information.

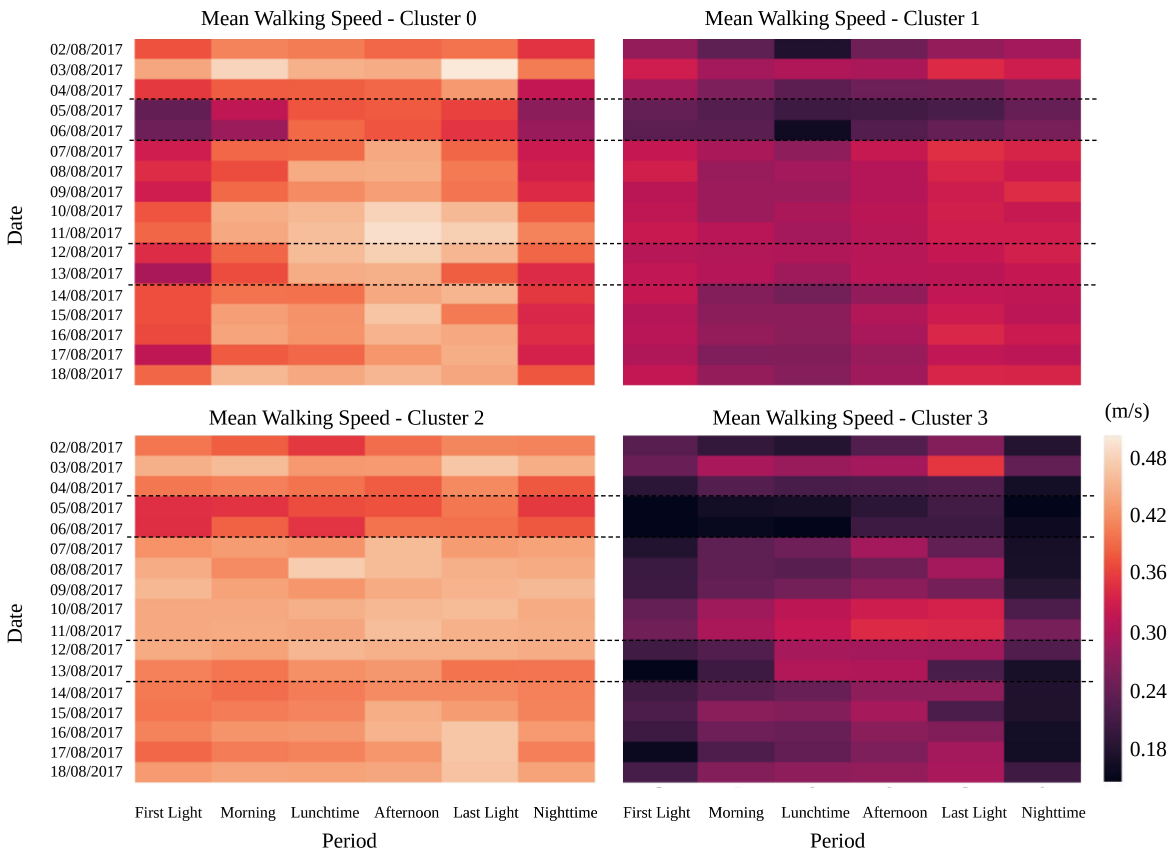


Figure 5-9: Heatmaps illustrating mean walking speeds in m/s as recorded for each cluster in August. The cluster number is indicated on the top of each graph. Dates are shown on the vertical axis, for dates 2nd of August (top) until 18th of August (bottom), while period is indicated on the horizontal axis, dashed lines indicating weekends

5.6.2 Hypothesis 2: Visibility as a driver for movement

Clusters 0 and 3 were revealed to have the lowest mean visibility values, indicating that user movement was within areas with limited space visibility (Figure 5-10). Therefore, the results indicate Cluster 0 and 3 as having behaviours that appear when an individual has an idea of where places are or has increased knowledge of the area, indicating that visibility is not a crucial parameter for their movement. Other needs, such as route efficiency, can better define walking patterns.

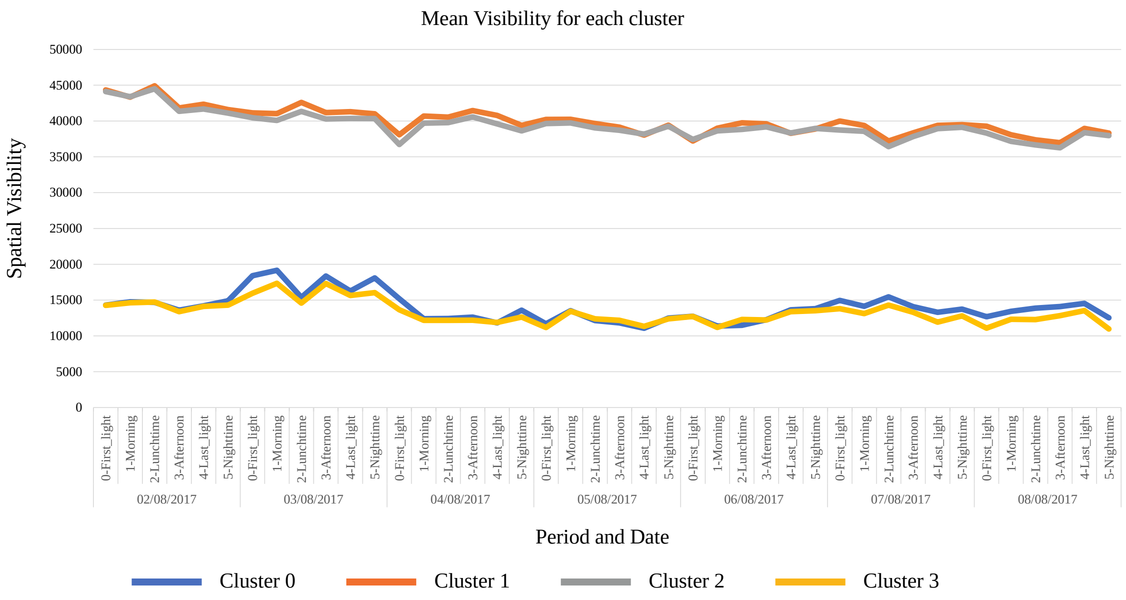


Figure 5-10 Line graph illustrating mean space visibility for each cluster in a typical week in August (2nd to 8th). Mean values are grouped by period with colours representing clusters.

5.6.3 Hypothesis 3: Knowledge of the area based on number of unique recorded devices

Results indicated that Cluster 1 and 2 have the highest percentage of unique devices recorded, from first to last light (05:24 to 21:28). This finding suggests that most of the points recorded in these clusters are potentially users with only limited knowledge of the area (Figure 5-11). Figure 5-11 shows each cluster's recorded devices in a typical week in August. This graph indicates an increase in the number of unique users during weekends, while there is a considerable decline of unique users on a Monday (7th of August, a UK Summer Day). Cluster 3 indicates that most of the unique users visit the area during weekends. However, due to the lack of weekend days (only two weekends were incorporated within August due to data source restrictions), these devices are highlighted as unique within the existing dataset.

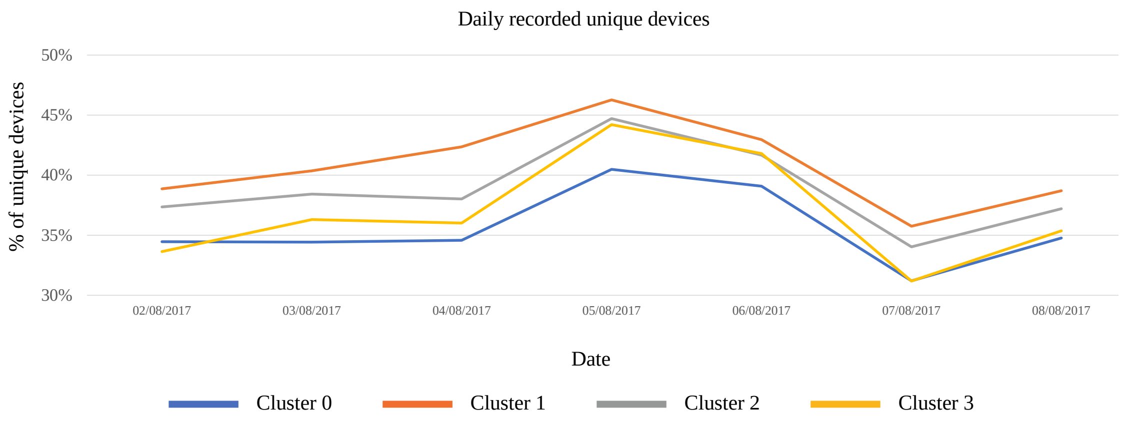


Figure 5-11: Unique devices recorded for each cluster in a typical week in August. Values are daily and count devices from first to last light (05:24 to 21:28). Colours represent clusters. 5th and 6th of August are Saturday and Sunday respectively.

5.7 Discussion

This paper broadens the understanding of the complex, multi-layered relationships occurring between urban space and pedestrians via the combination of quantitative location data pertaining to end-users, as well as quantitative spatial attributes, such as visibility. This novel analytical approach provides a new understanding of behaviours exhibited in an urban space while implying that although spatial configuration influences walking behaviour, such information crucially affects the quality of the walking experience and individual preferences. The ‘Big Data’ approach adopted further permits automation to be applied to the analysis, making the approach suitable for large-scale area analysis.

In this paper, an unsupervised ML framework methodology is proposed for identifying the pedestrians' intention while walking in urban spaces. Different contexts can reveal different behaviours, as urban environments vary significantly. The results illustrate how the employment of unsupervised ML algorithms can reveal categories of pedestrian behaviours that can directly inform urban models, rather than depending on biased observational data or existing literature.

It is suggested that the combination of the type of activity undertaken, and the knowledge of the area represent the key behaviours found in walking patterns in the context of a high-street. These can be grouped into two activity categories: (i) Utilitarian walking (with motivation/ destination) and (ii) Leisure walking (no motivation) (Ki & Lee, 2021). The first category includes learning and journey efficiency, where the user has specific destinations to reach with varying levels of area knowledge (expert and novice strider (Davies, 2007)). The second category includes wandering and open-ended journeys, where travellers explore already known areas of the city or discover them as they walk (expert and novice stroller (Davies, 2007)). Further to this, results indicated a consistent link between spatial visibility and previous experience of the area. The greater the familiarity a pedestrian has with the surrounding space, the less spatial visibility is needed for its pedestrian movement (Figure 5-12).

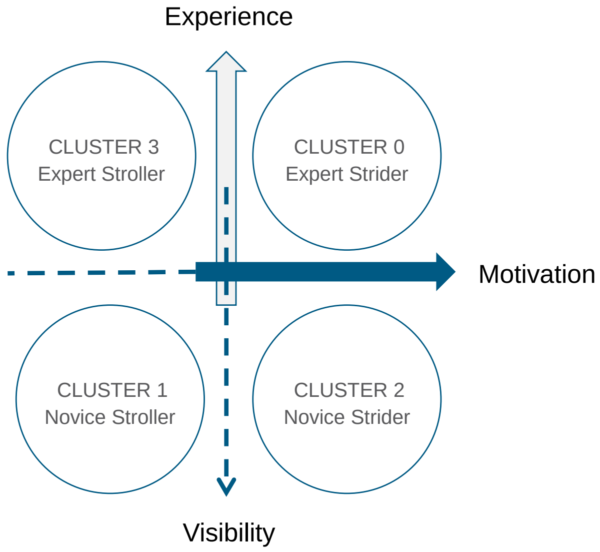


Figure 5-12: Clustering results against key behaviours identified.

Validation of area knowledge was undertaken utilising two hypotheses: (i) visibility as a driver for movement and (ii) knowledge of the area based on the number of unique recorded devices. It is hypothesised that the reason why a significant percentage of unique recorded devices falls within the rest of the clusters is due to the point-by-point analysis was employed, indicating that behaviours change as someone moves within the space. Distractions and content divert pedestrians' plans to minimise distance and effort and vice versa (Al-Widyan, et al., 2017). The exploratory journey may become goal-oriented when the awakening of interest occurs, aiming to reach the specific destination-area of interest. Similarly, the walking behaviour may be altered to an exploratory journey, evoked by the emotional qualities of an environment. Furthermore, once the individual achieves a visual connection with their final destination, the level of area familiarity can change from no knowledge to increased knowledge, as it enhances a faster acquisition of the cognitive map of the destination (legibility) (Zacharias, 2001).

Employment of BDAs and, more specifically, unsupervised ML algorithms in the context of walking activities in urban space present several challenges. These challenges include the selection of suitable datasets and variables, and to overcome these, domain knowledge is required. Pedestrian activity at a specific place and time is influenced by numerous factors related to the urban built environment, such as network connectivity, spatial visibility, and quality of view, introducing challenges in the types of data needed. Due to the high expense of primary data collection, data sources must be chosen carefully to ensure the study question will be addressed.

5.8 Conclusion

This study presents a novel means to assess pedestrian routing in urban environments. The authors contribute to knowledge by providing a framework methodology to classify behaviours and spatial configuration interactions, utilising machine learning algorithms and location data, as derived from Wi-Fi tracking techniques. The results indicate that the framework methodology developed provides a consistent method for identifying pedestrian categories across various periods, and it can be directly applied to urban design approaches. By investigating walking patterns in conjunction with spatial attributes, such as spatial visibility, insights are revealed concerning individual preferences and behaviours of end-users and utilisation of the urban space. The present study adds to the wayfinding and transportation existing literature by creating a methodological framework which can be utilised for observations in large-scale contexts, where traditional approaches currently fail.

Implications/ limitations of this study include the fact that data collection was not continuous due to technical issues with missing entries for some of the days. This study's collection of pedestrian movement was restricted to a specific number of days, spanning from August to October, introducing further limitations relative to neglected spatio-temporal aspects of human behaviour. An additional limitation was the low resolution of input data that can introduce significant errors in a streetscape when assessing walking behaviours in the context of the micro level design. Furthermore, the employment of unsupervised ML algorithms in the context of walking activities in urban space presents limitations, such as model overfitting, where a model represents noise or random errors rather than revealing actual patterns inherent in data. ML algorithms provide improved performance with more data. However, increasing model complexity can result in deterioration in performance.

Future research can focus on the analysis of additional datasets or the collation of further real-time data that could enable investigation of variations from exogenous parameters, such as weather conditions or one-time incidents, leading to increased robustness and more accurate predictions. Moreover, exploration of data collection techniques is needed to compare such methods related to pedestrian movement tracking, such as GPS devices and video recordings. Such approaches can achieve their full potential if integrated into the urban design process, requiring significant changes in how design is implemented. Hence fundamental changes are necessary in educational and behavioural contexts. Finally, a better understanding of the key characteristics of the identified behaviours (clusters) and the spatial attributes that encourage or discourage their presence is needed. Comparing speed profiles, occupancy patterns, and weather information can provide beneficial insights. Such frameworks could improve the post-COVID re-evaluation of urban spaces, especially in similar contexts where retail activities decline. Most likely, one of the key priorities for urban city centres, as these recover, would be their economic development. Nevertheless, as the urban geometry of cities varies significantly, one common approach is not easy to be adopted. Therefore, context-specific approaches, such as the framework methodology proposed in this study, are required to enable effective planning and response.

REFERENCES

Abu-Bakar, H., Williams, L. & Hallett, S., 2021. Quantifying the impact of the COVID-19 lockdown on household water consumption patterns in England. *npj Clean Water,* 4(13).

Accuware Inc, 2017. *Accuware.* [Online]
Available at: https://accuware-inc.com/
[Accessed 19 08 2021].

Ahrné, K., Bengtsson, J. & Elmqvist, T., 2009. Bumble bees (Bombus spp) along a gradient of increasing urbanization. *PLoS ONE,* 4(5): e5574.

Al-Widyan, F., Al-Ani, A., Kirchner, N. & Zeibots, M., 2017. An effort-based evaluation of pedestrian route choice.. *Scientific Research and Essays,* 12(4), pp. 42-50.

Amin, A., Anwar, S., Adnan, A., Nawaz, M., Howard, N., Qadir, J., Hawalah, A. & Hussain, A., 2016. Comparing Oversampling Techniques to Handle the Class Imbalance Problem: A Customer Churn Prediction Case Study. *IEEE Access,* Volume 4, pp. 7940-7957.

Andersson, E., Barthel, S. & Ahrné, K., 2007. Measuring social–ecological dynamics behind the generation of ecosystem services. *Ecological Applications,* Volume 17, pp. 1267-1278.

Angelelli, F., Morrow, J. & Greenwood, C., 2018. *The potential application of Wi-Fi data in the development of agent based pedestrian models.* Dublin, Ireland, The Association for European Transport.

Aschwanden, G., Wijnands, J.S., Thompson, J., Nice, K.A., Zhao, H. & Stevenson, M., 2019. Learning to walk: Modeling transportation mode choice distribution through neural networks. *Environment and Planning B: Urban Analytics and City Science,* 48(1), pp. 186-199.

Askarizad, R. & Safari, H., 2020. The influence of social interactions on the behavioral patterns of the people in urban spaces (case study: The pedestrian zone of Rasht Municipality Square, Iran). *Cities,* Volume 101, Article 102687.

Batty, M., 2018. Digital twins. *Environment and Planning B: Urban Analytics and City Science,* 45(5), pp. 817-820.

Batty, M., Jiang, B. & Thurstain-Goodwin, M., 1998. *Local movement: agent-based models of pedestrian flows.,* London, UK: Centre for Advanced Spatial Analysis (UCL).

Bennett, D.A., 2001. How can I deal with missing data in my study?. *Australian and New Zealand Journal of Public Health,* 25(5), p. 464–469.

Bitgood, S., 2010. An Analysis of Visitor Circulation: Movement Patterns and the General Value Principle. *Curator: The Museum Journal,* 49(4), pp. 463-475.

Boniface, S., Scantlebury, R., Watkins, S. & Mindell, J., 2015. Health implications of transport: Evidence of effects of transport on social interactions. *Journal of Transport & Health,* 2(3), pp. 441-446.

Bozovic, T., Hinckson, E. & Smith, M., 2020. Why do people walk? role of the built environment and state of development of a social model of walkability. *Travel Behaviour and Society,* Volume 20, pp. 181-191.

Brown, G., Schebella, M. & Weber, D., 2014. Using participatory GIS to measure physical activity and urban park benefits. *Landscape and Urban Planning,* Volume 121, pp. 34-44.

Brum-Bastos, V., Long, J. & Demšar, U., 2018. Weather effects on human mobility: a study using multi-channel sequence analysis. *Computers, Environment and Urban Systems,* Volume 71, pp. 131-152.

Calinski, T. & Harabasz, J., 1974. A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods,* 3(1), pp. 1-27.

Capitanio, M., 2019. Attractive streetscape making pedestrians walk longer routes: The case of Kunitachi in Tokyo. *Journal of Architecture and Urbanism,* 43(2), pp. 131-137.

Carlson, L., Hölscher, C., Shipley, T. & Dalton, R., 2010. Getting Lost in Buildings. *Current Directions in Psychological Science,* 19(5), pp. 284-289.

Celebi, M., Kingravi, H. & Vela, P., 2013. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Systems with Applications,* 40(1), pp. 200-210.

Chang, H.-H. & Chen, S., 2009. Consumer perception of interface quality, security, and loyalty in electronic commerce. *Information & Management,* 46(7), pp. 411-417.

Chang, X., Wu, J., He, Z., Li, D., Sun, H. & Wang, W., 2020. Understanding user's travel behavior and city region functions from station-free shared bike usage data. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 72, pp. 81-95.

Choi, E., 2012. *Walkability as an Urban Design Problem: Understanding the activity of walking in the urban environment (Licentiate dissertation).* [Online] Available at: Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-102182
[Accessed 2021].

Choi, E., 2014. Walkability and the complexity of walking behavior. *A/Z ITU Journal of the Faculty of Architecture,* 11(2), pp. 87-99.

Clifton, K., Smith, A. & Rodriguez, D., 2007. The development and testing of an audit for the pedestrian environment.. *Landscape and Urban Planning,* 80(1-2), pp. 95-110.

Consolvo, S., Everitt, K., Smith, I. & Landay, J., 2006. *Design requirements for technologies that encourage physical activity.* Chicago, USA, Association for Computing Machinery, p. 457–466.

Cornell, E., Sorenson, A. & Mio, T., 2003. Human Sense of Direction and Wayfinding.. *Annals of the Association of American Geographers,* 93(2), p. 399–425.

Cranfield University, 2017. *Supercomputer powers up at Cranfield University.*
[Online]
Available at: https://www.cranfield.ac.uk/press/news-2017/supercomputer-powers-up-at-cranfield-university
[Accessed 04 06 2021].

Davies, J., 2007. *Yellow Book: A prototype wayfinding system for London,* London: Transport for London by Applied Information Group.

De Arruda Campos, M., 1997. *Strategic space: patterns of use in public squares of the city of London.* London, Space Syntax Laboratory Bartlett School of Graduate Studies, University College London.

De Montigny, L., Ling, R. & Zacharias, J., 2012. The effects of weather on walking rates in nine cities. *Environment and Behavior,* 44(6), pp. 821-840.

Deluka-Tibljaš, A., Šurdonja, S., Ištoka Otković, I. & Campisi, T., 2022. Child-Pedestrian Traffic Safety at Crosswalks—Literature Review.. *Sustainability,* 14(3), p. 1142.

Díaz-Álvarez, A., Clavijo, M., Jiménez, F., Talavera, E. & Serradilla, F., 2018. Modelling the human lane-change execution behaviour through Multilayer Perceptrons and Convolutional Neural Networks. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 56, pp. 134-148.

Dridi, M., 2015. Simulation of High-Density Pedestrian Flow: A Microscopic Model. *Open Journal of Modelling and Simulation,* Volume 3, pp. 81-95.

Duives, D., Wang, G. & Kim, J., 2019. Forecasting pedestrian movements using recurrent neural networks: An application of crowd monitoring data. *Sensors,* 19(2), p. 382.

El Naqa, I. & Murphy, M., 2015. What Is Machine Learning?. In: *El Naqa I., Li R., Murphy M. (eds) Machine Learning in Radiation Oncology..* Switzerland: Springer, Cham., pp. 3-11.

European Parliament and of the Council, 2016. *General Data Protection Regulation (GDPR),* Brussels: Official Journal of the European Union.

Feng, Y., Duives, D., Daamen, W. & Hoogendoorn, S., 2021. Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment,* Volume 187, Article 107329.

Forgy, E., 1965. Cluster Analysis of Multivariate Data: Efficiency vs Interpretability of Classifications. *Biometrics,* Volume 21, pp. 768-780.

Forsyth, A., Hearst, M., Oakes, J. & Schmitz, K., 2008. Design and destinations: Factors influencing walking and total physical activity. *Urban Studies,* 45(9), p. 1973–1996.

Fuxjaeger, P. & Ruehrup, S., 2018. *Towards Privacy-Preserving Wi-Fi Monitoring for Road Traffic Analysis.* [Online]
Available at: https://www.researchgate.net/publication/305877717

Gehl, J., 2010. *Cities for People.* Washington, DC: Island Press.

Gehl, J., 2011. *Life Between Buildings: Using Public Space.* Copenhagen, Denmark: The Danish Architectural Press.

Gehl, J. & Gemzoe, L., 1996. *Public Spaces. Public Life..* Copenhagen, Denmark: The Danish Architectural Press and Royal Danish Academy of Fine Arts, School of Architectural Publishers.

Gehl, J. & Svarre, B., 2013. *How to Study Public Life.* Washington, DC: 2nd ed. Island Press.

Gibson, E., 1988. Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology,* 39(1), pp. 1-42.

Golbeck, J. & Hansen, D., 2013. A method for computing political preference among Twitter followers. *Social Networks,* Volume 36, pp. 177-184.

Gonzalez, M. C., Hidalgo, C. & Barabasi, A., 2008. Understanding individual human mobility patterns. *Nature,* Volume 453, pp. 779-782.

Gunn, L. et al., 2017. Designing healthy communities: Creating evidence on metrics for built environment features associated with walkable neighbourhood activity centres. *International Journal of Behavioral Nutrition and Physical Activity,* Volume 14, p. 164.

Hair, J. F., Black, W., Babin, B. & Anderson, R., 2014. *Multivariate data analysis.* Harlow: Pearson Education Limited.

Han, J., Kamber, M. & Pei, J., 2012. *Data Mining: Concepts and Techniques. A volume in The Morgan Kaufmann Series in Data Management Systems.* Waltham, USA: Elsevier.

Hillier, B. & Hanson, J., 1984. *The Social Logic of Space.* Cambridge: Cambridge University Press.

Hillier, B., Penn, A., Hanson, J., Garjewski, T. & Xu, J., 1993. Natural Movement: Or, Configuration and Attraction in Urban Pedestrian Movement. *Environment and Planning B: Planning and Design,* Volume 20, p. 29 – 66.

Imants, P., Theeuwes, J., Bronkhorst, A. & Martens, M., 2021. Effect of multiple traffic information sources on route choice: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 81, pp. 1-13.

Istrate, A.-L.; Bosák, V.; Nováček, A.; Slach, O., 2020. How Attractive for Walking Are the Main Streets of a Shrinking City?. *Sustainability,* 12(15), p. 6060.

Järv, O. et al., 2012. Mobile phones in a traffic flow: A geographical perspective to evening rush hour traffic analysis using call detail records. *PLoS ONE,* 7(11): e49171.

Jin X., Han J., 2011. K-Medoids Clustering. In: *Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning.* Boston, MA: Springer.

Joliffe, I. & Morgan, B., 1992. Principal component analysis and exploratory factor analysis. *Statistical Methods in Medical Research.,* 1(1), pp. 69-95.

Jones, P., Marshall, S. & Boujenko, N., 2008. Creating more people-friendly urban streets through 'link and place' street planning and design. *IATSS Research,* 32(1), pp. 14-25.

Kötteritzsch, A. & Weyers, B., 2016. Assistive Technologies for Older Adults in Urban Areas: A Literature Review. *Cognitive Computation,* 8(2), p. 299–317.

Kürkçüoğlu, E. & Akin, O., 2013. The effects of water elements in urban space perception: A case study in Üsküdar Municipality Square. *A/Z ITU Journal of the Faculty of Architecture,* 10(1), pp. 159-175.

Ki, D. & Lee, S., 2021. Analyzing the effects of Green View Index of neighborhood streets on walking time using Google Street View and deep learning. *Landscape and Urban Planning,* Volume 205, Article 103920.

Klein, R. W., Koeser, A. K., Hauer, R. J., Hansen, G. & Escobedo, F. J., 2016. Relationship between perceived and actual occupancy rates in urban settings. *Urban Forestry & Urban Greening,* Volume 19, pp. 194-201.

Kontokosta, C., Hong, B., Johnson, N. & Starobin, D., 2018. Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities. *Computers, Environment and Urban Systems,* Volume 70, pp. 151-162.

Krizek, K., Forysth, A. & Schively Slotterback, C., 2009. Is There a Role for Evidence-Based Practice in Urban Planning and Policy?. *Planning Theory & Practice,* Volume 10, pp. 459-478.

Kuliga, S., Nelligan, B., Dalton, R.C., Marchette, S., Shelton, A.L., Carlson, L. & Hölscher, C., 2019. Exploring Individual Differences and Building Complexity in Wayfinding: The Case of the Seattle Central Library. *Environment and Behavior,* 51(5), pp. 622-665.

Lee, K-S., 2011. Interrogating 'Digital Korea': Mobile Phone Tracking and the Spatial Expansion of Labour Control. *Media International Australia,* 141(1), pp. 107-117.

Lee, J. M., 2020. Exploring Walking Behavior in the Streets of New York City Using Hourly Pedestrian Count Data. *Sustainability,* Volume 12.

Liao, H., Dong, W., Peng, C. & Liu, H., 2016. Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. *Cartography and Geographic Information Science,* 44(6), pp. 474-490.

Lin, Y., Jessurun, J., de Vries, B. & Timmermans, H., 2011. *Motivate: Towards context aware recommendation mobile system for healthy living.* Dublin, Ireland, IEEE, p. 250–253.

Li, R. & Klippel, A., 2012. Wayfinding in libraries: Can problems be predicted?. *Journal of Map & Geography Libraries,* 8(1), pp. 21-38.

Lynch, K., 1960. *The Image of the City.* Cambridge: MIT Press, ISBN-13: 9780262620017, ISBN 0262620014.

Malleson, N., Vanky, A., Hashemian, B., Santi, P., Verma, S.K., Courtney, T.K. & Ratti, C., 2018. The characteristics of asymmetric pedestrian behavior: A preliminary study using passive smartphone location data. *Transactions in GIS,* Volume 22, pp. 616-634.

Martín, J., Khatib, E.J., Lázaro, P. & Barco, R., 2019. Traffic Monitoring via Mobile Device Location. *Sensors,* 19(20), p. 4505.

Mauri, C., 2003. Card loyalty. A new emerging issue in grocery retailing. *Journal of Retailing and Consumer Services,* 10(1), pp. 13-25.

Ma, X., Wu, Y-J., Wang, Y., Chen, F. & Liu, J., 2013. Mining smart card data for transit riders' travel patterns. *Transportation Research Part C: Emerging Technologies,* Volume 36, pp. 1-12.

Mehta, V., 2009. Look closely and you will see, listen carefully and you will hear: Urban design and social interaction on streets. *Journal of Urban Design,* 14(1), pp. 29-64.

Mercieca, J., Kaparias, I., Bell, M. & Finch, E., 2011. *Integrated street design in high-volume junctions: the case study of London's Oxford Circus.* Athens, Greece, City Research Online.

Milligan, G. & Cooper, M., 1985. An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika,* 50(2), pp. 159-179.

Montello, D., 2005. Navigation. In: *P. Shah (Ed.) & A. Miyake, The Cambridge Handbook of Visuospatial Thinking.* s.l.:Cambridge University Press, p. 257–294.

Moore, S., 2017. *Opportunities for conversational AI in government.* [Online] Available at: https://www.gartner.com/smarterwithgartner/opportunities-for-conversational-ai-in-government/
[Accessed 27 03 2022].

Moreira, F. & Ferreira, M., 2016. Teaching and learning requirement engineering based on mobile devices and cloud: a case study. In: D. Fonseca & E. Redondo, eds. *Handbook of Research on Applied E-Learning in Engineering and Architecture Education.* Pennsylvania, USA: IGI Global, p. 1190–1217.

Namratha Reddy, T. & Supreethi, K., 2017. *Optimization of K-means algorithm: Ant colony optimization.* Erode, India, s.n.

O'Sullivan, D., Morrison, A. & Shearer, J., 2000. Using desktop GIS for the investigation of accessibility by public transport: An isochrone approach. *International Journal of Geographical Information Science,* Volume 14, pp. 85-104.

Peng, C., Harwell, M., Liou, S. & Ehman, L., 2006. Advances in missing data methods and implications for educational. In: S. Sawilowsky, ed. *Real data analysis.* North Carolina: Information Age Pub, p. 31–78.

Pettit, C., Lieske, S. & Leao, S., 2016. Big bicycle data processing: From personal data to urban applications. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci,* 3(2), pp. 173-179.

Pikora, T., Bull, F. C. L., Jamrozik, K., Knuiman, M., Giles-Corti, B. & Donovan, R.J., 2002. Developing a reliable audit instrument to measure the physical environment for physical activity. *American Journal of Preventive Medicine,* 23(3), pp. 187-194.

Pollard, J., Spencer, T. & Jude, S., 2018. Big Data Approaches for coastal flood risk assessment and emergency response. *WIREs Clim Change,* 9(5).

Reddy, G., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G. & Baker, T., 2020. Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access,* Volume 8, pp. 54776-54788.

Reid, W., 2005. *Millennium Ecosystem Assessment: Ecosystems and Human Well-being -- Synthesis.* Washington, DC: ISLAND PRESS.

Resch, B., Puetz, I., Bluemke, M., Kyriakou, K. & Miksch, J., 2020. An interdisciplinary mixed-methods approach to analyzing urban spaces: The case of urban walkability and bikeability. *International Journal of Environmental Research and Public Health,* 17(19), p. 6994.

Rousseeuw, P., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics,* Volume 20, pp. 53-65.

Scikit-learn 0.19.1 documentation, 2018. *Sklearn.preprocessing.MinMaxScaler.* [Online]
Available at:
http://scikitlearn.org/stable/modules/generated/sklearn.preprocessing.MinMaxS

caler.html
[Accessed 2018].

Shi, Q. & Abdel-Aty, M., 2015. Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies,* Volume 58, pp. 380-394.

Szczepanek, R., 2020. Analysis of pedestrian activity before and during COVID-19 lockdown, using webcam time-lapse from Cracow and machine learning. *PeerJ,* Volume 8, Article 10132.

The Crown Estate, 2021. *The Crown Estate.* [Online]
Available at: https://www.thecrownestate.co.uk/
[Accessed 19 8 2021].

Thorndike, R., 1953. Who belongs in the family?. *Psychometrika,* Volume 18, p. 267–276.

Till, J., 2007. *Architecture depends.* Cambridge (MA & UK): The MIT Press.

Turner, A., Doxa, M., O'Sullivan, D. & Penn, A., 2001. From Isovists to Visibility Graphs: A Methodology for the Analysis of Arcthiectural Space.. *Environment and Planning B,* Volume 28, pp. 103-121.

Van Dijk, J., 2018. Identifying activity-travel points from GPS-data with multiple moving windows. *Computers, Environment and Urban Systems,* Volume 70, pp. 84-101.

Van Rossum, G. & Drake Jr, F., 1995. *Python reference manual.* Amsterdam: Centrum voor Wiskunde en Informatica Amsterdam.

Varoudis, T., 2017. *depthmapX - Multi-Platform Spatial Network Analysis Software.* [Online]
Available at: https://varoudis.github.io/depthmapX/
[Accessed 19 08 2021].

Vinutha, H. P., Poornima, B. & Sagar, B., 2018. Detection of Outliers Using Interquartile Range Technique from Intrusion Dataset. In: *Satapathy S., Tavares*

J., Bhateja V., Mohanty J. (eds) Information and Decision Sciences. Advances in Intelligent Systems and Computing. Singapore: Springer, pp. 511-518.

Wang, H., Ma, C. & Zhou, L., 2009. *A Brief Review of Machine Learning and Its Application.* International Conference on Information Engineering and Computer Science, IEEE, pp. 1-4.

Wang, W., Lo, S., Liu, S. & Kuang, H., 2014. Microscopic modeling of pedestrian movement behavior: Interacting with visual attractors in the environment. *Transportation Research Part C: Emerging Technologies,* Volume 44, pp. 21-33.

Weisman, J., 1981. Evaluating Architectural Legibility: Way-Finding in the Built Environment. *Environment and Behavior,* 13(2), pp. 189-204.

Whyte, W. H., 1980. *The Social Life of Small Urban Spaces, Project for public spaces.* Washington, D.C.: Conservation Foundation.

Wirz, M., Franke, T., Roggen, D., Mitleton-Kelly, E., Lukowicz, P. & Tröster, G., 2013. Probing crowd density through smartphones in city-scale mass gatherings. *EPJ Data Science,* 2(1), pp. 1-24.

Ye, Y., Zeng, W., Shen, Q., Zhang, X. & Lu, Y., 2019. The visual quality of streets: A human-centred continuous measurement based on machine learning algorithms and street view images. *Environment and Planning B: Urban Analytics and City Science,* 46(8), pp. 1439-1457.

Yuan, C. & Yang, H., 2019. Research on K-value selection method of K-means clustering algorithm. *J,* Volume 2, p. 226–235.

Zacharias, J., 2001. Pedestrian behavior and perception in urban walking environments. *Journal of Planning Literature,* 16(1), pp. 3-18.

Zhang, J. & Leung, Y.-W., 2003. Robust clustering by pruning outliers. *IEEE Transactions on Systems, Man, and Cybernetics – Part B,* 33(6), pp. 983-999.

Zhang, N. et al., 2020. A Study on the Calculation of Platform Sizes of Urban Rail Hub Stations Based on Passenger Behavior Characteristics. *Mathematical Problems in Engineering,* Volume 7, pp. 1-14.

Zhang, Y. & Fricker, J., 2021. Investigating temporal variations in pedestrian crossing behavior at semi-controlled crosswalks: A Bayesian multilevel modeling approach. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 76, pp. 92-108.

Zhu, S., Wang, D. & Li, T., 2010. Data clustering with size constraints. *Knowledge Based Systems,* Volume 23, pp. 883- 889.

6 Insights into pedestrians' navigation in geo-temporal human behaviours: the case of a retail high-street

6.1 Abstract

Urban geometry plays a critical role in determining paths for pedestrian flow in urban areas. This chapter focuses on reviewing the key parameters affecting pedestrian movement behaviour, aiming to identify the way urban space attributes affect walking activities and types of behaviours, and more specifically, the impact of spatial visibility. Machine learning algorithms and a generalised additive model applied to location data have been employed. These data were derived from Wi-Fi tracking techniques collected in a retail high-street in London. Results of the data analysis indicated four types of behaviours, with varying levels of area knowledge. These behaviours are assigned the following 'personas': novice and experts, and likewise activity categories: striders and strollers. More specifically, novice behaviours are negatively influenced by reduced spatial visibility and increased distances between transportation links, while they are motivated by higher temperatures. Expert behaviours are encouraged by retail activities that support local users, such as office, leisure, and garden retail, while spatial visibility mainly affects their movement behaviour on weekdays during the summer months. Spatial visibility influences speed profiles and distances travelled throughout the study period, similarly, indicating that seasonality does not have a significant effect. Strider behaviours are influenced by spatial visibility concerning walking speed; the higher it becomes, the faster the users walk. Understanding the impact of the different physical attributes and environmental factors on pedestrian movement can lead to more accurate and robust models. These, in turn, can be applied by decision-makers in design processes to provide insights into pedestrians' navigation in geo-temporal human behaviours and the urban planning process.

6.2 Introduction

Walking is an essential mode of transportation, and pedestrian movement is a significant influencing parameter in the way cities are planned (Mendiola & González, 2021). Lack of understanding of what end-users need from their surrounding space can become an obstacle to designers in creating a place where social life can be encouraged. While a significant body of research exists investigating the effect of the built environment on walking patterns from diverse perspectives (Gehl, 2011; Gehl & Svarre, 2013; Zacharias, 2001; Mehta, 2009; Clifton, et al., 2007), very little has been made clear about how and to what degree it influences pedestrian behaviour (Choi, 2012; Feng, et al., 2021). Individual behaviour is determined by diverse and interrelated factors (e.g., layout, diversity of activities, individual characteristics). The ease of movement or aesthetics of the built environment may not always function as the dominant factor (Choi, 2012). The movement of a single individual at any time results from the combination of possible forces and circumstances (Dridi, 2015). However, concerning hindering walking, the built environment becomes the determining factor (Choi, 2012).

New data streams have renewed the interest in humans' mobility, resulting in studies aiming to understand better the influence of physical factors on people's walking behaviours (Zaki & Sayed, 2018; Liao, et al., 2016; Orellana & Wachowicz, 2011). For example, the reaction of pedestrian movement within a street network is affected by its connectivity (Hajrasouliha & Yin, 2015). Wayfinding studies suggest that there are two key categories affecting pedestrian movement, (i) spatial attributes such as visibility, layout, or diversity, and (ii) wayfinding support systems, such as signage and information boards (Weisman, 1981; Montello, 2005; Li & Klippel, 2012). Various authors to date have used visibility as a parameter to study architectural space, as walking response stimulates the use of space when direct visibility occurs due to its high impact on the spatial behaviour of a city (Braaksma & Cook, 1980; Parvin, et al., 2006; Hillier & Hanson, 1984; Turner & Penn, 1999; Othman, et al., 2020).

However, the majority of the studies utilising novel data sources, such as GPS devices, video recordings, or Wi-Fi tracking, have focused on traffic safety (Cervero & Duncan, 2003; Karbovskii, et al., 2019; Zhang, et al., 2020; Fernandez-Ares, et al., 2020) or retail footfall (Mumford, et al., 2021). These studies lack in identifying how and to what extent significant indicators in walkability studies, such as density, land-use diversity, form and scale, connectivity, weather, legibility, visibility, complexity, aesthetics, and many more, affect human behaviours. In addition, walkability urban studies have developed a simpler approach to pedestrian movement and accessibility measurement, focusing on the route street characteristics (Hillier & Hanson, 1984; Nakamura, 2016). These studies have not considered the quality attributes of such routes, neglecting parameters such as attractiveness, the character of a place, liveability, or cleanliness (Smith, et al., 1997). To date, there has been limited research undertaken investigating the way architectural and urban design decisions may result in specific visual and spatial patterns, leaving a gap in knowledge, which can introduce social consequences (Hillier, et al., 1993; Parvin, et al., 2008).

6.3 Urban space attributes influence on walking activities and types of pedestrian behaviour

Considering the impact of the urban environment on pedestrian movement, a review of existing research has been conducted to identify critical factors affecting walking patterns. Much of the literature recognises that pedestrians are not a homogenous group and should not be treated as one (Klingsch, et al., 2008; Li, et al., 2021). Several studies acknowledge that multiple physical and personal factors affect the movement of an individual, such as pedestrian characteristics, environmental surroundings, and traffic flow (Dridi, 2015). This study focuses on the physical attributes of external spaces that may influence walking behaviours, but not specifically investigating the effect of occupancy.

6.3.1 Occupancy

Most correlation studies on walkability assess walking behaviours by measuring and analysing the total walking activity, often referred to as occupancies. Occupancy provides researchers with the opportunity to understand the levels of usage of space (Choi, 2012; Fernandez-Ares, et al., 2020; Dridi, 2015) or as a proxy to attractiveness (Hart, et al., 2014; Mumford, et al., 2021). However, it should be noted that it does not necessarily mean that a high occupancy level is connected to high-quality and people-friendly environments.

Conventional measures used to assess street networks for pedestrians utilize levels of service (LOS), a theory proposed by Fruin (1971), where walking space is measured against its pedestrian flow capacity. According to his concept, increased pavement width and walking space are recommended for shoppers. However, although pavement width is highly correlated to pedestrian flow volume (Nakamura, 2016), pavement expansion does not serve as a success indicator on its own, generating increased demand for walking. In terms of quantity, occupancy is mainly influenced by the provision of key destinations (points of interest) and convenient access to these destinations. (Edwards & Tsouros, 2008; Choi, 2012). Previous research investigating elements that attract users in cities argued that "*entertainment or cultural representation may figure prominently in the image of the pedestrian environment and become an important reason for visiting and staying*" (Zacharias, 2001, p. 12).

6.3.2 Form & scale

The key hypothesis for studies on urban design and walkability is that the street configuration is the most important contributing factor to pedestrian movement and accessibility (Hillier & Hanson, 1984; Nakamura, 2016). Street configuration includes street network connectivity, in which well-connected streets are most likely to draw more people. This approach translates complex street information into a single behavioural principle of individuals' preference for increased street network legibility. Observational studies argue that walking distance is overrated in complex network layouts (Canter & Tagg, 1977; Sadalla & Nagel, 1980; Nakamura, 2016). Other studies support that in street connectivity, the number of changes in direction between street segments is the parameter that influences pedestrian flow volume (Hillier, 1996; Hillier, 2009). However, configuration analysis does not account for the quality of chosen routes. Therefore, such analyses can help estimate flows of movement but lack in understanding the qualitative aspects of such routes. Configuration analysis also has limitations related to sufficient consideration of impacts, such as route alteration or time spent in an area, due to key destinations attracting pedestrians. Therefore, review of travel goals, such as going to work, strolling, or shopping, and locations of such destinations is needed to evaluate the extent of influence on pedestrian movement.

Wayfinding studies support two types of travel goals (behaviours); those with a specific destination and those with an exploratory purpose (Gibson, 1988). Such studies identify that the spatial characteristics of a place and the wayfinding support system can heavily affect pedestrian movement (Gärling, et al., 1986; Weisman, 1981; Dalton, 2003). These studies consider visual cues of great importance. They create a “strong image” to any given observer, generating an image of their surrounding environment, which helps pedestrians navigate and understand the city (Lynch, 1960). However, other work argued against such theories, stating that such mental images are rather conceptual than perceptual, highlighting how urban structures act as stimuli due to their role as symbols rather than facilitating movement (Gottdiener & Lagopoulos, 1986).

Spatial movement is encouraged or discouraged by a range of other parameters, such as spatial visibility and its connection to wayfinding (Gath-Morad, et al., 2021; Wiener, et al., 2009). Unlike the configurational studies, spatial visibility assessments consider the cognitive complexity of pedestrian movement. Interest in identifying the link between the built environment and travel behaviour emerged first in the late 1990s in urban planning and, more specifically, in urban design and transportation planning (Handy, 1996; Crane & Crepeau, 1998). Some studies reported attempt to provide a comprehensive theory reflecting pedestrians’ navigational behaviour in cases where there is no specific destination. These studies base conclusions on simulations or survey data interpretation, with the assumption that movements are directed along the lines of sight, wherein the points in which more lines of vision meet as more likely to be chosen by individuals (Batty, et al., 1998; Parvin, et al., 2008; Wang, et al., 2017). Such studies supported that movement and visibility are not separate processes or functions but are interlinked to produce a series of vistas, viewing the world as people perceive it (Gibson, 1979; Ingold, 2000).

For an area to be “human scale”, it should also relate to the ability of a moving pedestrian to perceive small and subtle details better than those who use faster-moving modes of transportation/ locomotion, such as driving cars (Tibbalds, 2001; Clifton, et al., 2007). These intimate details are the ones that determine how people use a space and feel in it. The “appeal” of the streets is important, especially in commercial areas, as they attract the people to the area (Shamsuddin, et al., 2010). Moreover, it is important to introduce personalised elements, such as street-frontage with sign implementation, window design, and others (Mehta, 2009). Bosselmann (1998) was an early proponent that movement speed influences people’s perceptions of a place, as these being influenced by their slow mode of travel.

6.3.3 Walkability, accessibility, and diversity of activities

Drawing on the liveable neighbourhood concept, introducing the ten-minute walking distance of local amenities rule of thumb (Western Australian Planning Commission, 1997; Western Australian Planning Commission, 2007), research has shown that most people will consider walking up to 800 m (10 minutes) to a train station or town centre. Van Nes and Stolk (2012) argue, ‘*when implementing and improving sustainable means of public transportation, urban functions such as dwellings, shops, services, workshops, and offices have to be in short walking distances from stations and the street network must be easily understandable for wayfinding*’ (Van Nes & Stolk, 2012, p. 1). However, a recent study from Nakamura (2016) conducted in the West End area in London, utilising empirical data, concluded that stations are the most critical contributor to pedestrian flows as the nearest destination. At the same time, the study suggested that attractions are more important for multiple nearby destinations than actual stations (Nakamura, 2016).

Urban designers are concerned with diverse elements, such as the character of a place, variety, visual order, and cohesiveness, when examining city centres from a walkability perspective. Cohesiveness requires the creation of a place that has a centre of gravity (Lynch, 1960; Bohl, 2002). For example, this could either be a high street or a square, which naturally draws people in. These centres need to serve a sense of orientation and the capacity for wayfinding. Van Nes and Stolk (2012) argued that the spatial configuration and the street network in the stations' vicinity are dependent, highlighting the importance of the urban grid supplementing transportation nodes. However, the environment has to offer enough complexity to create exciting cues. Ground floor land uses are vital to the importance of the cohesiveness and attractiveness of these centres. Tibbalds (2001) argues that a city "*draws its vitality from the activities and uses in the buildings lining its streets…façades and activities provided at street-level–closest to the eye-level– are particularly important*" (Tibbalds, 2001, p. 40). Gehl (2010) also describes how the treatment of the city's edges, particularly the lower floors of buildings, has a decisive influence on life in the city space.

When modelling pedestrian flow, two key parameters are considered important: speed and density of individuals. The confidence of a pedestrian's sense of direction is reflected in various indicators, amongst them, the walking speed of individuals (Cornell, et al., 2003). An increased number of pedestrians may disrupt walking patterns, potential collisions with others, or feelings of discomfort (Gehl, 2010). Increased pedestrian traffic can directly impact walking speed, causing pedestrians to slow down. Literature suggests that one additional influencing parameter on walking speed is the purpose of the trip. Gehl and Svarre (2013), in their work, divided everyday activities into three key categories: necessary, optional, and social, focusing on when these activities are most likely to happen while other researchers divide activities into two key categories, such as utilitarian and leisure, based on the purpose of the trip (Ki & Lee, 2021). Empirical observations suggest that utilitarian activities, such as going to work or commuting, generally have higher speeds. In contrast, leisure activities, such as shopping or walking during leisure time, have lower speeds (Choi, 2012).

Recreational shopping has become a form of public life, with retail activities generating most of the trips in city centres, considered the most significant attractors of pedestrian traffic (Pawsey, 1985; Pushkarev & Zupan, 1975; Barton, et al., 2003). The overall spatial configuration of the street network proves to be a more robust correlate of walking than local street-level attributes. At the same time, only average sidewalk width appears to be a significant correlate of walking among the streetscape measures (Alfonzo, et al., 2008). However, the most significant and consistent correlate of the distribution of flows is the number of recreational uses at the segment level. (Özbil, et al., 2019)

As Jan Gehl argues (2010), possibilities are provided at the macro level, but the battle is fought at the micro-level. Gehl's work (2011) focused on how senses are central to how people perceive information. He argued that any movement speed beyond walking and running limits the amount of information that can be absorbed. In an empty space, the level of information that can be absorbed differs from that of a crowded space. However, the movement speed of an individual can differ and be influenced by many spatial and environmental parameters, such as age, the purpose of walking activity, time of the day, weather information, and density.

6.3.4 Safety & security

People's needs in pedestrian navigation can be divided into three layers: a physical sense, a physiological safety, and a mental satisfaction layer, according to Maslow's theory (Zhixiang, et al., 2015). When humans are called to make decisions about travel modes, they collect information related to security risks and safety to the specific travel modes. High visibility areas are linked to better visual environments with no obstructions and are believed to reduce the fear of crime (Asami, et al., 2002). However, feelings of safety can be linked to the number of people present, reflecting Jacobs' concept of "eye-on street", where the more people on the street, the better the surveillance will be (Jacobs, 1961). Man-made environments impact the sight of line and street configuration, and design can impact space visibility. Several studies concluded that high visibility areas are more likely to be visited due to safety and increased social interaction (Jacobs, 1961; Qiu, et al., 2013; Hillier & Sahbaz, 2011).

6.3.5 Seasonality

Environmental parameters are recognised as key criteria influencing walking activities within a public space (Gehl, 2010). Most of the studies reviewed weather conditions and their effect on route choices and speed profiles, linking them to the experience of comfortable conditions and psychological support (Gehl, 2011; Vasilikou & Nikolopoulou, 2020; Han, et al., 2012). Other studies focused on physical activity levels and health implications (Chan & Ryan, 2009; Forsyth, et al., 2009). Few studies looked at the potential impacts of seasonality on active travel and transportation, while others emphasized the behavioural change due to climatic conditions (Hong, J., 2016). For example, important correlations were found between weather-related parameters such as temperature, rain, and humidity, and walking and cycling activities (Noland & Ishaque, 2006; Humpel, et al., 2004).Accordingly, weather-related data are introduced into this study to permit these effects to be examined.

Although the literature suggests that strong correlations exist between walking patterns and seasonality, the effect of seasonality from the perspective of "time within the year" is not well examined. Some studies that looked at seasonality from the perspective of winter, or summer, reviewed its effects against the likelihood of physical activities' occurrence (Merchant, et al., 2007; Matthews, et al., 2001). Empirical observations suggested that reduced speed has been observed when the temperature reaches 25 ^{o}C, while higher walking speeds occur when temperatures are low (0 ^{o}C) (Choi, 2012). Nevertheless, such studies were based on survey interviews and empirical data rather than large-scale walking activity monitoring, leaving an additional gap in the existing literature.

Furthermore, the effect of seasonality can be relevant from the perspective of "time within the day". Walking behaviour is influenced by the type of activities, and the type of activity can vary throughout the day, i.e., morning commute to work or lunchtime break. Various studies explore the health impact of lunchtime walking activities, however, daily patterns in relation to their influence on walking activities are not thoroughly investigated (Sianoja, et al., 2018; Kubba, 2017).

6.3.6 Aesthetics

Another important influencing parameter of human movement is the aesthetic qualities of a space, which are the most undefined of the dimensions described, leaving another knowledge gap. Their influence on human behaviour is yet to be revealed and is difficult to measure. Gehl (2010) describes how the treatment of building edges has a decisive influence on the liveability of a space. He specifically discusses the opportunities that ground floors have to offer to pedestrians and the transparency of the facades. However, observational studies revealed that although less in frequency, pedestrians observe and interact with the building facades as a total (Choi, 2012).

Lynch (1960) argued that imageability as a physical environment quality creates a strong image for an observer. More specifically, he mentioned, "*It is that shape, colour, or arrangement which facilitates the making of vividly identified, powerfully structured, beneficial mental images of the environment*" (Lynch, 1960, p. 9). Ewing & Handy (2009) further argued that the complexity of the environment, referring to the visual richness of a place, is an important parameter related to walkability. They further discussed that complexity depends on various factors, such as the amount and types of buildings, architectural diversity and ornamentation, landscape elements, and others.

The existing literature often ignores the impact of spatial visibility on walking patterns (Carr, et al., 2011; Frank, et al., 2010), or it focuses on the effect of visual connectivity in specific aspects, such as occupancy (Hajrasouliha & Yin, 2015), leaving a gap in knowledge. This study hypothesizes that pedestrians' decisions to move from one point to another are primarily shaped by the visual experience of their surrounding environment. As the factors influencing pedestrian movement may differ based on the context of an area (Nakamura, 2016), attention is paid to the travel patterns within a retail high-street location in central London. This study focuses on reviewing key parameters affecting pedestrian movement, as identified by the literature (Southworth, 2005; Whyte, 1980), in relation to actual movement data. The aim is to identify the way urban space attributes affect walking activities and types of pedestrian behaviour, and more specifically, the impact that spatial visibility has on them. This chapter provides an in-depth analysis of pedestrians' affective experience of urban design qualities related to the spatial visibility, configuration of the built environment, and key attractors, based on Wi-Fi location data collected in a retail high-street in London. Further, the research expands existing knowledge by offering insights into complex relationships between urban design qualities and pedestrian movement, focusing on aspects that may hinder or support walking intentions.

6.4 Study area

The chosen study area is Oxford Circus, located in the West End, in central London. The study area was chosen firstly due to data availability and secondly due to its complexity. The area includes a wide range of activities, from retail to offices, high connectivity, and direct connections to nightlife. Oxford Circus connects two of the most famous retail streets, Oxford and Regent Street. The intersection of these two streets consists of an interesting case study (Figure 6-1) as it has undergone a recent transformation to an open diagonal crossing and introduced space sharing as a means of enhancing quality environments, while it also has the highest pedestrian volumes recorded (Mercieca, et al., 2011).

Oxford Street is recognised as a global retail destination and is one of the longest continuous retail streets in Europe. It significantly contributes to the area's economy, while it is also an important transportation corridor, connecting the city of London to the West, the Midlands, and the North of England. Regent Street has a frontage of c.2km with over 150 retail and catering outlets and 700 small and medium-sized businesses (Transport for London, 2011). Due to its conflicted twin roles, the area has been of interest to several researchers and practitioners. These studies focus on traffic and pedestrian safety, air pollution, and pedestrianisation design, aiming to address the complexity of the area better and create safer and liveable environments for its users (Van Veldhoven, et al., 2019; Turner & Giannopoulos, 1974; Mercieca, et al., 2011).

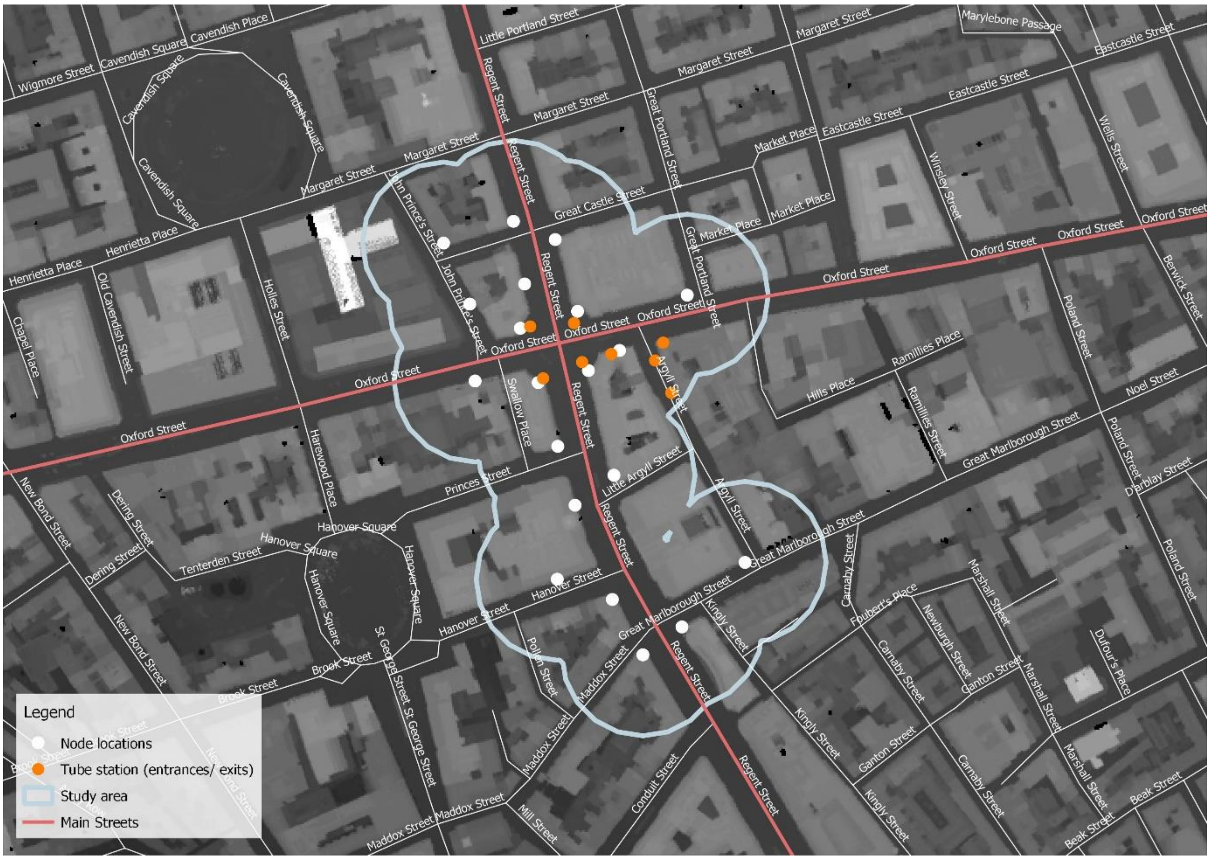


Figure 6-1: Study area and the distribution of Wi-Fi nodes

6.5 Materials and methods

This study employed the framework methodology developed in Chapter 5, while it further explored the individual behaviours extracted via the adopted model. The methodology followed is described in Figure 6-2, while Section 6.5.1 provides an overview of the framework methodology employed.

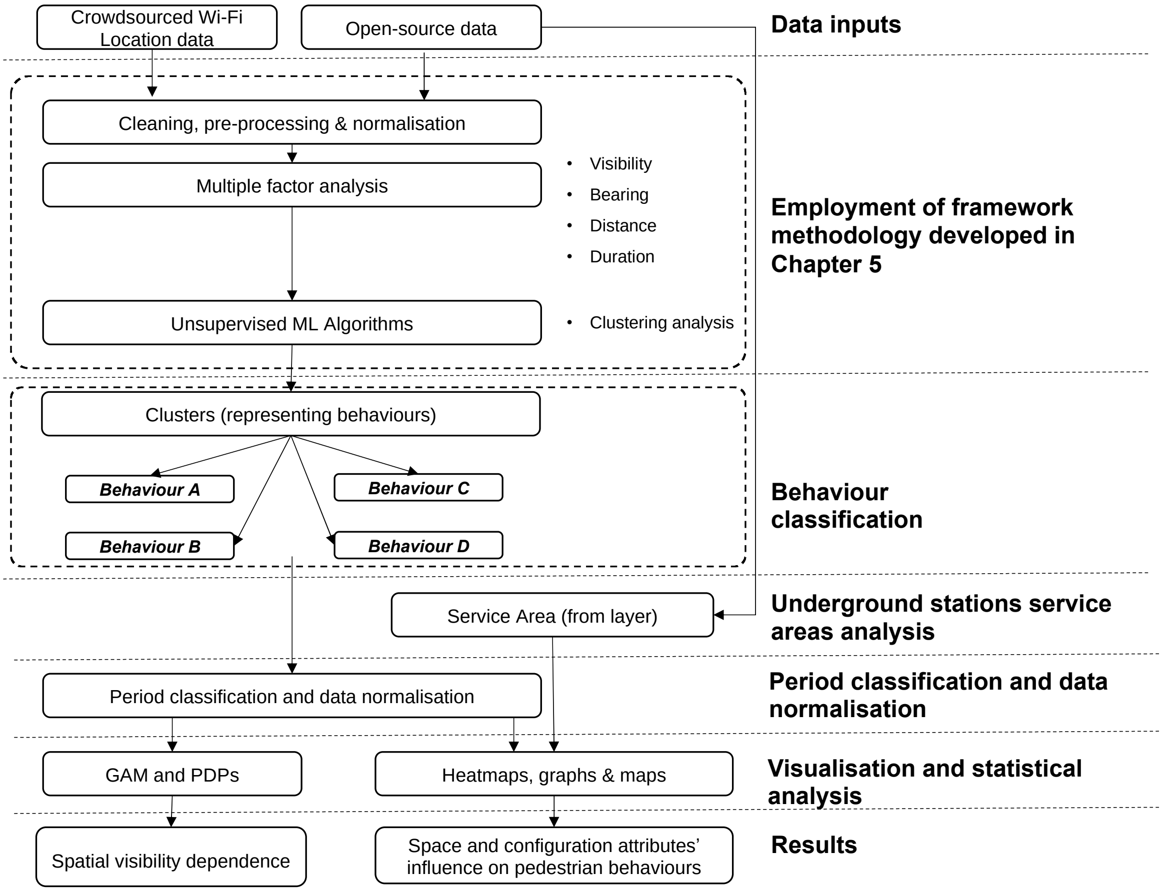


Figure 6-2: Overview of the methodology followed

6.5.1 Revealing existing walking behaviours via K-means analysis: Data collection & assimilation

This study employed the framework methodology developed in Chapter 5. This section provides an overview of the model, including the steps and data types used and adopted for this study. Pedestrian data have been given for this study from The Crown Estate, and they have been obtained by Wi-Fi tracking over two typical weeks (Angelelli, et al., 2018). In total, this study utilises location data for 22 different days spread across August, September, and October 2017. A multi-field .csv file was provided as a post-processed output information from the technology provider, removing all privacy-related information (Accuware Inc, 2017). The data captured was unique Media Access Control (MAC) address-unique identifiers, signal strength, X Y Z coordinates location, and a timestamp. Signal strength was also recorded and used for triangulation purposes to derive location. Data frequency varied between 1-60 seconds, dependent upon the type of devices, manufacturer, or activity level (Accuware Inc, 2017).

Weather information was sourced from Weather Underground, utilising a private weather station in proximity (Station ID: ILONDON636). Precipitated Rate (in) was also collected but not used as there was no rain during the period of the collected data. Data assimilation, processing, modelling, and visualisation were completed using python software. Internal recorded locations were excluded from this analysis, as this research does not aim to understand interactions between internal and external environments.

As people move in space, visible areas and points within the surrounding environment change dynamically. In wayfinding studies, visibility in a space is critical to determine pedestrian pathways. The stated hypothesis of this study is that movements are directed along the lines of sight, as described in Sections 6.3.1 and 6.3.6. Therefore, the visibility graph analysis (VGA) was employed via the open-source software DepthmapX_net_035 (DepthmapX development team, 2017). Values represent the number of points that have visibility towards the specific grid cell on pedestrian eye-level. The analysis utilised a 2km radius in a 2x2m grid to derive results that are not distorted due to the small scale. In previous research, the minimum distance thresholds used range from 300m up to 1km; hence the chosen scale is acceptable (Ahrné, et al., 2009).

The final dataset assessed consisted of the following variables: Duration (s), Distance (m), Speed (m/s), Bearing (degrees), Spatial visibility (VGA), Humidity (%), Wind speed (mph), Solar radiation (w/m^2), Temperature (^oC) and Hours (h). The positional factors of duration, distance, speed and bearing were computed from the timestamped point location in the source data. Location data were assessed using a data analysis framework methodology developed based on previous research (Chapter 5). Analysis was performed on a daily resolution as behaviour may vary within a week. Unsupervised machine learning algorithms were used to identify the number of existing walking behaviour in the study area. More specifically, cluster analysis via K-means analysis was performed (Yuan &

Yang, 2019) for the following four variables: Duration (s), Distance (m), Speed (m/s), Bearing (degrees), Spatial visibility (VGA).

According to TFL's research, there are four different and distinct types of journeys, each with specific travel characteristics, thus: Novice strider, Expert strider, Novice stroller, and Expert stroller (Davies, 2007). Cluster analysis was employed to reveal pedestrian behaviours. Each cluster represents one behaviour. These behaviours, based on the combination of the type of activity undertaken, and the knowledge of the area, are classified as: Cluster 0: Expert Strider, Cluster 1: Novice Stroller, Cluster 2: Novice Strider, and Cluster 3: Expert Stroller and are utilized for this study (Chapter 5). Within this categorisation, knowledge of the area is incorporated for all types of activities. Data points were mapped against all collected parameters and divided into clusters for the analysis. Data points for each cluster were assessed and visualised separately. Data exploration was undertaken to explore differences in cluster behaviours. Visualisation of clustering results using mean values and the python package "seaborn" was initially undertaken, investigating the effect of the built environment on walking patterns (Waskom, 2021).

6.5.2 Underground stations service areas analysis

QGIS software was employed to assess the study area using the algorithm "Service Area (from layer)" to better understand how underground stations' locations respond concerning the street network. More specifically, the QNEAT package was used to extract areas of 800m walking distance (QGIS Python Plugins Repository, 2020). The assessment was conducted using two separate layers of information obtained from the Points of Interest (POIS) of the Ordnance survey (Ordnance Survey, 2021). The first layer used all data points for the London underground entrances located on the two high streets of Oxford and Regent Street. The second layer used only the data points classified as "Oxford Circus station". The assessment was undertaken using three key radius distances: 400m, 800m, 1200m, which respond to 5, 10, and 15 minutes of walking distance respectively These values were selected as they are commonly used values in walkability studies and, more specifically in the liveable neighbourhood concept (Western Australian Planning Commission, 1997; Western Australian Planning Commission, 2007).

6.5.3 Period classification and data normalisation

The period classification was done based on the August 1st sun cycle in the U.K. For consistency purposes, the same was calculated for all the days. Classification used is as follows: 'First light': from 05:24 to 08:40, 'Morning': from 08:40 to 12:10, 'Lunchtime': from 12:10 to 14:00, 'Afternoon': from 14:00 to 20:47, 'Last light': from 20:47 to 21:28 and 'Nighttime': from 21:28 to 05:24. These time-of-day classifications were computed for each of the timestamped point locations recorded in the source data. All values presented were normalised against the number of hours and minutes per period for comparison purposes. Following that, they have also been normalised, where appropriate, against the number of hours recorded per period and against street length, which forms part of the study area. Due to the study area extent covering different lengths in each street segment, for comparison purposes, street length normalisations were applied with Regent Street calculated length 372m and Oxford Street calculated length 236m.

6.5.4 Revealing spatial visibility dependence: visualisation and statistical analysis

Modelling the variables against the spatial visibility for each data point was undertaken to ascertain the effects of diverse identified parameters concerning spatial visibility. Data exploration was conducted using a generalised additive model (GAM) and visually assessing the partial dependence plots (PDPs) results. A GAM is valid for this type of analysis as it can reveal either monotonic, linear, or even complex relationships, based on how the individual variables respond to the changes of the dependent ones. GAMs were selected as the most appropriate method, as they are similar to Generalized Linear Models (GLM) but differ as they relax the linear assumption. Therefore, they potentially reveal non-linear relationships and significant data structures that would otherwise be missed (Wiley & Wiley, 2019). GAMs extend Generalised Linear Models (GLM) to include a smoothing basis function, enabling it to measure arbitrarily non-parametric relationships (Barton, et al., 2020). The mathematical expression of GAMs is shown in Equation 6-1.

Equation 6-1: Mathematical expression of GAMs

$$\mathbf{g(E[y|X]) = \beta_0+f_1(X_1) + f_2(X_2, X_3) +\ldots+f_M(X_N)}$$

where $X_T = [X_1, X_2, ..., XN]$ are independent variables, y is the dependent variable, and g() is the link function that connects the predictor variables to the expected value of the dependent one.

For the GAM modelling, it was initially assumed a Poisson distribution; however, the model wouldn't converge, meaning that the data is meaningless for such distribution. Therefore, a linear GAM exists in these data. However, linear models may also have log and inverse terms that follow other types of curves while remaining linear in their parameters (Kamal & Saxena, 2019).

From the 'pyGAM' package in python, the 'LinearGAM' function was used to identify links and normal distribution (Servén & Brummitt, 2018). The penalised regression spline (n=8) was employed for the smoothing basis function to reduce manual editing of different transformations for each variable. This is advantageous as it reduces computational costs while avoiding model overfitting. This is achieved due to the smoothing penalty applied on the coefficient estimation, generalising the smoothers by decreasing them towards zero (Barton, et al., 2020). This approach also aids the model in reducing the effects of concurvity, as the occurrence of variable dependence may result in considering false statistically significant effects due to poor parameter estimation and increased confidence intervals (Zhou, Y., 2018; Barton, et al., 2020).

For the calculation of concurvity, we used the 'mgcv' package in R-studio software, as a similar metric does not exist within the python implementation package (Wood, 2006). The use of 'concurvity()' function from the 'mgcv' package allows us to measure the way a smoothed variable can be estimated by another (Barton, et al., 2020). All variables are grouped together but calculated against each one to conclude into one value for concurvity. The metric computes three indices, varying from values 0 to 1: worst, observed, and estimate scenarios, where 0 = no concurvity and 1 = lack of identifiability between the variables. The latter is the most reliable measure (Wood, 2006). However, there are no universal criteria for concurvity thresholds, some studies suggest that when values are higher than 0.5, then they start to introduce considerable errors (He, S., 2004; Ramsay, et al., 2003). Concurvity test (Table 6-1) suggests that only Temperature (^{o}C) has an estimated concurvity value of approximately 0.59. Some studies suggest that the model requires inspection if the variable reaches a high value (>0.8) in the worst scenario (Ross, 2019). As Temperature (^{o}C) is an important parameter, with its value close to the threshold (red text in Table 6-1), without exceeding the suggested value in the worst scenario, this research included Temperature as a variable in the final models.

Table 6-1: Concurvity test for all variables

	s(Bearing)	s(Duration)	s(Distance)	s(Speed)	s(Humidity)	s(Wind)	s(Solar)	s(Temperature)	s(hours)
worst	0.268	0.130	0.378	0.292	0.595	0.034	0.553	0.611	0.778
observed	0.167	0.117	0.215	0.280	0.218	0.029	0.462	0.596	0.251
estimate	**0.116**	**0.097**	**0.302**	**0.162**	**0.486**	**0.027**	**0.306**	**0.588**	**0.349**

Partial dependence plots (PDPs) were employed to identify the effects of the covariates better. PDPs look for the variable dependence and allow for a visual assessment of the results, which help determine the variables that show the strongest effect due to the ease of interpretation (Fasiolo, et al., 2019). The final results are presented as PDPs using matplotlib.pyplot as an appropriate method. They help to visualize how the model variables work in a graphical exploratory way and how the predictions partially depend on the values of the input variables. PDPs evaluate the variable effects by displaying the change in the mean predicted number of spatial visibility as the variable interval changes over its distribution (Goldstein, et al., 2015) –"*mean centred since the smoothed variables must sum to zero in a GAM*"- ((Barton, et al., 2003, p. 31), accompanied by the 95% confidence intervals. The estimated p-value (<0.01) was used to determine if the variable significantly affects spatial visibility.

6.6 Results

The mobile data used incorporate 3,240,361 unique mobile users, spanning from August to October 2017. General occupancy patterns and the effect of spatial attributes have been extracted and are further analysed in the following sections.

6.6.1 Space and configuration attributes' influence on pedestrian behaviours

6.6.1.1 General occupancy patterns

General occupancy patterns extracted for August, including weekends, revealed that the greatest occupancy is observed during lunchtime period (12:10 to 14:00), while the lowest occupancy rates are during Nighttime for weekdays (from 21:28 to 05:24) and first light for weekends (from 05:24 to 08:40). The increased occupancy observed throughout the days during lunchtime implies the influence of active land uses and activities at this time of the day. Due to the strong retail character that the study area possesses, the influence of retail activity on occupancy patterns has been further observed during Sundays. Although retail opening times are extended during the weekdays, Sundays' retail activity hours occur from 12:00 to 18:00 in both high streets. Overall, Sundays present the lowest occupancy rates within a week (Figure 6-3), while it decreases substantially after 17:00, by almost 50%.

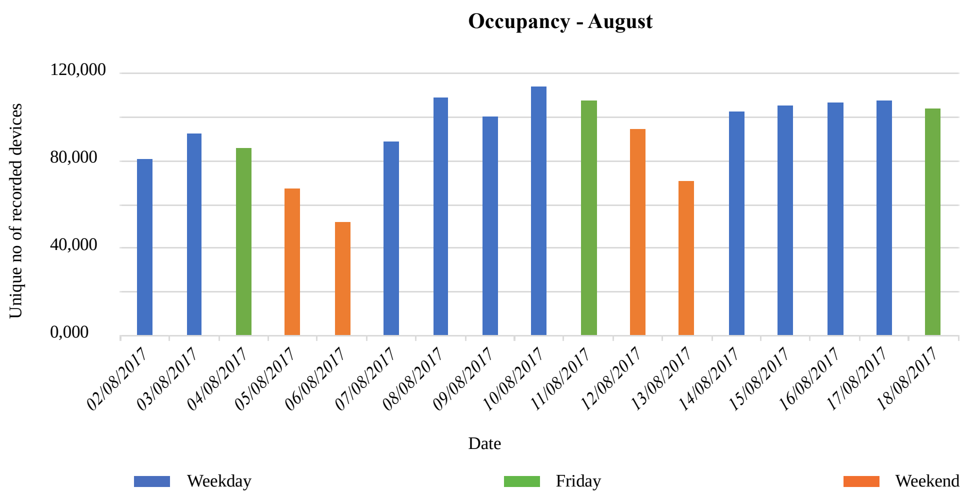


Figure 6-3: Average normalised total occupancy per day in August within study area

In addition, occupancy in the first week of August appears to be lower than the rest of the weekdays (Figure 6-3). Furthermore, general occupancy patterns extracted from the collected data points indicate that Oxford Street has a higher occupancy than Regent Street (Figure 6-4). This may occur due to the wide range of activities within or near the study area every year during summertime. Activities range from the closure of Regent Street for the “Summer Street” event to the “Fashion & Design Month”. These activities occur from the end of June until the end of September. For the year of the recorded data, the following events were recovered, which have affected occupancy rates during the first week of August. These are:

- Event 1: From midday until 16:00 on Saturday 12 August, delays to traffic and road closures on Regent Street, Piccadilly Circus, Haymarket, Pall Mall East, Trafalgar Square, and Whitehall.
- Event 2: From 02 July to 23 July, every Sunday from 12-6 pm was closed as of Summer Street by Regent Street. Special activities day 23 July.
- Event 3: Regent Street Stylists 26/06 to 31/08.
- Event 4: NFL on Regent Street 2017 (30/09).
- Event 5: September 2017 – Fashion & Design Month.
- Event 6: 11 August 2017 - Oxford Circus Tube station evacuated due to fire. The station was evacuated shortly before 09:00 BST, and trains did not stop at the station for nearly two hours.

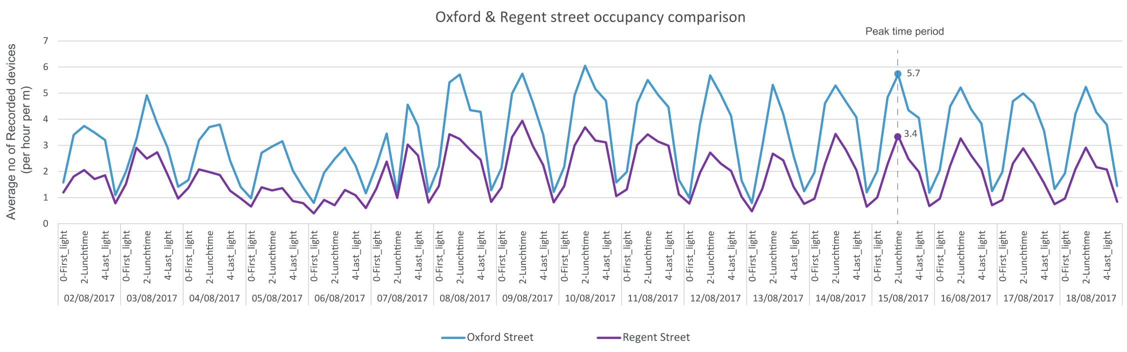


Figure 6-4 Recorded no of devices in Oxford and Regent Street, normalised per hour and per street length.

Occupancy patterns have been visualised, and heatmaps illustrating the number of recorded devices (users) for each cluster in August have been produced (Figure 6-5). Findings suggest that reduced occupancy has been observed for the first and third week of August, specifically for Clusters 0 and 3. It can be assumed that this was due to the increased occupancy that most likely occurred during July due to event 2, in combination with the summer holiday period throughout August.

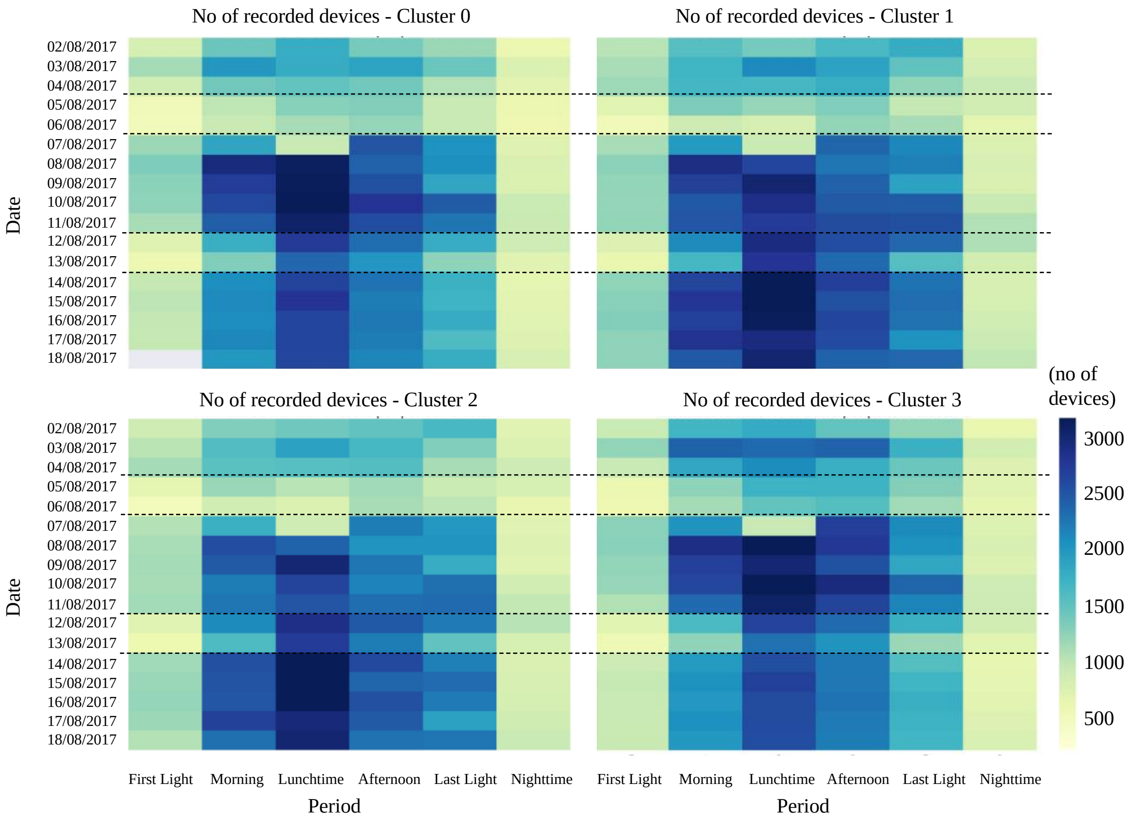


Figure 6-5 Heatmaps illustrating normalised number of recorded devices (users) for each cluster in August. Cluster number is indicated on the top of each graph. Dates are shown on the vertical axis, for dates 2nd of August until 18th of August, while period is indicated on the horizontal axis. Dash lines indicate weekends.

Mean temperature values recorded were generally high, at approximately 17.2 ^{o}C. However, the daily average temperatures dropped to 13.6 ^{o}C on the 9th of August (Figure 6-6). During these days, Cluster 0 and 3 occupancy increases but falls again when daily average temperatures rise on the 12th of August to 19.3 ^{o}C. The opposite occupancy effect is observed for Cluster 1 and 2 concerning temperature values. In addition, wind speed on the week starting at the 7th is low, ranging from 0 to 0.22 mph, while the week after, slightly stronger wind speeds occurred, with a maximum average daily value of 0.88 mph on the 12th of August (Figure 6-6). This finding indicates a relationship between weather conditions and occupancy patterns on novice behaviours (Cluster 1 and 2), implying that cooler temperatures and low wind speeds have a positive effect on these behaviours.

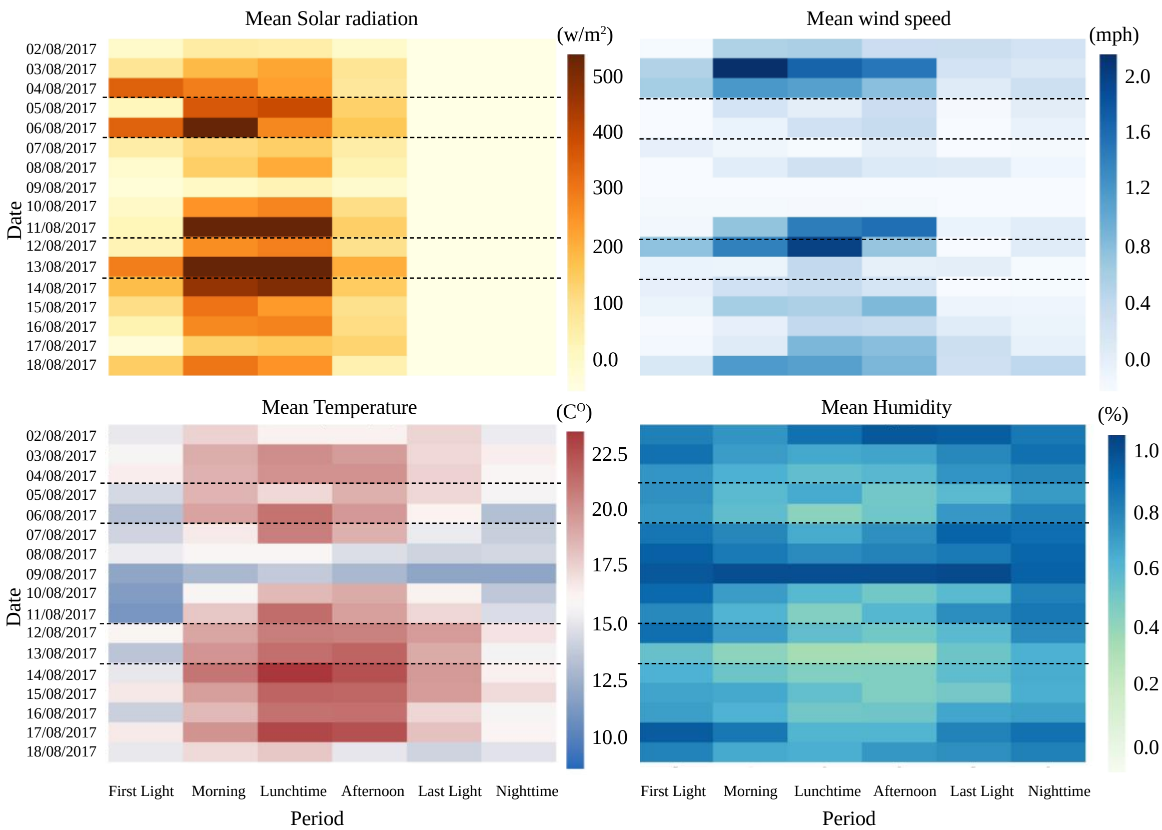


Figure 6-6 Heatmaps illustrating mean environmental values as recorded for month in August. Solar radiation in w/m2 (top left), mean air temperatures in Celsius (bottom left), mean wind speeds in mph (top right) and mean humidity in % (bottom right). Dates are shown on the vertical axis, for dates 2nd of August until 18th of August, while period is indicated on the horizontal axis. Dash lines indicate weekends.

6.6.1.2 Form and scale

Results derived from VGA indicate that increased visibility occurs throughout the length of Oxford Street, with the highest values concentrating in the junctions and crossroads, as though expected (Figure 6-7). Regent Street presents a different visibility profile, with relatively lower values recorded. The upper part of the street which connects Oxford Circus to Piccadilly Circus has increased visibility, while as the observer approaches towards the end of the street, lower visibility patterns are recorded. The reasoning behind this is primarily due to the street network curvature and the lack of other streets connected to it.

Figure 6-7 VGA results for study area in a radius of up to 2km. The greater the number, the greater spatial visibility is.

The view from Oxford Street towards Regent Street, on either North or South sides, is blocked due to the street network geometry and configuration and building structures located in the middle. Therefore, both directions look like “dead-ends” or present unknown destinations to people who have limited experience of the area, limiting the potential for these types of pedestrians. Therefore, this impact of visibility indicates the difference in occupancy patterns in the two main streets, as also revealed in Figure 6-4. It should be acknowledged though that the online maps and “*find my route*” apps (e.g., Google maps) have altered “*explorative*” trips, as people tend to follow a defined route to reach their desired destination blindly.

6.6.1.3 Walkability, accessibility, and density of activities

Mean daily walking speed values recorded in August range from 0,19 to 0,44 m/s (Figure 6-8). Although literature suggests an estimated speed of 1,33 m/s (Bohannon, 1997), such findings indicate that either external parameters are causing pedestrians to slow down, or that a great majority of the recorded pedestrians are not moving. However, although for the week beginning on the 7th, occupancy rates are higher by 7% in comparison to the one beginning on the 14th, higher walking speeds are recorded when increased occupancy occurs (Figure 6-5). This event occurs when average mean temperatures dropped from 17.2 ^{o}C on the 7th of August up to 13.6 ^{o}C on the 9th (Figure 6-6). These findings imply that other parameters, such as external weather conditions, can have a higher influence on walking patterns than merely the density of pedestrian traffic.

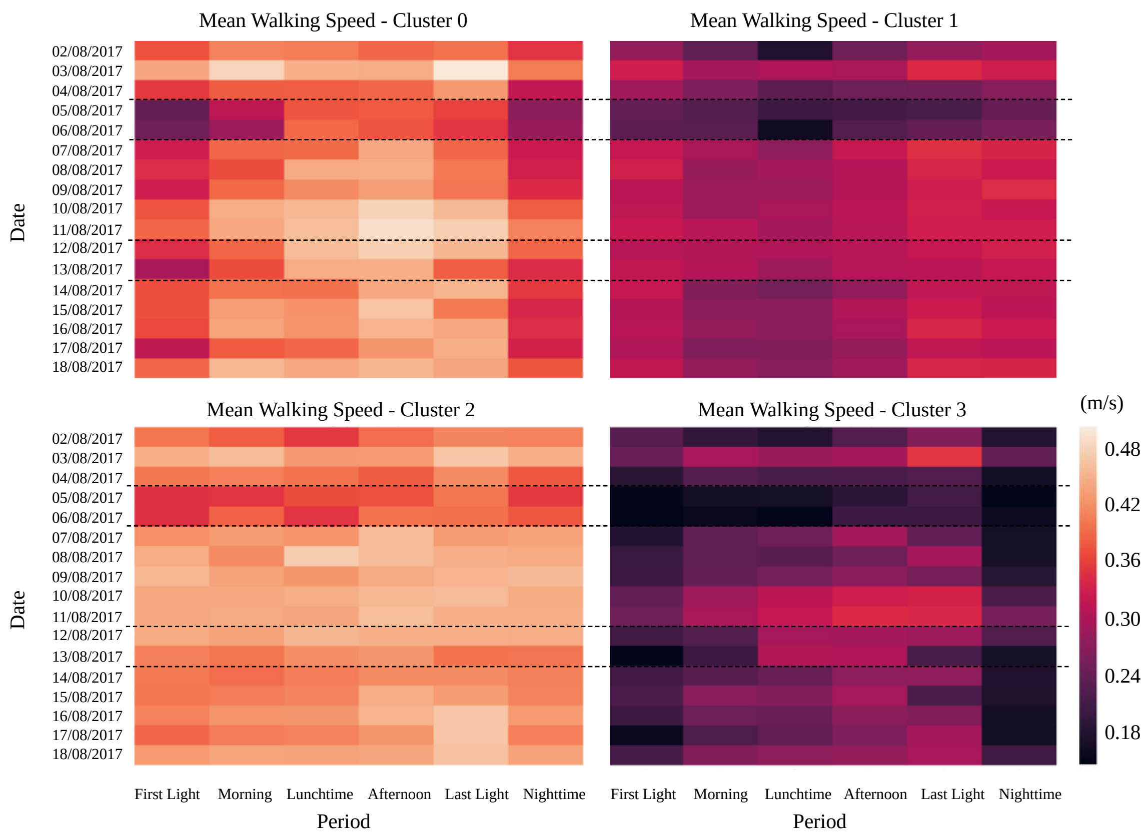
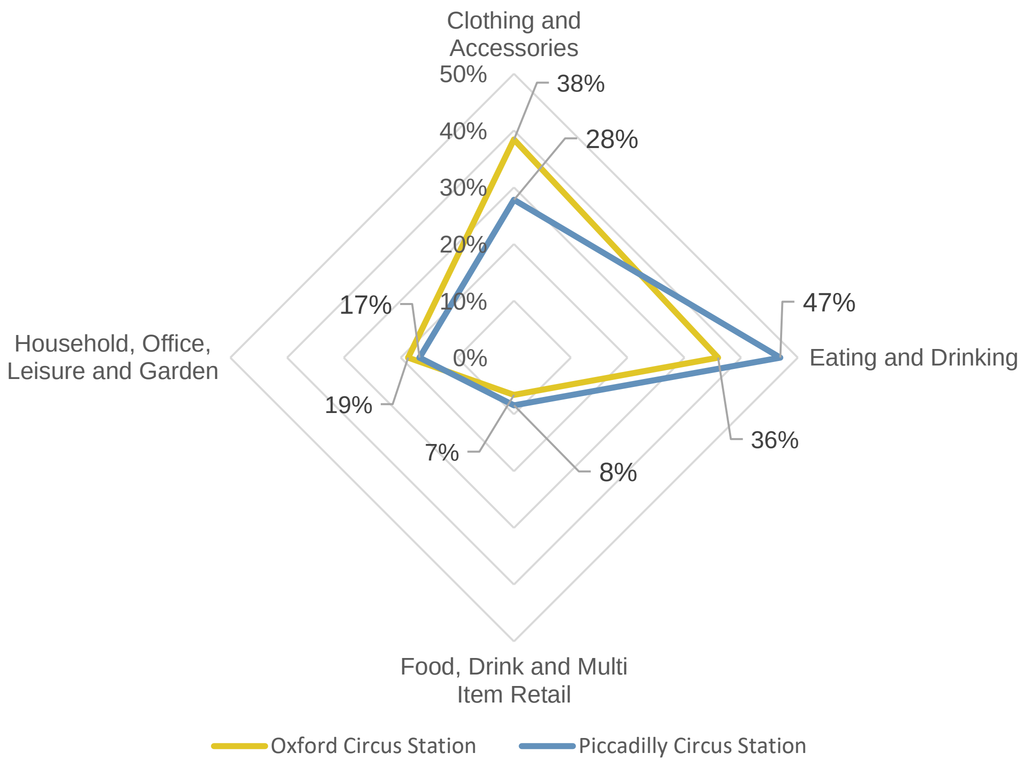


Figure 6-8 Heatmaps illustrating mean walking speeds in m/s as recorded for each cluster in August. Cluster number is indicated on the top of each graph. Dates are shown on the vertical axis, for dates 2nd of August (top) until 18th of August (bottom), while period is indicated on the horizontal axis. Dash lines indicate weekends.

Results derived from the service areas' assessment indicated that Oxford Street tube stations' service areas (from Marble Arch station to Tottenham court road station) fall within an 800 m radius of the destination. They can be reached by an 800 m walk along streets, while the walking distance between Oxford Circus underground station and Piccadilly Station is slightly longer and requires more than 10 minutes walking (Figure 6-9).

Figure 6-9 Service areas for 5, 10 and 15-minutes walking distance for Oxford Circus station only underground entrances/ exits.

It should be noted that Oxford Circus station entrances and exits are better located to the diverse retail activities provided within the area, in comparison to Piccadilly Circus. More specifically, Oxford Circus tube station is in proximity to 13% more overall retail provision (Figure 6-10), in addition to the better transportation network offering. The most significant difference is found in the Clothing & Accessories category, where provision is better for up to 37%.



Type of activity	Oxford Circus Station (No of points)	Piccadilly Circus station (No of points)	Difference (%)
Clothing and Accessories	807	509	37%
Eating and Drinking	755	859	-14%
Food, Drink and Multi Item Retail	138	154	-12%
Household, Office, Leisure and Garden	392	305	22%
Grand Total (count of points)	**2100**	**1831**	**13%**

Figure 6-10 Retail activities provision breakdown within 800m walking distance (10 minutes) from Oxford Circus station and Piccadilly station.

Data points looking at retail functions in the area were also assessed to better understand the retail provision of the two primary streets. Regent Street's retail activities are higher in items that address mainly "locals", as 30% of its total provision is household, office, leisure & garden retail. In comparison, Oxford Street offers 7% more retail in clothing & accessories and 4% more in Food, Drink & multi-item retail (Convenience stores, supermarkets, bakeries, Off Licences and Wholesalers, etc.) (Figure 6-11).

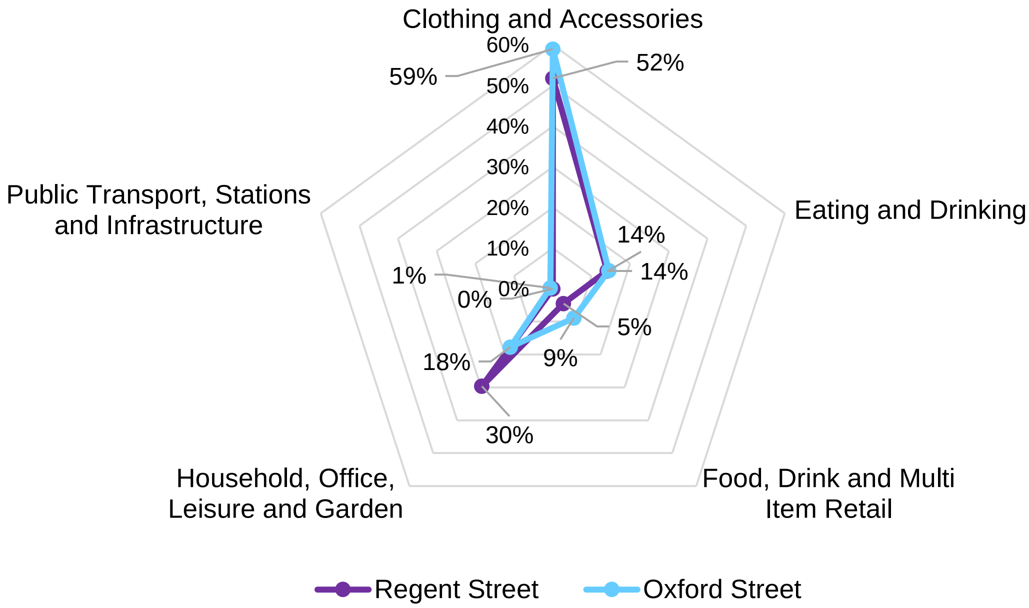
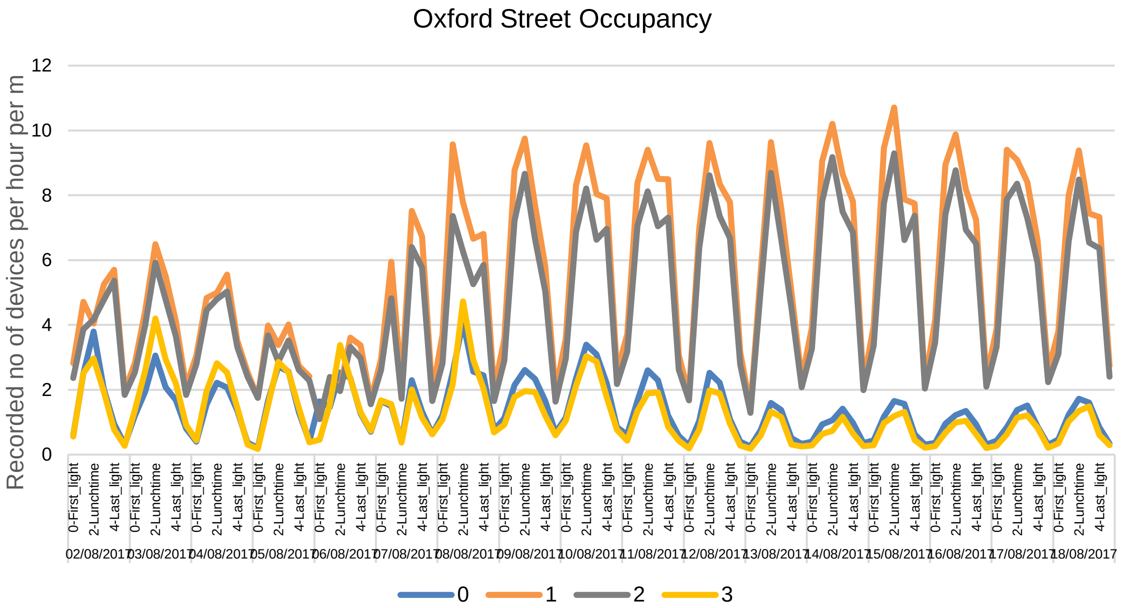


Figure 6-11 Overview of retail provision in Regent and Oxford Street.

Mapping cluster analysis results against the street network revealed that Oxford Circus is mainly occupied by Clusters 1 and 2, while Regent Street by Clusters 0 and 3 (Figure 6-12). Cluster 1 and 2 are novice behaviours, indicating that they are not regular users of the area. Average occupancy in Oxford Street for Clusters 0 and 3 has a 35.8% and 41.6% decrease respectively compared to Regent Street, while for Clusters 1 and 2 are significantly higher, with a 268.2% and 192.7% increase respectively (Table 6-2). These findings indicate that Oxford Street supports retail functions that apply to wider audiences, better supporting all types of behaviours found in the area. Regent Street and Oxford Street generally draw most of the users recorded, highlighting the need for fewer changes in direction between street segments and higher spatial visibility for wayfinding and safety purposes.



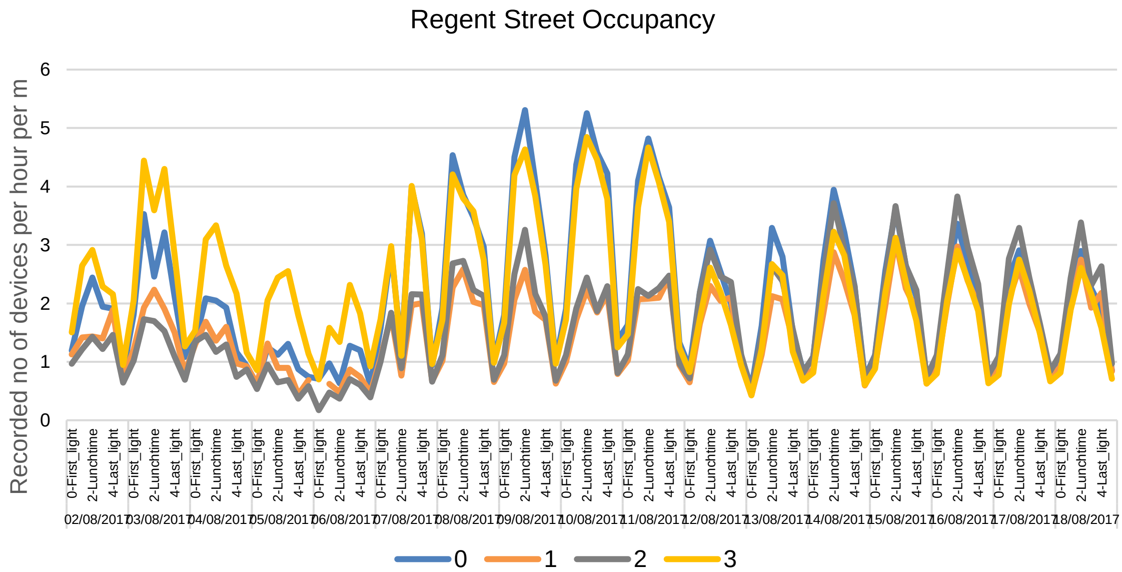


Figure 6-12 Average normalised number of recorded points for each cluster found in Oxford Street (up) and Regent Street (down). Colours represent clusters.

Table 6-2: Average occupancy values for the two high streets – Normalised per metres and per hour

Parameters	Cluster 0	Cluster 1	Cluster 2	Cluster 3
Type	Expert strider	Novice stroller	Novice strider	Expert stroller
Oxford Street	1.38	5.56	4.80	1.28
Regent street	2.15	1.51	1.64	2.19
% of difference	**-35.8%**	**268.2%**	**192.7%**	**-41.6%**

6.6.1.4 Safety and security

However, during first light (from 05:24 to 08:40) and night-time period (from 21:28 to 05:24), significant occupancy is also observed in secondary streets, indicating that for a substantial number of pedestrians, getting in and out of the area in the most efficient way serves a greater motivation than walking in the streets with the highest visibility (Figure 6-13). In addition, the location of transportation nodes has the most significant contribution to pedestrian flow volume, especially when there are no activities within an area. For example, John Prince's Street serves as a terminal for a range of different bus routes, and a constant higher occupancy is observed throughout the day. Nevertheless, its occupancy increases from last light to night-time and first light, indicating the need for 24hour transportation services to get in and out of the area. This finding is further supported since, during these periods, retail activities are not open. Therefore, the two main streets do not serve as a retail/ activity attractor, rather than a transportation service offering.

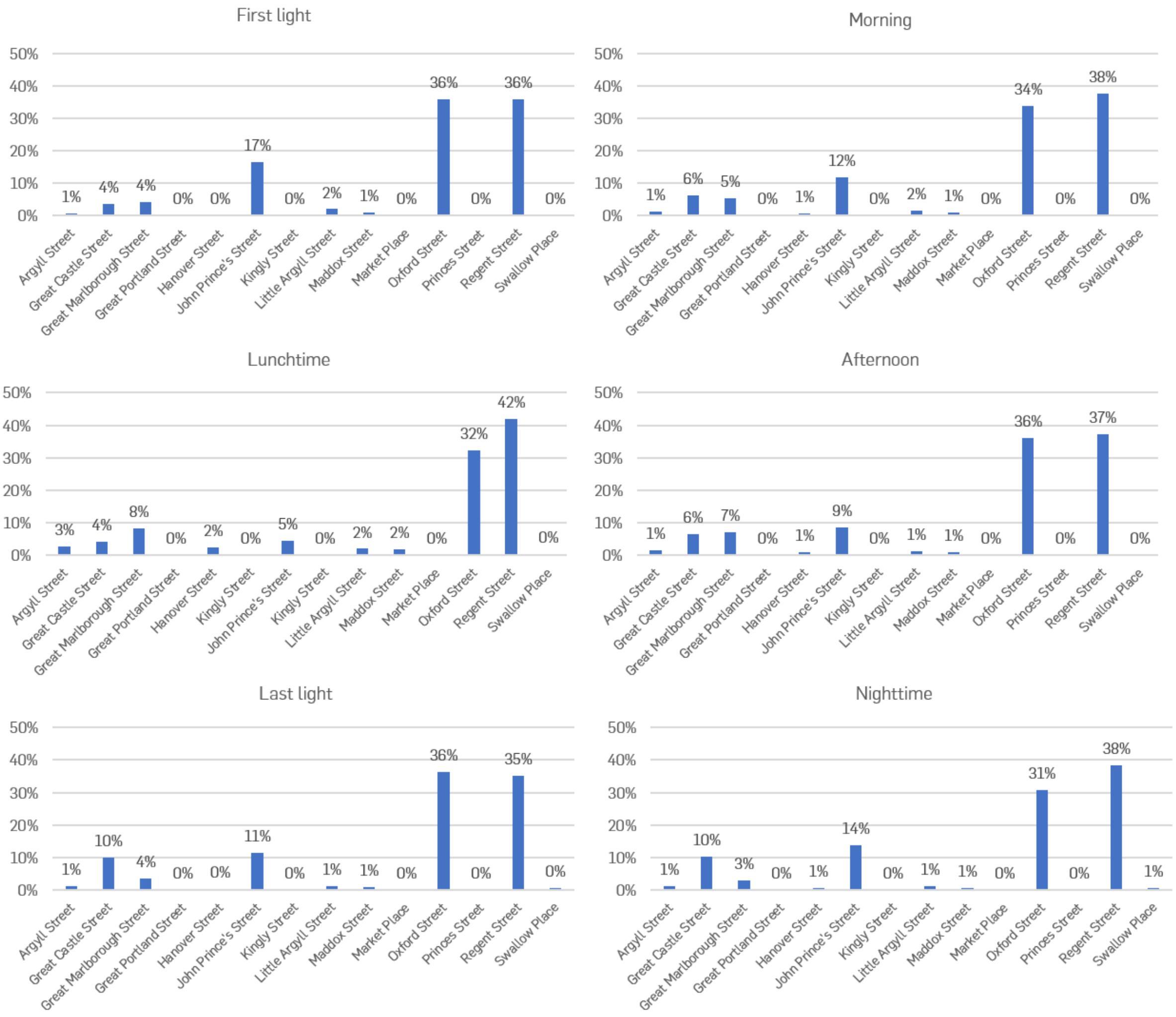
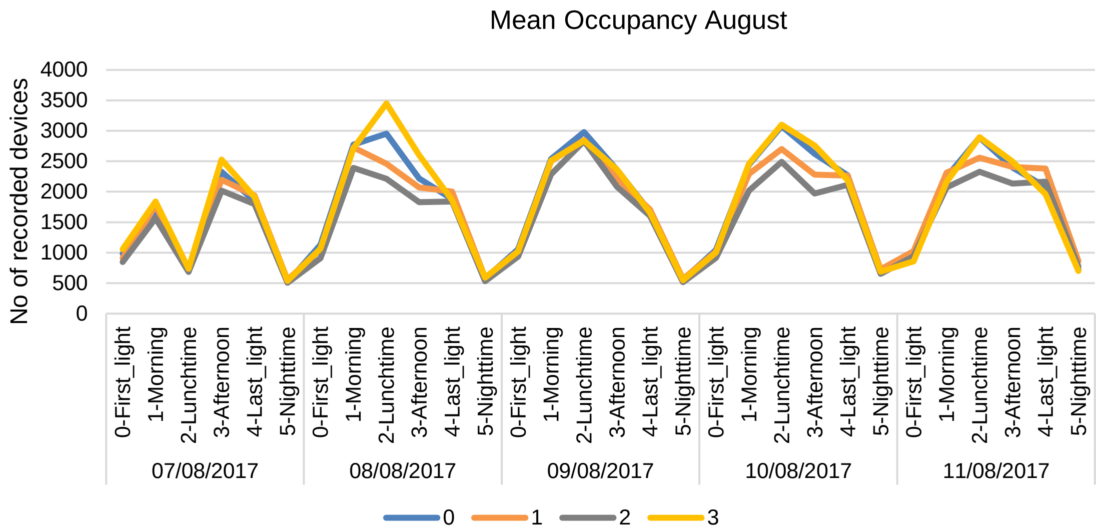


Figure 6-13 Example of street choice –normalised per hour in % of total occupancy– on the 7th of August (Monday).

6.6.1.5 Seasonality

Comparative analysis on occupancy patterns in August and October revealed that Clusters 1 and 2 are greatly influenced by the season, presenting an average reduction of 47% and 42%, respectively, in the number of recorded devices during October (Figure 6-14).



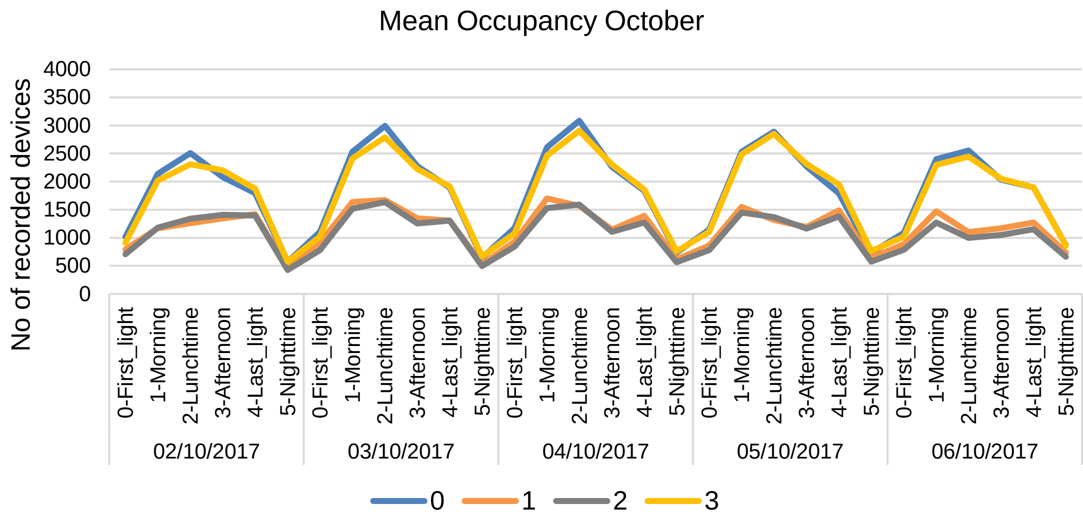
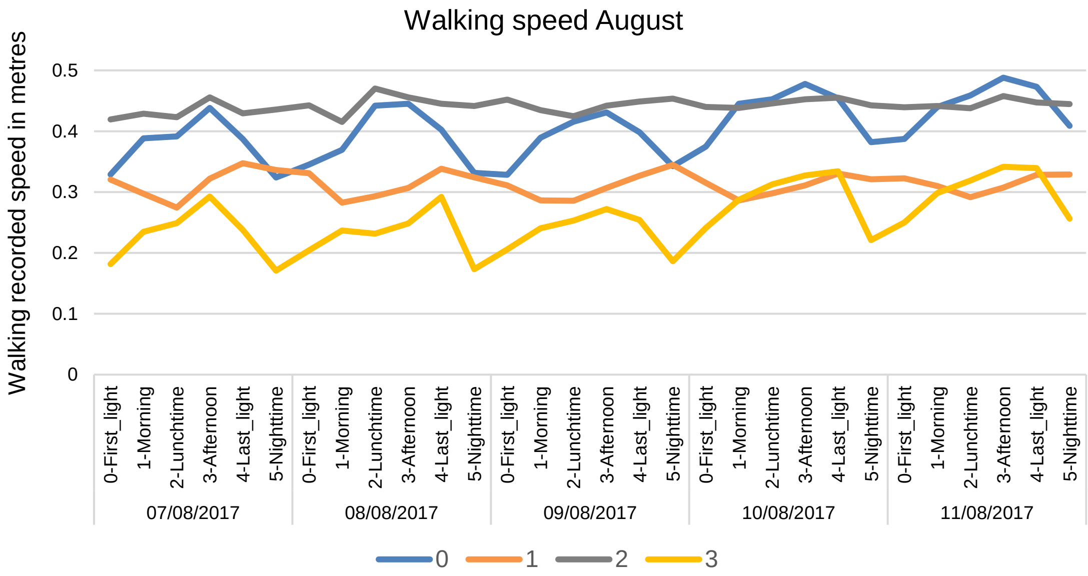


Figure 6-14 Mean recorded occupancy (number of recorded points) in a typical week for each cluster in August (up) and October (down).

Further analysis revealed that seasonality affects the individual walking characteristics, such as speed profiles, in which Cluster 1 and 2 have reduced mean walking speed in October. On the contrary, Clusters 0 and 3 showed an increase (Table 6-3). In addition, patterns differ within the day. For example, Cluster 1 has a stable speed profile throughout the day in August, while prominent peaks appear in October, in which speed increases from 20:47 until 08:24. For Cluster 2, though, although a slight decrease of 5% seems to take place in October, the speed profile is generally consistent (Figure 6-15). Cluster 0 has increased speed during afternoon time in August, while in October, these peaks appear to be after 20:47.

Table 6-3: Average walking speeds per cluster per month in a typical week (Monday to Friday)

	Clusters			
Month	**0**	**1**	**2**	**3**
August	0.394	0.291	0.420	0.235
October	0.436	0.269	0.401	0.246
Difference (%)	**+10%**	**-8%**	**-5%**	**+5%**



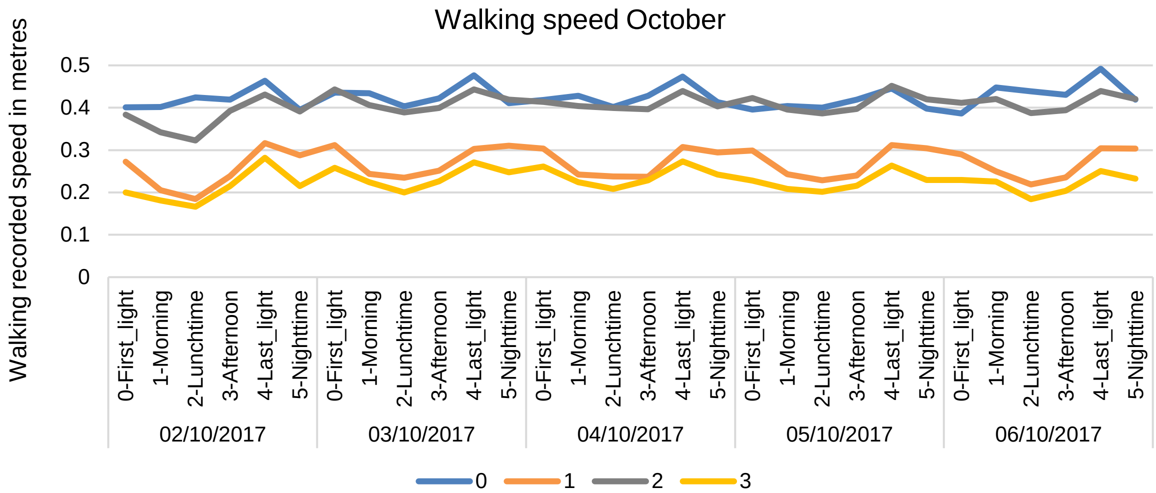


Figure 6-15 Mean recorded speed (m/s) in a typical week for each cluster found in August (up) and October (down).

Finally, all Clusters except from 2 showed a wider variance in their speed profiles, ranging from -0.15 to +0.15 m/s difference in August, while a greater consistency is observed in October. These findings indicate that seasonality has a significant effect on pedestrians' behaviours.

6.6.2 Spatial visibility dependence

This section investigates the relationship of spatial visibility against diverse parameters via GAM employment, aiming to address some of the complexity present to human behaviour, and the existence of non-linear relationships. Each plot (Figure 6-16 to Figure 6-18) shows a covariate and their partial dependence on space visibility in the context of the model. The y axis shows the mean number of observed spatial visibility, and the x-axis is the covariate interval. The red lines represent the 95% confidence interval.

Analysis of a typical week in August revealed that Cluster 3 is more likely to move in areas with increased visibility when the temperature is more than 17.5 ^{o}C, unlike the rest of the clusters, where visibility value decreases as the temperature rises (Figure 6-16). Clusters 0 and 1 are more likely to move in high visibility areas when the temperature is between 13 and 17 ^{o}C, while for Cluster 2, the higher the temperature, the lower the visibility. Results based on October (weekdays only) indicate that pedestrians seek highly visible areas when temperatures range from 12,5 to 18 ^{o}C (Figure 6-17).

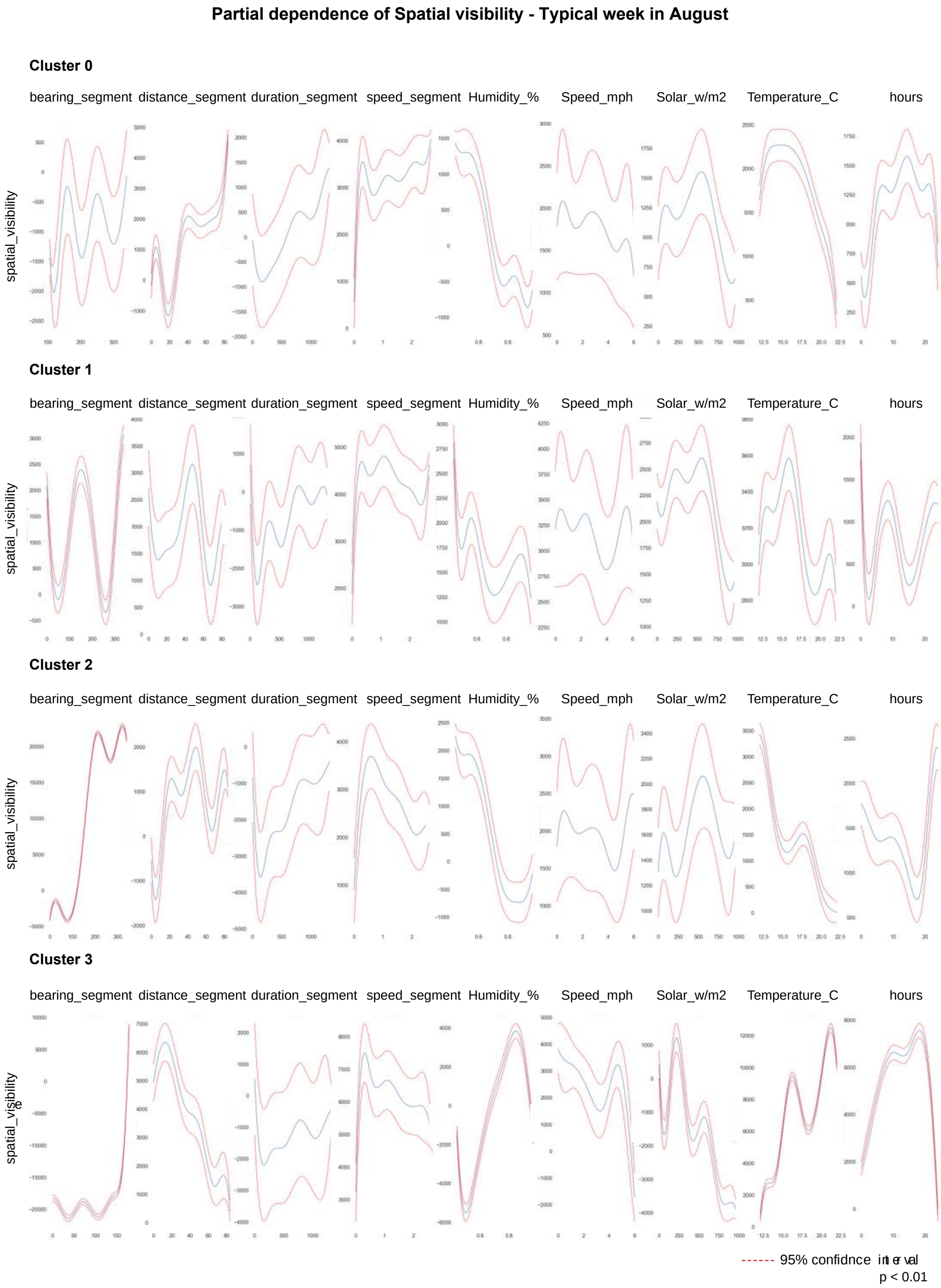


Figure 6-16 GAM feature effect of the variables for predicting the space visibility for a typical week in August (07th- 11th of August 2017).

Additionally, during weekdays, Cluster 0 and 3 (Expert behaviours) are mainly found in areas with great visibility from 08:00 until 20:00, with Cluster 0 showing the highest rates of predicted visibility during early afternoon hours (15:00) and Cluster 3 during late evening time (19:00) (Figure 6-17). During weekends, behaviours for Cluster 0 are altered, as they are found mainly in areas with high visibility in the afternoon and night-time hours (from 20:00 onwards) (Figure 6-17). The rest of the Clusters also differ, as the highest spatial visibility is predicted during the day, from 10:00 until 20:00, when it drops. The effect of seasonality is evident for Cluster 0, as predicted spatial visibility has no significant impact on their movement behaviour. For Cluster 1 and 2, similarities in their patterns are observed for August and October, where high predicted visibility values are observed during morning hours (before 10:00) and night-time periods (after 21:00). Cluster 3 appears to have a slight change in its patterns, as its peak seems to be during the early afternoon. However, the visibility scale has significantly dropped, from 8000 units in August to a maximum of 800 units in October, indicating the significance of visibility during summer months (Figure 6-17 and Figure 6-18).

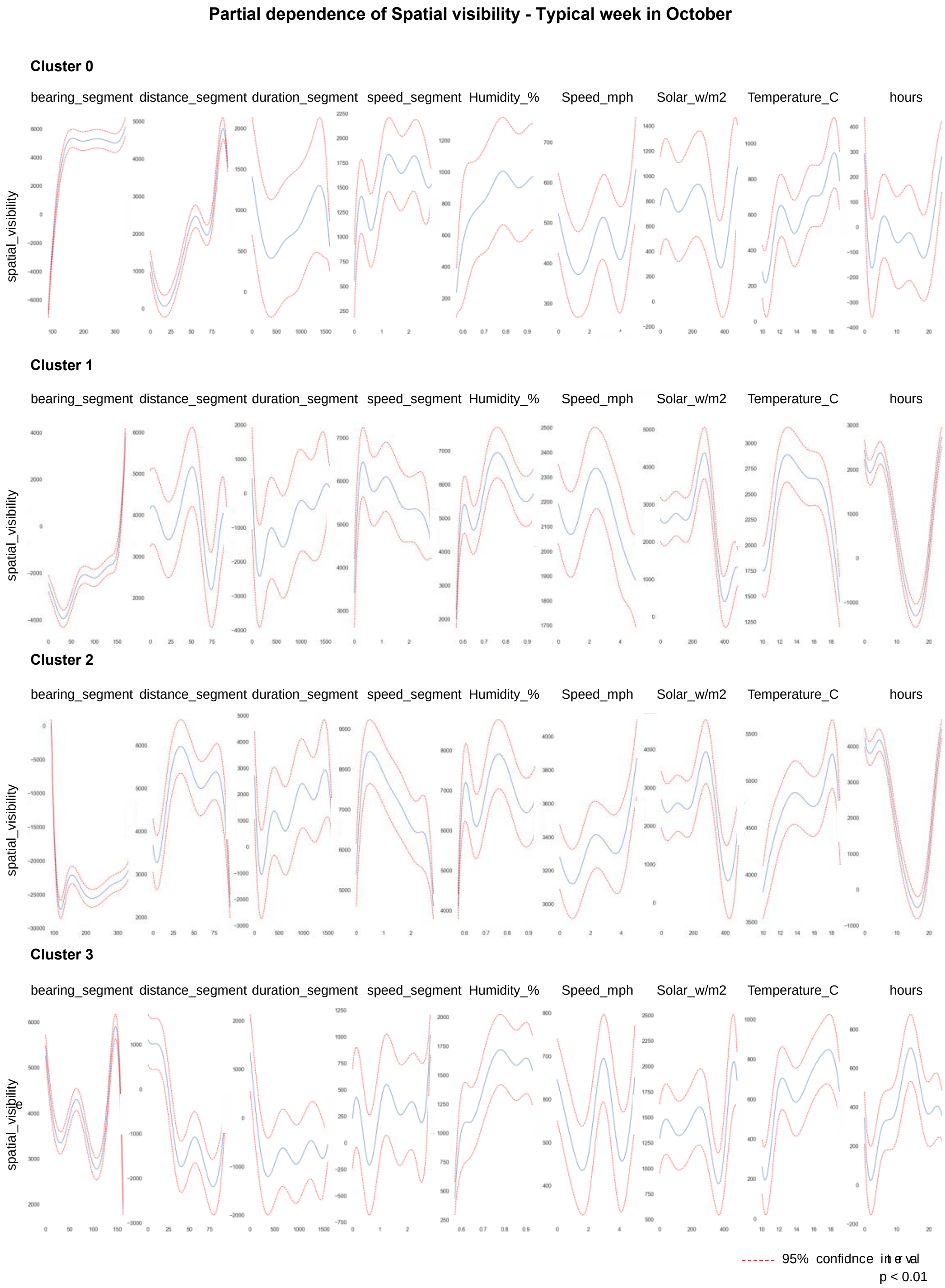


Figure 6-17 GAM feature effect of the variables for predicting the space visibility for a typical week in October (09th- 13th of October 2017).

Finally, speed and distances travelled are also influenced by spatial visibility. For Cluster 0, when a visibility value of more than 2000 is achieved, pedestrians are more likely to travel more than approximately 50m during August (Figure 6-16). However, their speed slightly increases as the predicted spatial visibility augments. The opposite is observed for Cluster 3, where high space visibility positively affects distances less than 25m. As predicted, spatial visibility decreases the speed increases (Figure 6-16). Similar speed profiles are observed for Clusters 1 and 2, although distances to be travelled present different predicted profiles. For Cluster 1, spatial visibility has a positive effect for distances higher than 25m, while it negatively affects distances higher than 50m or lower than 25m. For Cluster 2, though, predicted distances travelled increase as spatial visibility increases, while it drops for distances more than 50m and less than 75m. Seasonality does not have a significant effect, as similar patterns with slight differences appear for all clusters during these periods (Figure 6-17 and Figure 6-18). This finding suggests that spatial visibility does not influence the differences found in the speed profiles (Figure 6-16).

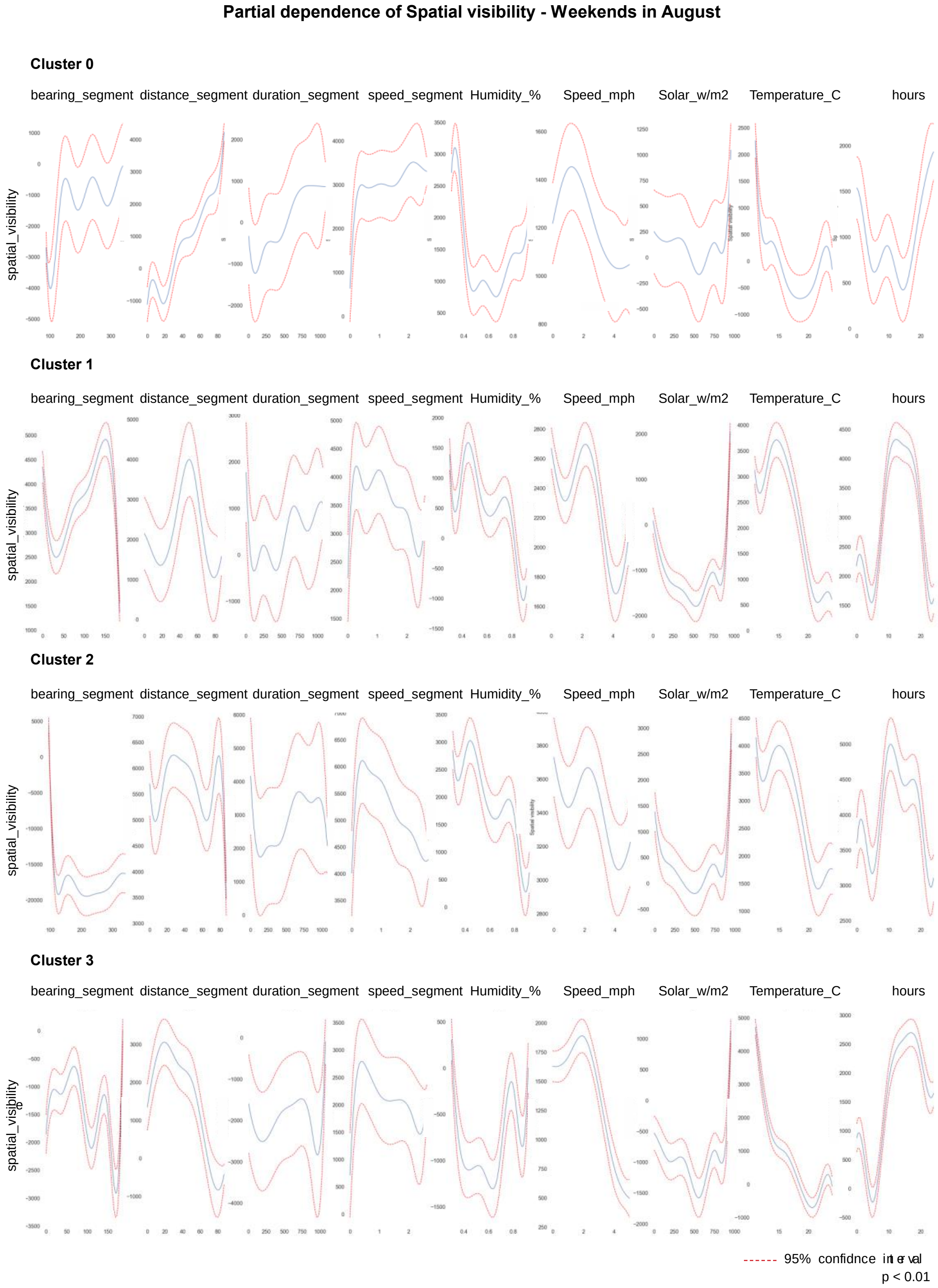


Figure 6-18 GAM feature effect of the variables for predicting the space visibility for weekends in August (05- 06th and 12th – 13th of August 2017).

6.7 Discussion

Urban design guidelines and studies are dependent on expert knowledge, existing literature, and experience rather than data-driven innovations (Choi, 2012). Some studies attempted to test factors related to occupancies and degree of spatial use (De Arruda Campos, et al., 2003). However, most of the factors considered are based on existing literature, often lacking data evidence, or relying on traditional collection and analysis techniques (Whyte, 1980; Feng, et al., 2021). Other researchers reviewed pedestrian behaviours in urban space utilising data-driven innovations. However, the majority of these studies were focused on emergency and evacuation scenarios or crowd behaviours (Wirz, et al., 2013; Shi & Abdel-Aty, 2015; Moreira & Ferreira, 2016; Martín, et al., 2019), leaving a gap in the exploration of sensorial navigation of space.

Due to the complexity of human behaviour, the way the built environment influences the walking experience differs based on the types of behaviours present in space. Results indicated that the built environment is not the only influencing parameter. Findings revealed that the environmental parameters might heavily influence the presence of specific behaviours in an urban environment, such as seasonality or weather, as these are related to the personal experience of a space. Another implication is that results indicated that seasonal changes do not solely reflect weather differences. Instead, seasonality may refer to the influences of the sensorial experiences of a place due to the seasonal context and state of mind. These findings imply that specific year periods attract novice behaviours, with the attractors relating to both weather impacts, such as higher temperatures and holiday seasons.

Additionally, the findings revealed that the spatial structure of urban environments plays a significant role in how urban densities are distributed in a city. Urban connectivity is correlated with movement densities. Results indicated that reduced visibility and increased walking distance between transportation links negatively influence solely novice behaviours (Clusters 1 and 2), suggesting how urban configuration and wayfinding can undermine specific behaviours found in urban space. This finding implies that street segments constituting the primary skeleton, holding the street network together (Oxford and Regent streets) are strongly associated with how people navigate urban space. Traditional models of movement consider distance and walking time; however, aspects of legibility and coherence of urban form are neglected. Integrating information of intelligibility, such as spatial visibility, can lead to enhanced urban form and function models. However, it should be noted that the effect of spatial structure is not the determining factor of pedestrian volume. Instead, it influences how pedestrian volume is distributed.

Various studies have already highlighted the importance of the availability of non-residential destinations, such as schools and transit stations (Cervero, R., 2002; Lee, et al., 2013) or retail functions (Pawsey, 1985; Pushkarev & Zupan, 1975; Barton, et al., 2003). Analysis results revealed that the location of transportation nodes is a significant pedestrian flow contributor, especially when other activities within an area are not active at a specific time within the day. Nevertheless, commercial activities on the ground floor level, such as shops and cafes, are more stimulating to the pedestrian, attracting people from the immediate surroundings. At the same time, transportation nodes offer accessibility to these uses from the wider area. From the analysis, though, it is evident that the type and characteristics of these retail uses can heavily influence pedestrian volumes recorded. Strategic design of the ground floor at a road segment scale is critical for urban vitality and sustainability (Gehl, 2010); therefore, retail activities should aim for an equal distribution of those supporting “local” users and “visitors”.

Results derived from GAM analysis revealed that both individual characteristics, such as time spent in an area or walking speed, and environmental parameters, such as weather and seasonality, are influenced by spatial visibility. In addition, results indicated that spatial visibility does not influence variances found in the speed profiles. Instead, they are mainly driven by seasonality. These findings imply the importance of the environmental parameters on walking speed, while it indicates that thermally comfortable environments are critical to the enjoyment of urban spaces.

The results from the data analysis support the notion that urban environments should be designed to support different types of behaviours, rather than assuming that spaces within one category, such as retail areas, are fundamentally similar and that "*one size fits all*". This study proves that urban areas of similar type, e.g., retail high streets, have a wide range of user behaviour variations supported via different space configurations and activities. Key differences were mainly found between novice and expert behaviours, with knowledge of the area (memorability) for the first group being considered a key driver of movement behaviour. Therefore, it can be argued that novice users respond to new and complex environments by "reading" urban elements, and most of the captured information is coming from visual stimuli unless exogenous parameters occur, e.g., suggestions from local users or walking within a group. Experts have built substantial knowledge bases that affect how they organise and interpret such details, resulting in efficient information processing.

This study has sought to bridge existing knowledge gaps and provided evidence as to how spatial visibility influences walking. This study demonstrates how BDAs can be employed to perform spatio-temporal analysis of human walking behaviour in urban environments to extract insights into space usability. Additionally, the spatio-temporal analysis illustrated how practitioners and researchers could better visualize results to understand walking behaviours in diverse urban environments.

6.8 Conclusion

This study focuses on reviewing key parameters affecting pedestrian movement aiming to identify the way urban space attributes affect walking activities and types of pedestrian behaviour and, more specifically, the effect of spatial visibility. The authors utilised machine learning (ML) algorithms and multivariate analysis on location data derived from Wi-Fi tracking techniques. This study further expands existing knowledge by offering insights into complex relationships between urban design qualities and pedestrian movement, focusing on aspects that may hinder or support walking intentions.

Implications/ limitations were encountered due to a lack of qualitative information of people's opinions, and findings regarding other influencing parameters, such as aesthetics, are not conclusive. Thus, this finding highlights the need for capturing and analysing comprehensive high-volume individual behavioural data sets that can be mapped against objective movement data to draw conclusions. In addition, further limitations for data collection techniques concerning large-scale monitoring systems include the fact that the pedestrian data used in this study were restricted to a limited number of days, with weekend recordings only for a period in August. Furthermore, data collected via Wi-Fi tracking techniques do require pedestrians to use a mobile phone with the Wi-Fi settings turned on. Hence, it may introduce additional limitations regarding the representativeness and completeness of the recorded datasets.

Future research can focus on collecting and analysing additional datasets qualitative information mapping against objective pedestrian movement data to enable further investigation on the subjective attributes of urban environments, such as aesthetics. Understanding the effects of the different physical attributes and environmental factors on pedestrian movement can lead to more accurate and robust models that decision-makers in urban design processes can apply. Such approaches could provide insights into pedestrians' navigation in geo-temporal human behaviours and the aggregation of the urban planning process within the building design sector industry. Realising how to design urban spaces better attuned to end-users' needs has become of great importance in post-COVID re-evaluation of urban spaces, as they can enable effective planning and response.

REFERENCES

Özbil, A., Gurleyen, T., Yesiltepe, D. & Zunbuloglu, E., 2019. Comparative Associations of Street Network Design, Streetscape Attributes and Land-Use Characteristics on Pedestrian Flows in Peripheral Neighbourhoods. *International Journal of Environmental Research and Public Health,* 16(10), p. 1846.

Accuware Inc, 2017. *Accuware.* [Online]
Available at: https://accuware-inc.com/
[Accessed 19 08 2021].

Ahrné, K., Bengtsson, J. & Elmqvist, T., 2009. Bumble bees (Bombus spp) along a gradient of increasing urbanization. *PLoS ONE,* 4(5).

Alfonzo, M., Boarnet, M.G., Day, K., Mcmillan, T. & Anderson, C.L., 2008. The Relationship of Neighbourhood Built Environment Features and Adult Parents' Walking. *Journal of Urban Design,* 13(1), pp. 29-51.

Angelelli, F., Morrow, J. & Greenwood, C., 2018. *The potential application of Wi-Fi data in the development of agent based pedestrian models.* Dublin, Ireland, The Association for European Transport.

Asami, Y., Kubat, A., Kitagaw, K. & IIda, S., 2002. *Introducing the third dimension on space syntax: application on the historical Istanbul.* Tokyo, Center for Spatial Information Science, University of Tokyo.

Barton, H., Grant, M. & Guise, R., 2003. *Shaping neighbourhoods. A guide for health, sustainability and vitality.* London: Spon Press.

Barton, N., Farewell, T. & Hallett, S., 2020. Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *Clean Water,* Volume 3, Article 13.

Batty, M., Jiang, B. & Thurstain-Goodwin, M., 1998. *Local movement: agent-based models of pedestrian flows.,* London, UK: Centre for Advanced Spatial Analysis (UCL).

Bohannon, R., 1997. Comfortable and maximum walking speed of adults aged 20-79 years: Reference values and determinants.. *Age and Ageing,* 26(1), p. 15–19.

Bohl, C., 2002. *Place Making.* Washington, DC: Urban Land Institute.

Bosselmann, P., 1998. A tribute to the work of Jan Gehl and Lars Gemzøe.. *Places,* 12(1), pp. 29-31.

Braaksma, J. & Cook, W., 1980. Human Orientation in Transportation Terminals. *Transportation Engineering Journal,* Volume 106, pp. 189-203.

Canter, D. & Tagg, S., 1977. Distance estimation in cities. *Environment and Behavior,* 7(1), pp. 59-80.

Carr, L., Dunsiger, S. & Marcus, B., 2011. Validation of walk score for estimating access to walkable amenities. *British Journal of Sports Medicine,* 45(14), pp. 1144-1148.

Cervero, R., 2002. Built Environments and Mode Choice: Toward a Normative Framework. *Transportation Research Part D: Transport and Environment,* 7(4), p. 265–284.

Cervero, R. & Duncan, M., 2003. Walking, Bicycling, and Urban Landscapes: Evidence From the San Francisco Bay Area. *Am J Public Health,* 93(9), p. 1478–1483.

Chan, C. & Ryan, D., 2009. Assessing the effects of weather conditions on physical activity participation using objective measures. *International Journal of Environmental Research and Public Health,* 6(10), p. 2639–2654.

Choi, E., 2012. *Walkability as an Urban Design Problem: Understanding the activity of walking in the urban environment (Licentiate dissertation).* [Online] Available at: Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-102182
[Accessed 2021].

Clifton, K., Smith, A. & Rodriguez, D., 2007. The development and testing of an audit for the pedestrian environment. *Landscape and Urban Planning,* 80(1-2), pp. 95-110.

Cornell, E., Sorenson, A. & Mio, T., 2003. Human Sense of Direction and Wayfinding. *Annals of the Association of American Geographers,* 93(2), p. 399–425.

Crane, R. & Crepeau, R., 1998. Does Neighborhood Design Influence Travel? A Behavioral Analysis of Travel Diary and GIS Data. *UC Irvine: Center for Activity Systems Analysis.*

Dalton, R., 2003. The secret is to follow your nose: Route path selection and angularity. *Environment and Behavior,* 35(1), pp. 107-131.

Davies, J., 2007. *Yellow Book: A prototype wayfinding system for London,* London: Transport for London by Applied Information Group.

De Arruda Campos, M. B., Chiaraida, A., Smith, A., Stonor, T. & Takamatsu, S., 2003. *Towards a 'walkability index'.* Strasbourg, France, Association for European Transport (AET).

DepthmapX development team, 2017. *depthmapX (Version 0.6.0) [Computer software].* [Online]
Available at: https://github.com/SpaceGroupUCL/depthmapX/
[Accessed 17 10 2021].

Dridi, M., 2015. Simulation of High-Density Pedestrian Flow: A Microscopic Model. *Open Journal of Modelling and Simulation,* Volume 3, pp. 81-95.

Edwards, P. & Tsouros, A., 2008. *A healthy city is an active city: a physical activity planning guide.* Copenhagen, Denmark: WHO Regional Office for Europe.

Ewing, R. & Handy, S., 2009. Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. *Journal of Urban Design,* 14(1), pp. 65-84.

Fasiolo, M., Nedellec, R., Goude, Y. & Wood, S., 2019. Scalable visualization methods for modern generalized additive models. *Journal of Computational and Graphical Statistics,* 29(1), pp. 78-86.

Feng, Y., Duives, D., Daamen, W. & Hoogendoorn, S., 2021. Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment,* Volume 187, Article 107329.

Fernandez-Ares, A., Garcia-Sanchez, P., Arenas, M. G., Mora, A. M. & Castillo-Valdivieso, P. A., 2020. Detection and Analysis of Anomalies in People Density and Mobility through Wireless Smartphone Tracking. *IEEE Access,* Volume 8, pp. 54237-54253.

Forsyth, A., Oakes, J., Lee, B. & Schmitz, K., 2009. The built environment, walking, and physical activity: Is the environment more important to some people than others?. *Transportation Research Part D: Transport and Environment,* 14(1), p. 42–49.

Frank, L., Sallis, J. & Saelens, B., 2010. The development of a walkability index: Application to the neighborhood quality of life study.. *British Journal of Sports Medicine,* 44(13), pp. 924-933.

Fruin, J., 1971. *Pedestrian Planning and Design.* New York: Metropolitan Association of Urban Designers and Environmental Planners.

Gärling, T., Böök, A. & Lindberg, E., 1986. Spatial orientation and wayfinding in the designed environment: A conceptual analysis and some suggestions for postoccupancy evaluation. *Journal of Architectural and Planning Research,* 3(1), p. 55–64.

Gath-Morad, M., Thrash, T., Schicker, J., Hölscher, C., Helbing, D. & Melgar, L.E.A., 2021. Visibility matters during wayfinding in the vertical. *Sci Rep,* Volume 11, p. 18980.

Gehl, J., 2010. *Cities for People.* Washington, DC: Island Press.

Gehl, J., 2011. *Life Between Buildings: Using Public Space.* Copenhagen, Denmark: The Danish Architectural Press.

Gehl, J. & Svarre, B., 2013. *How to Study Public Life.* Washington, DC: 2nd ed. Island Press.

Gibson, E., 1988. Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology,* 39(1), pp. 1-42.

Gibson, J., 1979. *The Ecological Approach to Visual Perception.* Boston, USA: Routledge.

Goldstein, A., Kapelner, A., Bleich, J. & Pitkin, E., 2015. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *Journal of Computational and Graphical Statistics,* 24(1), p. 44–65.

Gottdiener, M. & Lagopoulos, A., 1986. *The City and the Sign.* New York Chichester, West Sussex: Columbia University Press.

Hajrasouliha, A. & Yin, L., 2015. The impact of street network connectivity on pedestrian volume. *Urban Studies,* 52(13), pp. 2483-2497.

Handy, L., 1996. Urban Form and Pedestrian Choices: Study of Austin Neighborhoods. *Transportation Research Board,* 1552(1).

Han, J., Kamber, M. & Pei, J., 2012. *Data Mining: Concepts and Techniques. A volume in The Morgan Kaufmann Series in Data Management Systems.* Waltham, USA: Elsevier.

Hart, C., Stachow, G., Rafiq, M. & Angus, L., 2014. *The customer experience of town centres,* Loughborough: Loughborough University.

He, S., 2004. *Generalized Additive Models for Data With Concurvity: Statistical Issues and a Novel Model Fitting Approach.* Pittsburgh: University of Pittsburgh.

Hillier, B., 1996. *Space is the Machine.* Cambridge: Cambridge University Press.

Hillier, B., 2009. *The genetic code for cities – is it simpler than we thought?.* Delft, Netherlands, University of Delft.

Hillier, B. & Hanson, J., 1984. *The Social Logic of Space.* Cambridge: Cambridge University Press.

Hillier, B., Penn, A., Hanson, J., Garjewski, T. & Xu, J., 1993. Natural Movement: Or, Configuration and Attraction in Urban Pedestrian Movement. *Environment and Planning B: Planning and Design,* Volume 20, p. 29 – 66.

Hillier, B. & Sahbaz, O., 2011. Safety in Numbers: High-resolution Analysis of Crime in Street Networks. In: *The Urban Fabric of Crime and Fear.* Dordrecht: Springer, pp. 111-137.

Hong, J., 2016. How does the seasonality influence utilitarian walking behaviour in different urbanization settings in Scotland?. *Social Science & Medicine,* Volume 162, pp. 143-150.

Humpel, N. et al., 2004. Perceived environment attributes, residential location, and walking for particular purposes. *American Journal of Preventive Medicine,* Volume 26, pp. 119-125.

Ingold, T., 2000. *The Perception of the Environment: Essays on Livelihood, Dwelling and Skill.* London: Routledge.

Jacobs, J., 1961. *The Death and Life of Great American Cities.* New York: Random House..

Kamal, R. & Saxena, P., 2019. *BIG DATA ANALYTICS: Introduction to Hadoop, Spark, and Machine-Learning.* s.l.:McGraw-Hill Education.

Karbovskii, V., Severiukhina, O., Derevitskii, I., Voloshin, D., Presbitero, A. & Lees, M., 2019. The impact of different obstacles on crowd dynamics. *Journal of Computational Science,* Volume 36, Article 100893.

Ki, D. & Lee, S., 2021. Analyzing the effects of Green View Index of neighborhood streets on walking time using Google Street View and deep learning. *Landscape and Urban Planning,* Volume 205.

Klingsch, W., Rogsch, C., Schadschneider, A. & Schreckenberg, M., 2008. *Pedestrian and Evacuation Dynamics 2008.* Verlag Berlin Heidelberg: Springer.

Kubba, S., 2017. *Handbook of Green Building Design and Construction: LEED, BREEAM, and Green Globes.* Second edition ed. s.l.:Butterworth-Heinemann.

Lee, C., Zhu, X., Yoon, J. & Varni, J., 2013. Beyond Distance: Children's School Travel Mode Choice. *Annals of Behavioral Medicine,* 45(1), p. 55–67.

Liao, H., Dong, W., Peng, C. & Liu, H., 2016. Exploring differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers. *Cartography and Geographic Information Science,* 44(6), pp. 474-490.

Li, R. & Klippel, A., 2012. Wayfinding in libraries: Can problems be predicted?. *Journal of Map & Geography Libraries,* 8(1), pp. 21-38.

Li, S., Blythe, P., Zhang, Y., Edwards, S., Xing, J., Guo, W., Ji, Y., Goodman, P. & Namdeo, A., 2021. Should older people be considered a homogeneous group when interacting with level 3 automated vehicles?. *Transportation Research Part F,* Volume 78, pp. 446-465.

Lynch, K., 1960. *The Image of the City.* Cambridge: MIT Press, ISBN-13: 9780262620017, ISBN 0262620014.

Martín, J., Khatib, E.J., Lázaro, P. & Barco, R., 2019. Traffic Monitoring via Mobile Device Location. *Sensors,* 19(20), p. 4505.

Matthews, C., Freedson, P.S., Hebert, J.R., Stanek, E.J., Merriam, P.A., Rosal, M.C., Ebbeling, C.B. & Ockene, I.S., 2001. Seasonal variation in household, occupational, and leisure time physical activity: longitudinal analyses from the seasonal variation of blood cholesterol study. *American Journal of Epidemiology,* 153(2), pp. 172-183.

Mehta, V., 2009. Look closely and you will see, listen carefully and you will hear: Urban design and social interaction on streets. *Journal of Urban Design,* 14(1), pp. 29-64.

Mendiola, L. & González, P., 2021. Urban Development and Sustainable Mobility: A Spatial Analysis in the Buenos Aires Metropolitan Area. *Land,* 10(2), p. 157.

Merchant, A., Dehghan, M. & Akhtar-Danesh, N., 2007. Seasonal variation in leisure time physical activity among Canadians. *Can. J. of Public Health,* Volume 98, pp. 203-208.

Mercieca, J., Kaparias, I., Bell, M. & Finch, E., 2011. *Integrated street design in high-volume junctions: the case study of London's Oxford Circus.* Athens, Greece, City Research Online.

Montello, D., 2005. Navigation. In: *P. Shah (Ed.) & A. Miyake, The Cambridge Handbook of Visuospatial Thinking.* s.l.:Cambridge University Press, p. 257–294.

Moreira, F. & Ferreira, M., 2016. Teaching and learning requirement engineering based on mobile devices and cloud: a case study. In: D. Fonseca & E. Redondo, eds. *Handbook of Research on Applied E-Learning in Engineering and Architecture Education.* Pennsylvania, USA: IGI Global, p. 1190–1217.

Mumford, C., Parker, C., Ntounis, N. & Dargan, E., 2021. Footfall signatures and volumes: Towards a classification of UK centres. *Environment and Planning B: Urban Analytics and City Science,* 48(6), pp. 1495-1510.

Nakamura, K., 2016. The spatial relationship between pedestrian flows and street characteristics around multiple destinations. *IATSS Research,* Volume 39, p. 156–163.

Noland, R. & Ishaque, M., 2006. Smart bicycles in an urban area: evaluation of a pilot scheme in London. *Journal of Public Transportation,* 9(5), pp. 71-95.

Ordnance Survey, 2021. *Points of Interest.* [Online]
Available at: https://www.ordnancesurvey.co.uk/business-government/products/points-of-interest
[Accessed 15 10 2021].

Orellana, D. & Wachowicz, M., 2011. Exploring patterns of movement suspension in pedestrian mobility. *Geographical Analysis,* 43(3), pp. 241-260.

Othman, F., Yusoff, Z. & Salleh, S., 2020. Assessing the visualization of space and traffic volume using GIS-based processing and visibility parameters of space syntax. *Geo-spatial Information Science,* 23(3), pp. 209-221.

Parvin, A., Min, A. & Beisi, J., 2008. Effect of visibility on multilevel movement: a study of the high-density compact built environment in Hong Kong. *Urban Des Int,* Volume 13, p. 169–181.

Parvin, A., Ye, A. & Jia, B., 2006. *Visual Accessibility and Pedestrian Movement: A Study of the Compact Spatial Environment in Hong Kong.* Guangzhou, China, UNESCO.

Pawsey, M., 1985. *The pedestrian and his environment.* Brunswick, Australia, Institution of Engineers.

Pushkarev, B. & Zupan, J., 1975. *Urban space for pedestrians: A report of the Regional Plan Association.* Cambridge: The MIT Press.

QGIS Python Plugins Repository, 2020. *QNEAT3.* [Online] Available at: https://plugins.qgis.org/plugins/QNEAT3/ [Accessed 15 10 2021].

Qiu, L., Lindberg, S. & Nielsen, A., 2013. Is biodiversity attractive?—On-site perception of recreational and biodiversity values in urban green space. *Landscape and Urban Planning,* Volume 119, pp. 136-146.

Ramsay, T., Burnett, R. & Krewski, D., 2003. The Effect of Concurvity in Generalized Additive Models Linking Mortality to Ambient Particulate Matter. *Epidemiology,* 14(1), p. 18.

Ross, N., 2019. *2- Interpreting and Visualizing GAMs.* [Online] Available at: https://noamross.github.io/gams-in-r-course.

Sadalla, E. & Nagel, S., 1980. The perception of traversed distance. *Environ. Behav.,* Volume 12, pp. 65-79.

Servén, D. & Brummitt, C., 2018. *pyGAM: Generalized Additive Models in Python.* [Online]
Available at: https://zenodo.org/record/1208724#.YkBlSi1Q30o
[Accessed 27 03 2022].

Shamsuddin, S., Abdul Rahman, N. & Sulaiman, A., 2010. *How walkable is our city? Its influence in creating sustainable city centre design.* Malaysia, Universiti Sains Malaysia.

Shi, Q. & Abdel-Aty, M., 2015. Big Data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies,* Volume 58, pp. 380-394.

Sianoja, M., Syrek, C.J., de Bloom, J., Korpela, K. & Kinnunen, U., 2018. Enhancing daily well-being at work through lunchtime park walks and relaxation exercises: Recovery experiences as mediators. *Journal of Occupational Health Psychology,* 23(3), pp. 428-442.

Smith, T., Nelischer, M. & Perkins, N., 1997. Quality of an urban community: a framework for understanding the relationship between quality and physical form. *Landscape and Urban Planning,* Volume 39, pp. 229-241.

Southworth, M., 2005. Designing the Walkable City. *Journal of Urban Planning and Development,* 131(4), pp. 246-257.

Tibbalds, F., 2001. *Making People Friendly Towns: Improving the Public Environment in Towns and Cities.* London: London Spon Press.

Transport for London, 2011. Regent Street - Consolidation and collaboration. London. TRID. (1971). A rational urban cartage system. *Transportation Research Board,* 11(10), pp. 15-39.

Turner, A. & Penn, A., 1999. *"Making Isovists Syntactic: Isovist Integration Analysis.* Universidad de Brasilia, Brazil, Space syntax network.

Turner, E. & Giannopoulos, D., 1974. Pedestrianisation: London's Oxford Street experiment. *Transportation,* Volume 3, p. 95–126.

Van Nes, A. & Stolk, E., 2012. *Degrees of sustainable location of railway stations: Integrating space syntax and node place value model on railway sations in the province of North olland's strategic plan for 2010-2040.* Santiago de Chile, Chile, s.n.

Van Veldhoven, K., Kiss, A., Keski-Rahkonen, P., Robinot, N., Scalbert, A., Cullinan, P., Chung, K., Collins, P., Sinharay, R., Barratt, B.M., Nieuwenhuijsen, M., Rodoreda, A.A., Carrasco-Turigas, G., Vlaanderen, J., Vermeulen, R. & Portengen, L., 2019. Impact of short-term traffic-related air pollution on the metabolome - Results from two metabolome-wide experimental studies.. *Environment international,* Volume 123, p. 124–131.

Vasilikou, C. & Nikolopoulou, M., 2020. Outdoor thermal comfort for pedestrians in movement: thermalwalks in complex urban morphology. *International Journal of Biometeorology,* Volume 64, p. 277–29.

Wang, W., Lo, S. & Liu, S., 2017. A cognitive pedestrian behavior model for exploratory navigation: Visibility graph based heuristics approach. *Simulation Modelling Practice and Theory,* Volume 77, pp. 350-366.

Waskom, M., 2021. seaborn: statistical data visualization. *Journal of Open Source Software,* 6(60), p. 3021.

Weisman, J., 1981. Evaluating Architectural Legibility: Way-Finding in the Built Environment. *Environment and Behavior,* 13(2), pp. 189-204.

Weisman, J., 1981. Evaluating Architectural Legibility: Way-Finding in the Built Environment. *Environment and Behavior,* 13(2), pp. 189-204.

Western Australian Planning Commission, 1997. *Liveable Neighbourhoods Community Design Code,* Perth: Western Australian Planning Commission.

Western Australian Planning Commission, 2007. *Liveable Neighbourhoods Community Design Code: A Western Australian Government Sustainable Cities Initiative,* Perth: Western Australian Planning Commission.

Whyte, W. H., 1980. *The Social Life of Small Urban Spaces, Project for public spaces.* Washington, D.C.: Conservation Foundation.

Wiener, J., Büchner, S. & Hölscher, C., 2009. Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Computation ,* 9(2), pp. 152-165.

Wiley, M. & Wiley, J., 2019. GAMs. In: *Advanced R Statistical Programming and Data Models.* s.l.:Apress, p. 165–224.

Wirz, M., Franke, T., Roggen, D., Mitleton-Kelly, E., Lukowicz, P. & Tröster, G., 2013. Probing crowd density through smartphones in city-scale mass gatherings. *EPJ Data Science,* 2(1), pp. 1-24.

Wood, S., 2006. *Generalized Additive Models: An Introduction with R.* Boca Raton, Florida: Chapman and Hall/CRC.

Yuan, C. & Yang, H., 2019. Research on K-value selection method of K-means clustering algorithm. *J,* Volume 2, p. 226–235.

Zacharias, J., 2001. Pedestrian behavior and perception in urban walking environments. *Journal of Planning Literature,* 16(1), pp. 3-18.

Zaki, M. & Sayed, T., 2018. Automated Analysis of Pedestrian Group Behavior in Urban Settings. *IEEE Transactions on Intelligent Transportation Systems,* 19(6), pp. 1880-1889.

Zhang, N., Chen, F., Zhu, Y., Peng, H., Wang, J. & Li, Y., 2020. A Study on the Calculation of Platform Sizes of Urban Rail Hub Stations Based on Passenger Behavior Characteristics. *Mathematical Problems in Engineering,* Volume 7, pp. 1-14.

Zhixiang, F., Qingquan, L. & Shih-Lung, S., 2015. What about people in pedestrian navigation?. *Geo-spatial Information Science,* 18(4), pp. 135-150.

Zhou, Y., 2018. *Deterioration and Optimal Rehabilitation Modelling for Urban Water Distribution Systems.,* Delft: Delft University of Technology.

7 The challenges of implementing evidence-based strategies to inform building and urban design decisions: a view from current practice

7.1 Abstract[1]

Purpose: This study aims to raise awareness of the key challenges, opportunities, and priorities for evidence-based strategies' application to inform building and urban design decisions.

Design/methodology/approach: The study employs deductive qualitative content and manifest analysis, utilising semi-structured interviews undertaken with building and urban design professionals who represent a UK-based organisation.

Findings: The challenges associated with the practical implementation of frameworks, potential application areas, and perceived areas of concern have been identified. These not only include the need to practically test their use, but also to identify the most appropriate forums for their use. Participant responses indicate the need to further develop engagement strategies for their practical implementation, clearly communicating the benefits and efficiencies to all stakeholders.

Research limitations/implications: Implications/ limitations of this study come with the fact that some of the respondents may possess inadequate professional experience in properly evaluating all the questions. Additionally, the information gathered is restricted to the UK geographical context, as well as coming from one organisation, due to data accessibility.

Practical implications: The findings of the study can be adopted by designers in the strategic definition level to overcome key challenges associated with the use of evidence-based strategies, enhancing their decision-making processes.

[1] Abstract follows structure of the journal required format, as described in Section 1.8

Originality/value: As a theoretical contribution to knowledge, this study enhances the body of knowledge by identifying the challenges associated with the practical implementation of evidence-based strategies to inform building and urban design decisions. In practice, the findings aid urban planners, designers, and academics in embedding and adopting strategies that enhance decision-making processes.

7.2 Introduction and theoretical background

Building and urban design hold the prospect of enhancing quality of life for its users while also help in addressing global issues found in urban systems (Imants, et al., 2021; Chang, et al., 2020). Design process is considered as a creative problem, where its solving includes a series of steps or sequenced activities, that eventually can lead from the initial concept to realisation (Koberg & Bagnall, 1981; Ledbury, 2018). There are several interpretations as to what design process means, and several models have been proposed to divide the design process into steps, such as the “Double Diamond Model” or “Design Thinking Theory” (Ledbury, 2018; Meinel & Leifer, 2011). In practice, the stages of a design project are defined to help organise the process of briefing, design, and construction. These stages are described in the Royal Institute of British Architects Plan of Work (RIBA PoW) (RIBA, 2020).

Contemporary practice in building and urban design industry still relies on belief rather than evidence (Brown & Corry, 2011). However, these sources do not clearly articulate relevant planning and design research, nor does research heavily impacts practice. Both the profession and the discipline follow the culture of “non-reporting”, missing the opportunity of monitoring build projects to identify whether they have achieved their stated objectives, and thus, avoid mistakes while introducing design efficiencies (Ahern, 1999). Medicine is one of the first disciplines that has transitioned from utilising “theoretical foundations” to practicing evidence-based medicine (e.g., (Rosenberg & Donald, 1995)). However, several difficulties have already been identified in implementing evidence-based approaches, such as practitioner resistance to embrace it, lack of communication between stakeholders and academics, or a clear definition of what counts as evidence (Abruzzini & Abrishami, 2022; Stanitsas & Kirytopoulos, 2022).

Implementing and dissolving an innovative idea (innovation process) in building and urban design requires substantial resources and the role of private development (Forsyth, 2007). Innovation process can be defined as an iterative process aimed at the creation of new products, processes, knowledge, or services using new or even existing knowledge (Kusiak, 2009). Emergent technologies and new trends can create new opportunities and help in overcoming such challenges, but they also pose new challenges to designers (Semeraro, et al., 2021). More specifically, "Big Data" tools are recognised as a new generation of technologies, designed to extract value from enormous volumes and varieties of data, enabling capturing, analysis, and knowledge-discovery in high-velocities (Esteves & Curto, 2013). The evolving trend of big data-driven innovation is leading towards the development of data-driven commodities and services and can empower data-driven planning (Gahm, 2020). Big Data Approaches (BDAs) refer to the combination of diverse datasets and related technologies to extract insights on complex systems via novel organisational and analytical capabilities (Pollard, et al., 2018). Data-driven innovation (DDI) entails exploitation of any type of data in the innovation process to generate positive economic and social impacts (Jetzek, et al., 2014). Nevertheless, the lack of a systematic definition, a unified meaning of what "Big Data" is, shared amongst academia, industry, business, and media, is adding a mystery around its concept (Ward & Barker, 2013; Chen, et al., 2014). Due to the ambiguity of what it represents, its meaning has been converted into a "buzz" word (Power, 2014). The solution (final design product) should be based on evidence and knowledge, increasing the trustworthiness in decision-making (Fredriksson, 2017; Power, et al., 2019). To achieve this, practitioners should be allowed to use research results and evidence as a basis for their designs. Thus, there is a need for developing their own frameworks of reporting evidence in design.

This paper aims to raise awareness of the key challenges, opportunities, and priorities for evidence-based strategies' application to inform building and urban design decisions. The study employs deductive qualitative content and manifest analysis, utilising semi-structured interviews undertaken with building and urban design professionals who represent a United Kingdom (UK) based organisation. The originality of this study lies in the contribution to new knowledge by reviewing existing literature and revealing building and urban design professionals' views on the challenges associated with the practical implementation of frameworks. The significance of the findings is reflected in the highlighted potential application areas and perceived areas of concern for future development. The outcomes present opportunities for evidence-based strategies to be effectively followed by designers.

7.3 Research Methodology

7.3.1 Data collection via semi-structured interviews

Design professionals from a design, engineering and project management consultancy in UK were interviewed via a semi-structured approach. The study followed a deductive qualitative content analysis, where the interview guide was pre-defined based on previous research and the desire to explore topics based on these findings (Barton, et al., 2021; Neuendorf, 2016). The research conducted in Chapters 3 and 5, defined the structure of the data framework presented in Figure 7-3, while it informed the selection of participants, as described in Section 7.3.2.

The semi-structured interview approach was selected due to its benefits of encouraging participants to diverge and elaborate beyond a certain point, revealing additional information of relevance (Adams, 2015). There are several limitations when conducting semi-structured interviews, mainly associated with respondents fully comprehending the questions asked (Oltmann, 2016). Addressing this, the selected participants were practitioners with a depth of professional knowledge in the design of buildings and places. Furthermore, to avoid bias introduced by the interviewer, questions were asked in ways to avoid increased involvement in the discussion or further commenting (Adams, 2015; Ritchie, et al., 2003).

Interviews were recorded and transcribed with participants' consent obtained, and anonymity assurance was provided. This study also employs manifest analysis, utilising direct quotation of participants (Gopaldas, 2016), providing detailed understanding on a subject area (Bengtsson, 2016). Two pilot semi-structured interviews were completed prior the commencement on the main cohort of participants to validate the included questions, concepts, and language and to determine suitability of research instruments (Malmqvist, et al., 2019). The collection of data via semi-structured interviews was undertaken between August and October 2021. Each interview was restricted to approximately 45 minutes to ensure willingness of participation, fatigue reduction and to achieve a higher response rate (Barton, et al., 2021).

7.3.2 Stakeholder group

The stakeholders' group is composed by practitioners with experience in the fields of design and construction of buildings and places. The collected sample included 15 participants from one UK-based organisation, which is within the acceptable range of interviewees as defined by the literature (Bengtsson, 2016; Galvin, 2015). The number of participants is determined based on conclusive responses to the questions regarding the analysed themes and the sample has reached the point of information saturation (Ritchie, et al., 2003; Fusch & Ness, 2015; Frank, 2000). The study followed a convenience sampling through direct approach, targeting participants based on professional relativity and participation willingness (Brodaty, et al., 2014). This method was selected due to the benefits it presents, avoiding complications of dealing with a randomised sample and to obtain information and trends relevant to decision-making within the design process. Key criteria, as gathered from existing literature, were participant different levels of hierarchy, sector focus, and experience in DDI and BDAs (West, et al., 2008; Harris, 2012). The participant details are provided in Table 7-1.

Table 7-1: Participants of the semi-structured interviews

ID	Sector	Role	DDI or BDAs Experience
P1	Commercial & Workplace	Architect	Partial
P2	Education	Project/ Lead Architect	None
P3	Master planning & Urban Design	Technical Director	Beyond role
P4	Master planning & Urban Design	Associate Director	Partial
P5	Research & Innovation and Building information modelling	BIM Coordinator/ Digital Developer	Expert
P6	Rail	Associate Director	Beyond role
P7	Rail	Senior Architect	Partial
P8	Research & Innovation	Senior Design Researcher	Expert
P9	Commercial & Residential	Project Architect	Partial
P10	Building information modelling	Associate Director	Expert
P11	Master planning & Urban Design	Urban Designer	Partial
P12	Commercial & Residential	Architectural Assistant	None
P13	Education & Residential	Technical Director	Partial
P14	Rail	Architectural Assistant	Partial
P15	Education	Architect	None

7.3.3 Semi-structured interviews guide design

An interview protocol was developed, outlining the main questions, while several secondary questions occurred based on the initial responses to the main ones. These questions are structured based on the key themes outlined in Figure 7-1, divided further by sub-themes for a detailed understanding of the challenges and opportunities present in existing approaches and the future implementation of new frameworks. The key themes explored are: Existing Design process investigation, DDI & BDAs potential, and Practical application & Framework validation.

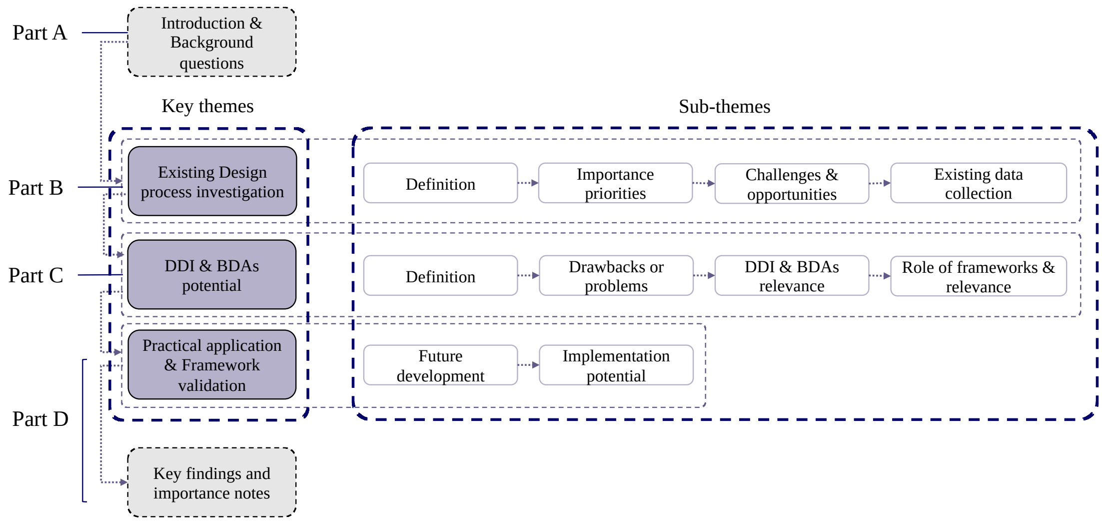


Figure 7-1: Systematic map of the key themes explored during the semi-structured interviews.

An initial introduction to the research and the format of the interview was provided to participants and several background questions were asked (Figure 7-1. Part A). The purpose of the first part of questions asked was to identify how involved the participants are with the design process and at which level they are using diverse types of data. Part A questions also allowed the later questions in Part B to be placed into the context of each participant's role (Figure 7-1. Part B). Furthermore, it allowed terminology used to be defined, ensuring participants fully acknowledge the content of the questions that will follow. The questions included in Part B, C and D were used to explore the key themes and investigated the potential of new BDA approaches to be implemented in practice, serving the aim of this paper. Further to this, some of the work produced as part of this research, was communicated via graphs, and it is further explained in the next sections (Figure 7-2 and Figure 7-3).

7.4 Results

7.4.1 Challenges and opportunities derived from existing design processes

The purpose of the semi structured interviews in Part B was to reveal the challenges and opportunities designers face in existing design process. Based on the responses received by the participants, results indicated an in-depth understanding of what "*design process*" means to them, with a great majority referring to an "*iterative*" process. Nevertheless, it was suggested that there is not one single methodology to be followed, rather it is just a narrative or a set of guidance documents, often used as advice. When discussed with participants, several suggested that key steps exist when rationalising the design process, as noted below:

*"*So*, it's really reflecting the work process that we traditionally follow in design. Well, in all the projects, it's data collection, then it's diagnostic, and then it's prototyping on the project and development to the project. Those are the four main steps. Nothing more specific."* [P4].

However, participants that formed part of the Rail sector referred to "*a scientific and structured process in the early stages of the projects, to come down to a single preferred option*" ([P7]), or to a "*heavily standardised and very prescriptive process that needs to follow client's design standards*" ([P6]). Similar comments were received by those involved with the Building Information Modelling (BIM) service line, highlighting that the end-goal of their service line is to streamline the design process, enabling designers "*to be a lot more prescriptive and linear and record the process in which various stakeholders delivered and developed their design.*" [P10]. Finally, one of the participants referred to the design process as “*pre-conceived*” and by further expanding “*There's a lot of repetition. There's not much flexibility in design.*” [P15].

Figure 7-2 describes the stages of design, research, and development process, spanning from the generic design stages up to detailed design tasks, as identified in literature (Ledbury, 2018; Meinel & Leifer, 2011). RIBA stages are mapped against each step for ease of interpretation of the participants. Following the presentation of Figure 7-2 to the participants, they were then asked to highlight the important parts of the design process and identify the parts that are not aligned with their internal design methodologies. Three participants commented on the presented process, highlighting key differences, while the rest of the participants felt that Figure 7-2 accurately represents the steps usually followed when designing a building or a place.

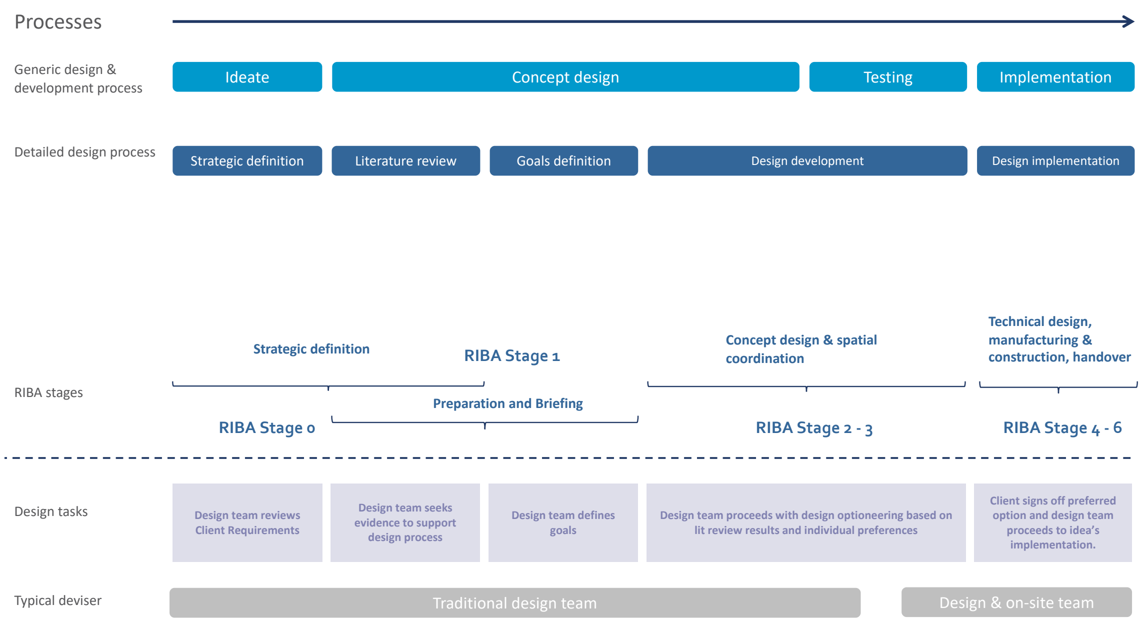


Figure 7-2: Design, research, and development stages.

More specifically, regarding the highlighted differences, [P3] felt that “*All the tasks fit within the strategic definition. Mini scheme on its own.*”, highlighting the iterative nature of the design process. [P1] felt that “*testing comes in tandem with concept design. You will never do the testing after you develop a concept design*.”, commenting on the need for a design validation process prior to a finalised concept design. Finally, [P4] identified a missing step in the beginning of that process; "*That's probably what's missing, as in understanding the context, understanding the client.".*

Majority of the participants felt that all the steps within the design process are equally important. However, a great majority of the participants indicated that the biggest impact of decision-making is in the initial stages of design. They highlighted its linear structure, indicating that the outcome of the first step (Figure 7-2- Ideate) will be carried on in the following stages of design, up to the implementation (Figure 7-2– Implementation). Others identified important aspects of the design process in need to be considered, rather than important steps. These can be divided into three themes: Resourcing, Information and Understanding, as noted by several participants:

"In strategic definition and briefing, there is sometimes a lack of knowledge of where you would get information or what type of information will be useful at that stage outside of people's experience". [P8]

"Client requirements. I think it's really important to have a very clear understanding of the brief." [P11].

"We all make assumptions; we form our viewpoint depending on precedent and what we've seen before. Understanding the opportunity that data can bring us up, because sometimes it surprises us." [P13].

Majority of the participants highlighted the fact that in initial stages of design they follow a two-stage approach to collect the required information. Designers rely mainly on the client to provide information, which is then followed by a desktop study from the design team, a site visit or by direct commission of external consultants to obtain missing information. Therefore, design teams consider that "*the only source of information is from the client and standard policies*" [P1] or "*Personal education*" [P15]. According to the authors' understanding, this implies that in the initial stages of design, publicly accessible information is not used, indicating designers' over-reliance to the client's brief and policies, which can be linked to several associated factors, e.g., competence and skills, experience, project budget, areas of interest etc. This is further supported by one of the participants:

"*Publicly accessible information at the RIBA stage Zero, but I think design teams tend to maybe kind of rush that and will do things like just use Google Maps. Will take an image of like a street pattern and then just infer on top of that. And I think that that is a challenge because you're not using accurate information necessarily*." [P8].

The radical alteration of the designer's practice has been identified as a continuous challenge since the early 90s. Schön (1991) claimed that professionals are expected to solve tasks they are not educated to handle, and are required to generate technological change to be aligned with the expectations and demands that technology itself has generated. Contemporary practices still struggle to accommodate change, as the complexity and time-consuming processing involved with "new ways of working" force designers to rely on existing approaches and workarounds to accelerate their work.

Furthermore, participants discussed wider the key challenges and opportunities of the approaches they use, summarised in Table 7-2. To highlight important aspects of the discussion in existing design approaches, the key findings are separated in four categories: Data, People, Tools and Brief. Table 7-2 also summarises the priorities for future efforts to be focused on, to improve existing design processes.

Table 7-2: Summary of key findings including challenges and opportunities in existing design approaches of buildings and places

Challenges	Opportunities	Findings & priorities for future efforts to improve design processes
Data		
Data management Data analysis Data sourcing Data accuracy Lack of information	Open-source data Data availability Feedback collection	• A consistent approach on how to collect the data first. • Validate the data from a quality perspective, not just from what it has been produced against the contractual, but against a set of requirements in terms of what the data means for the success of the project. • Encouraging conversation at the beginning of the process to leverage data that exist or the client to get information from other parts of the local authorities.
People		
Designer's (users) culture Client familiarity Understanding of client's business model Data Literacy	Centralised clients	• Choosing the right methodology for the clients, not necessarily changing the design process. • Putting in a broader context the nature of the problem, so it becomes clearer and with the breadth of the solutions. • Depends on the geographical context - introduce the benefits of sharing the data. • Having people with the experience of bringing that data or information into those different stages. • Bring people in with enough experience around construction and design, and data literate. • Robust leadership and a robust set of stakeholders that can supervise and sit across. • Validation at every single stage against all the stakeholders. • Design methodology captured in a way that could be audited.
Tools		
Multiple forms of communication Limited software experience Resources and skills	Open-source tools Design validation	• Tools with successful interface interaction with the client. • Low cost and effective tools and information, freely accessible geospatial information, and platform to use. • Reduction of manual processing. • Reduction of data sizes or training on how to utilise heavy files of information.
Brief		
Restricted budgets Time Complex stakeholder and procurement route	Future potential	• In the long run, a data-driven approach would be more efficient in the design process - therefore encouraging implementation in early stages. • Utilizing what is produced out of expensive requirements into contracts and scope in a meaningful way. • Digitize the construction process and the design. • Aligning stakeholders' expectations. • Sharing of information to avoid abortive work. • Asking the intelligent questions early on.

Results indicated the lack of a monitoring process of existing design approaches. More specifically, participants highlighted the need for the design process to be defined and communicated, including any potential methodologies. This can enable the designers to choose amongst a range of methodologies and identify the most suitable one for their projects. This can further help in articulating the benefits of the diverse methodologies, which can then be communicated to the clients, encouraging conversation at the beginning of the process.

7.4.2 Perceived drawbacks and priorities in DDI and BDAs implementation

The purpose of the semi structured interviews in Part C was to identify participant experience in DDI and BDAs implementation, followed by a discussion of the perceived drawbacks and priorities of such approaches. Several participants felt that "*it does not mean anything specific*" [P2] or that "*Big data is a fuzzy name – very abstract.*" [P3]. Participants with a wider knowledge around the use of data, classified as "Experts" in Table 7-1, also received this concept with various interpretations, as noted below:

"*These very large repositories, having more data all in one place and sort of timestamped."* [P5].

"*A really large number of data points, over 10,000, that provides you with a much more granular understanding of something. It could grow into a really big dataset, so it gives you much more kind of detailed information."* [P8].

"*From my perspective, means anything that doesn't sit specifically to a direct input, but it's more related to a pattern and it's modern drive out of a series of other datasets that are trying to connect. But it's the connectivity of data for other purposes than what it was intended for."* [P10].

Other interpretations included "*A system that stores raw data. which cannot necessarily be handled manually."* [P12], or "*Access to the large and open variety of information that can be extracted from a multiple from a variety of sources. Although they are not coordinated.*" [P4].

As identified from the literature, contemporary practices assume of having an in-depth appreciation of people's point of view on "Big data", as it forms a big part of the industry's agenda. However, the lack of a systematic understanding of how people perceive data results in the missed opportunity for stakeholders to express their opinion on data, how can these be embedded in their processes and what their impressions and challenges are.

One of the key findings includes that even though participants had a general understanding of what "*Big Data*" is, all of them felt that DDIs and BDAs are "*definitely relevant*" [P2] to their work. The reasoning behind that positive response was explained by several participants, as noted below:

"*It's important to have an overall view of the bigger picture of things before you start designing, so it's just another facet that feeds into the design process."* [P7].
"*Yes, because I know there could be a lot of more efficient ways that big data and digital emphasis could sort of improve efficiency on a project. So, I do think it's got a part of it. I just think a lot of people struggle to understand what big data is and what digital is on the railway."* [P6].

Due to the challenges identified in existing processes around people and skillsets, as summarised in Table 7-2, participants highlighted the fact that in many cases there are specific people with specialised roles responsible for the incorporation of diverse data sources and their findings within the design, as a separate team within the organisation. Nevertheless, based on participants' commentary, this barrier is now recognised by the organisation and moving towards a *"more integrated approach. The team is separated from the people who analyse the data – and keep feeding this information. But because they are also trained architects in most cases, it would be way easier to do it as part of the team."* [P1]. This has been recognised as a key challenge within the organisation, stating that "*Needs to be part of the design team. The design team might not be experts in that, but they need to have sufficient understanding to identify the need. It is difficult to keep up with the emergent technologies and they need to educate themselves.*" [P3].

Participants felt positive towards the implementation and adoption of new "*data frameworks*", while they recognised the value of such evidence-based strategies as part of the wider planning and design process, rather than just the project design. For example, "*A series of frameworks might help identifying different clients.*" [P1] or "*It doesn't necessarily need to relate to design information, but it really could relate to things like program or fees or resources or skills before design.*" [P9].

Participants were presented with the framework in Figure 7-3, enabling further discussion on the implementation challenges of similar evidence-based strategies in the design process. The proposed framework describes the steps to be undertaken on the top, with detailed information for each step described below. It generally recognises the importance of three distinct layers of information which are pre-processed and analysed utilising DDI and BDAs, leading to the extraction of insights. These can then be fed to the design team, enabling the definition of the design indicators and success factors at the end of this framework, which are then implemented directly in the traditional design process.

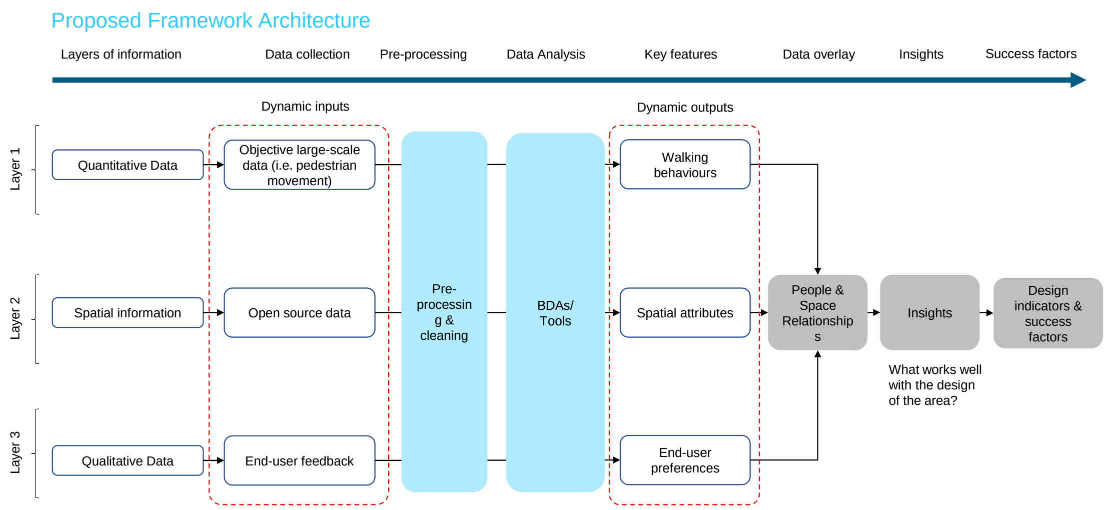


Figure 7-3: Proposed framework architecture example, as presented to the semi-structured interviews.

Participants expressed a positive feeling overall in the presented example or to similar evidence-based strategies potential as part of their design process. Participants did not alter their initial views around frameworks, following their exposure on Figure 7-3. On the contrary, the presentation of a framework example enabled some of the participants with no familiarity of DDIs and BDAs to materialise what the role of such evidence-based strategies could be and express their views towards potential drawbacks or problems in their future implementation. A summary of the findings and overall discussion is presented in Table 7-3. The same structure in the categories of Data, People, Tools and Brief is followed.

Table 7-3: Summary of drawbacks or problems participants foresee with evidence-based strategies implementation.

Drawbacks or problems	Findings & priorities for future efforts for evidence-based strategies implementation
Data	
Key Performance Indicators (KPIs) should be in place	• Data that can be turned into modelling.
Data accessibility	• Use the findings to review and analyse. • Consider it as an additional source of useful information for designers. • Centralise the repository of the information that is produced from a design stage through to construction and beyond.
People	
Complexity of implementation	• Needs to be used intelligently, not just applied to a project.
Fit in clients' aspirations	• It can serve as evidence when pitching to the clients.
Specialised roles required	• Useful as an overall view of the bigger picture.
Evidence the benefits	• Has to be dealt by someone who knows how to extrapolate but has to be also someone who understands the project needs.
Culture of users: nervousness around the use	• Finding a common language that can be shared.
Tools	
Bias in the analysis	• Improve efficiency. • Bring predictability in what designers do.

Brief	
Be organic, not solid, and linear	• Applied in the early stages of design- crucial decisions early in the process.
Increased budget	• Informs the general agenda.
Time constraints	• Potential in minimizing error and changes.
Communicate relevance to projects	• Everybody working in this environment has to have the ability or picking up that data in the shortest time possible.
	• Use lessons learnt to speed up the process.
	• Produce findings against a series of criteria

Participants highlighted the need for specialised skills and roles for the effective implementation of such approaches, while they have also referred to the need of clearly communicating any potential benefits of such approaches. DDI and BDAs have been perceived as an opportunity to inform the general agenda of their work and to access insights which have been previously hidden or required extensive manual work. In general, participants expressed their positive views towards new approaches, however, they have indicated that the application of such frameworks can reveal further opportunities and potential benefits.

The results further disclose the underlying concern for a lacking part of the stakeholder basic knowledge of general principles and methods around DDI and BDAs. More specifically, findings imply that the human component can be underestimated as a part of the design process. Design processes can become of increased complexity and by having multiple owners and leads creates challenges. The lack of a robust leadership in the initial stages of design results in a missed opportunity to introduce a set of requirements to define the role of “data” from a quality perspective. Furthermore, the absence of a clear definition of how information can be utilised in a meaningful way, embedded in an already complex procurement route, may lead to the introduction of additional costs as part of contracts and design scope. Therefore, the human aspect becomes a crucial factor when introducing novel approaches in design.

7.4.3 Strengths, opportunities, and future development of evidence-based strategies

Although evidence-based strategies' research potential in addressing design and planning issues is promising, technical and knowledge discovery challenges slow down their practical implementation, as indicated in Table 7-2 and Table 7-3. Nevertheless, all participants stated willing to test and adopt the presented framework or other similar ones. Participants noted that "*it is important to test things before you judge*" [P1] to "*get a clear understanding*" [P4]. However, few of the participants were hesitant to confirm that they would adopt such frameworks as part of their traditional approaches before comprehending what its role might be, as noted below:

"*Depends on how it would fit in. If you can demonstrate where the time saving is or where the efficiencies are, I have no reasons to say no."* [P2]

"*So, adopting it, it depends. If you are delivering a project and you're trying to test that on a project, that might be a challenge. A parallel thing where we test project that we have delivered, and we know already what's the performance in doing a traditional in going through traditional approach versus unimproved approach. That is something that is more realistic.*" [P4].

Other *participants* felt that the role of evidence-based strategies application lies with the ones defining the overall design process, such as the clients themselves or the client facing roles within the organisation, as noted below:

"*I can personally test it, but I don't think that I can make a call for how big the project is in terms of tests. For example, testing a new framework when it's the process of how we go about things is quite set in stone."* [P14]

"*It depends on the dynamics. So, depends on how your client is going to drive the collection of your data. That's where you're going to make the decision of which framework and which approach, you're going to use.*" [P1]

Several participants indicated that the overall presented framework "*is quite robust.*" [P14] while they highlighted the parts that they see as its strengths and opportunities. However, participants commented on additional features they would like to see added in future evidence-based strategies. The information is summarised in Table 7-4.

Table 7-4 Summary of strengths & opportunities and additional detail to be considered added in future evidence-based strategies as expressed by the participants

Strengths and opportunities
• *"The distinction of quantitative data, spatial information, and qualitative data"* [P13] • *"The data overlay and the insights' part, to understand that is an evolving process, rather than a single solution"* [P5] • *"Thinking about people, space, and relationship as a whole"* [P8] • *"The headings work really well as a general overview"* [P13] • *"Areas of innovation, especially ones that can help us communicate better"* [P12]
Future development/ Missing detail
• *"Site visits and collection of observational data"* [P1] • *"Allow for input and feedback from other disciplines and factors that would be part of the scheme"* [P6] • *"A good definition of what's a success factor"* [P3] • *"Evidencing if finding has a positive or a negative impact"* [P6] • *"The story should be at the beginning"* [P3] • *"Between layers of information and data collection, there needs to be an input of existing KPI's"* [P10] • *"Elements of predictability"* [P10] • *"Clear step, in which validation of information happens"* [P5] • *"Data sources step for documentation"* [P5] • *"Links to the building regulations and contractual requirements"* [P15]

Participants highlighted the importance of having such frameworks in place, however, there needs to be a strategy also in place, if this is to be presented back to the organisation. More specifically, the main concern voiced by one interviewee regarded a cultural issue often observed when introducing new workflows or methodologies, aimed to replace or add to existing ones:

"*I think people are just fixated on using that, but it will be hard to shift away from their habits that people use just. I think people find it hard to visualize something else that kind of does the same thing. It might be hard to either explain people why this sort of thing might be relevant in addition to the plan of work or something that might replace it*." [P15].

In a similar context, this was mentioned by another participant, however, highlighting an additional challenge; the level of experience in utilising such frameworks within the design process, as indicated below:

"*I don't think there's anything missing from the framework as such, but I don't know whether if you were going to do this in practice, whether you would almost need like a story at the bottom, where beforehand to be like this is the challenge. Why bother looking at those over other things? If people don't have previous experience of working with data, they sometimes find it abstract. How do you introduce people and design teams to this? But I don't see anything wrong with this."* [P8].

Several future development concepts were suggested by the participants; however, results indicate that the most critical one remains the stakeholder engagement. The development of evidence-based strategies is closely dependant to the end-users, and without their proper involvement in this process, frameworks can endlessly be altered to meet project needs. Instead, the focus should be turned on creating practical application examples to be tested, as the requirement of achieving specific outputs can highlight missing key features and enable users with no previous experience on such strategies to provide constructive feedback.

7.4.4 Potential areas of application and prioritised areas of concern

The results of the interviews indicate that designers feel that frameworks have a potential role in their existing design process. Throughout the interviews several parts of the design process that could provide potential areas of application for the frameworks were identified. However, several areas of concern associated with their practical implementation were also raised. This section discusses the key areas of concern as prioritised by the interviewers and possible ways of overcoming them.

In terms of design and planning, the optioneering development process present difficulties associated with creating evidence against each option tested and it is currently limiting their application at the strategic level (Figure 7-2- Strategic Level). Were the limitations as identified in Table 7-2 to be overcome, then frameworks could be applicable at the strategic level, re-enforcing what design process is already doing. Some interesting responses to arise from the interviews are noted below:

"*I think that this is a framework that sits in a stage rather than across all the stages. I guess you could use it in different ways at those different stages. Test your options against those, so you might come back to this at those later stages. In the rail sector in particular, they already do that kind of approach."* [P8].

"*Yes, I think it could probably go feed into all stages actually. I think this would be brilliant. in the early stages of a project in feeding into the design process for the rail sector."* [P6]

"*I would find these more useful during the early stages of the of the design process maybe. When you define what the concept is and what the dynamics of the project is."* [P9]

Furthermore, the results of the interviews articulate that designers feel that frameworks have a potential role in participatory design, hence acting as communication tools when design processes vary as much as identified in the organisation. As [P12] noted, they may offer a means to open dialogue between different disciplines by helping to explain and communicate in a "*language that* *everyone uses". In that way it can help to overcome issues arising from not using "a common template… Same approach or same tools. Then the information and the data can easily be connected afterwards.*" [P12]. Similarly, such frameworks have also a role to play in communicating project goals in the strategic level. Design process is very complex, and in many cases, designers feel overwhelmed from the breadth of information defined in a client brief. [P13] highlighted frameworks' importance as communication tools defining what is needed, as noted below:

"*To look at a site when you can't see the context, if you look at the brief, if you just had the site plan without the context, then you're not going to understand how you actually respond to that site. To be able to make it easier for people is what you essentially need."* [P13].

Another participant felt that frameworks have a role to play in the management of the design process, rather than the process itself. More specifically, [P10] notes that "*It doesn't necessarily need to relate to design information, but it really could relate to things like program or fees or resources or skills before design.".* Their potential is as visualisation tools to understand "*progress in your project*" [P15]. The participant further elaborated on the lack of interactive tools to ensure the efficient management of the project itself, noting that "*Buildings are so complex, you can't remember everything, and you need a database on every project to make sure you ticked off every box. And that currently doesn't really exist. It exists as a document, like a piece of literature."*

Participants were asked to close the interview by prioritising the most important point from the overall discussion (Table 7-5). It was interesting to find that the general concerns raised by the interviewees were similar throughout the semi-structured interviews, however, in terms of prioritising them, participants raised different points as of high priority. Results indicated that the variety of design methodologies undertaken, the challenges present in each role and sector, as well as the diversity in experiences and knowledge in big organisations, heavily affect the decision-making in design processes. This highlights even further the importance of framework implementation to streamline and enhance innovation in the design process of buildings and places. Further to this, 7 out of 12 interviewees' concern was related to the "People" category, emphasising the key role that the end-user holds in the effective implementation of evidence-based approaches in the design of buildings and places.

Table 7-5 Conclusion points and high importance priorities as perceived by the interviewees (key elements have been highlighted).

ID	Overarching Theme	Category	**Conclusion points and priorities** (as extracted via direct quotation)
P1	People/ Skillset	People	*"The whole process is important – every part is important. As a broader skill that everyone needs to have in the level of design and operation of a project is* ***to be able to understand and analyze data****".*
P2	Data management	Data	***"Management of data and how it is shared****. What changes it would make in knowledge sharing".*
P3	Integrated process	Brief	*"This process needs to be into the design process in a way that recognizes that* ***it's not distinct*** *and separate off from it".*
P4	Seamless process	People & Tools	***"It needs to be seamless****, but the fact how it's seamless and easy is to access the framework and use it".*
P5	Data availability/ analysis efficiency	Data & Brief	***"Availability of the data****, finding methods or people that can* ***analyse the data quickly*** *and make it available to decision makers".*
P6	Data sources	Data	*"It's* ***where you get your information from*** *because that forms your argument. Or your evidence".*
P7	Software	People & Tools	***"Having the right software*** *to work with just on the very basic level. This is shaping what we can achieve as well, and if they're restrictive, that just restricts the whole process and leads to. I think it probably leads to innovation being stifled by being hammered with delivering in a certain way".*

P8	Success factors definition	Tools	*"The endpoints, the **design indicators**, and **Success Factors**. How do we do that and then test our designs against it. Indicate like what the benefit of this is. You know it's to have the like this evidence-based data-driven design".*
P9	Feedback loop	Tools	***"Feedback.** Present your data and then have feedback and rethink about the whole cycle and how it works".*
P10	Big data perception	People	*"Probably the most important one is one of the first questions, which is under **understanding other people's point of view on big data**. There is an assumption because data is such a big part of what we do, whether in the industry or globally".*
P11	Process as a communication tool	People	*"This very **well-structured process** and it makes it more **clear and easier for everybody in the team to understand**".*
P12	Framework's importance as part of the design process	People	***"The importance of frameworks.** Because if frameworks mean achieving a level of organization, it can be either in projects or analysis or anything, and this can automate processes, save time, and produce more money for a company for example".*
P13	Timeframes	Brief	*"Making our teams understand the **importance of this data** and allowing them **the time** and the space to access it, beginning of project. Particularly on projects which are very fast track and where we might only have a week or so to do this at the beginning".*
P14	Transformation of existing design process	People	***"The use of data**, the use of big data or for being like heavily part of the design process; because otherwise I don't see how anything can work".*
P15	Efficiency of design process	Brief	*"Frameworks may give you an answer to a question that normally takes about four weeks to decide".*

7.5 Discussion

The design process contains a series of activities which may vary from one type of a building or a discipline to another (Ledbury, 2018). Early in the process, designers take crucial decisions, which often, these are not supported by evidence (Brown & Corry, 2011). Findings revealed that halfway through, designers may realise that the assumptions or questions asked in the beginning of the process do not adequately reflect the requirements of the client brief, or the process becomes “top-heavy” by analysing all available data in the first stages of design, wasting time and effort.

The availability of information and the synthesis of research findings to inform decisions in the early stages of design are needed to support the evolving complexity of design and construction practice. This study revealed that “Big data” approaches are relevant to the decision-makers in design, promising to tackle the lack of information and bring predictability in their work. However, they are not yet effectively implemented in their existing approaches. Due to the technological changes, the collection of data becomes easier, however challenges around their analysis and interpretation arise (Reddy, et al., 2020). The analysis suggests that the design and construction sector has already started the transition from document-based to model-based data by embedding specialised roles in the design processes. Nevertheless, the change to digital ways of working has been slow, with users still conditioned to work with traditional sources of information, such as Portable Document Formats (PDFs) and drawings and limited to what it is provided by the client. The complexity and long processing times involved with evidence-based strategies force users to shift back to existing approaches and workarounds to expedite their work. Findings indicate that current practice does not have policies in place to monitor existing approaches, missing the opportunity to identify issues and save a significant amount of capital throughout the stages and minimise error and changes. Thus, the development of frameworks to report evidence in design becomes a prerequisite.

A common misconception is the fact that having access to a plethora of data means that an in-depth understanding and visibility of the problems in a project can be easily acquired, and subsequently, decision-makers can resolve any issues before they arise. Potentially the problem is “Big data” itself, and the fact that data in isolation does not replace the human element of behaviour, nor communicates elements that are culturally ingrained around how people behave. Monitoring and capturing the relationship of people’s engagement with the evidence-based approaches and understanding their point of view on data is the key priority, as people hold the prospect for the effective implementation of such strategies in practice.

Dossick and Neff (2010) in their work indicated the importance of cultural differences among the diverse stakeholders, stifling collaborative work, caused by the existence of multiple design processes, domain-specific tools, and modelling practices. However, the findings of this research indicate that, although the design process can vary significantly, it responds to an overarching structure which defines the diverse steps to be undertaken. Results also highlight the individuality of each project, and the way client requirements can be analysed under several different points of view. Accordingly, following the participants’ views, frameworks have the capacity to define the design process steps, while introducing efficiencies throughout the process. Nonetheless, for the efficient framework implementation in policy and practice there is a need for two distinct approaches: initially, the introduction of a general framework responding to the overarching structure of the design process, and secondary, via multiple adopted frameworks in accordance with the needs of the diverse stages, activities, or steps, responding to the uniqueness of each project. To achieve this, it requires the practical testing of frameworks within the design approaches, via which the most appropriate forums for their use can surface. A set of common practices and a larger vision for the identified challenges around data creation, management, use and interpretation, should be laid out in the beginning of each project, ensuring that every decision-making step is recorded and conducted without loss of information in between.

Gaining understanding of human behaviour is a highly complex challenge, requiring the inclusion of the wider group of decision-makers involved in the design of buildings and places (Stanitsas & Kirytopoulos, 2021). By adopting such approach, a broader cross-section of stakeholder groups and individuals can be consulted, incorporating their greater knowledge and experience as decision-makers in all stages of design. The development of engagement strategies for the practical implementation of evidence-based approaches is an essential component unlocking their potential to be effectively utilised by designers.

The opportunity to test and monitor the framework implementation lies in the strategic definition level, where the benefit of adopting new approaches can be identified and communicated to the diverse stakeholders. Practitioners should focus beyond the individual scope of their organisations towards the common goals of the project. Initially, there needs to be a focus on data collection, and communicate the importance of the data and continuous data capture as part of this process.

The study's distinctiveness resides in its contribution to new knowledge by presenting the perceptions of building and urban design experts on the challenges involved with the application of frameworks. Furthermore, designers gain insights into the effective implementation of evidence-based strategies as a result of the findings.

7.6 Conclusion

This paper aims to raise awareness of the key challenges, opportunities, and priorities for evidence-based strategies' application to inform building and urban design decisions. The study employs deductive qualitative content and manifest analysis, utilising semi-structured interviews undertaken with building and urban design professionals who represent a UK-based organisation. Participant responses indicate that whilst evidence-based strategies potentially have a wide range of uses in the building and urban design industry, at the present time there remains a need to further develop engagement strategies for their practical implementation, clearly communicating the benefits and efficiencies to all stakeholders.

Together with the challenges associated with the practical implementation of frameworks, several potential application areas, and perceived constraints of high priority for future development have been identified, which need to be resolved for the evidence-based strategies to be effectively utilised by designers. These not only include the need to practically test their implementation, but also to identify the most appropriate forums for their use. The collection of data and use of modern analytics and methods in early stages of design offers opportunities to improve existing design approaches, helping building and urban design organisations make more informed and efficient decisions, considering future demands in ever-changing environments. The recipients of the findings will be the urban planners, designers and academics who are interested in delivering urban environments aligned to the end-user needs, employing evidence-based strategies in their decision-making processes.

Implications/ limitations of this study come with the fact that some of the respondents may possess inadequate professional experience in properly evaluating all the questions. Additionally, although participants were selected with different levels of experience in DDI or BDAs, the initial questions were exploring their understanding, identifying areas of concern regarding their responses, or further explanation of concepts where necessary. Finally, the information gathered is restricted to the UK geographical context, as well as coming from one organisation, due to data accessibility. Therefore, they are not generalisable and future research should explore gathering information from a wider audience in a global geographical context which could fine tune the results for specific contexts. Another line of inquiry for future research could be the analysis of the applicability of these frameworks before and during the implementation of a project. Further research should be conducted on the identified approaches for different types of projects to validate their usage as generic or project and sector specific.

REFERENCES

Abruzzini, A. & Abrishami, S., 2022. Integration of BIM and advanced digital technologies to the end of life decision-making process: a paradigm of future opportunities. *Journal of Engineering, Design and Technology,* 20(2), pp. 388-413.

Adams, W., 2015. Conducting semi-structured interviews.. In: *Handbook of Practical Program Evaluation(Newcomer, K. E., Hatry, H. P. &Wholey, J. S., eds.).* Hoboken, NJ, USA: John Wiley & Sons, Inc, p. 492–505.

Ahern, J., 1999. Landscape Ecological Analysis: Issues and Applications. In: *Spatial concepts, planning strategies and future scenarios: a framework method for integrating landscape ecology and landscape planning. J.M. Klopatek, R.H. Gardner (Eds.).* New York: Springer-Verlag, pp. 175-201.

Barton, N., Hallett, S. & Jude, S., 2021. The challenges of predicting pipe failures in clean water networks: a view fromcurrent practice. *Water Supply,* 22(1), p. 527–541.

Bengtsson, M., 2016. How to plan and perform a qualitative study using content analysis. *NursingPlus Open,* Volume 2, p. 8–14.

Brodaty, H., Mothakunnel, A., de Vel-Palumbo, M., Reppermund, S., Kocha, N.A., Savage, G., Trollor, J.N., Crawford, J. & Sachdev, P.S., 2014. Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging. *Annals of epidemiology,* Volume 24, pp. 63-71.

Brown, R. & Corry, R., 2011. Evidence-based landscape architecture: The maturing of a profession. *Landscape and Urban Planning,* 100(4), pp. 327-329.

Chang, X., Wu, J., He, Z., Li, D., Sun, H. & Wang, W., 2020. Understanding user's travel behavior and city region functions from station-free shared bike usage data. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 72, pp. 81-95.

Chen, M., Mao, S. & Liu, Y., 2014. Big data: A survey.. *Mobile Networks and Applications,* 19(2), pp. 171-209.

Dossick, C. & Neff, G., 2010. Organizational Divisions in BIM-Enabled Commercial Construction. *Journal of Construction Engineering and Management,* 136(4), pp. 459-467.

Esteves, J. & Curto, J., 2013. A risk and benefits behavioral model to assess intentions to adopt big data. *Journal of Intelligence Studies in Business,* 3(3), p. 37–46.

Forsyth, A., 2007. Innovation in Urban Design: Does Research Help?. *Journal of Urban Design,* 12(3), pp. 461 - 473.

Frank, L., 2000. Land Use and Transportation Interaction: Implications on Public Health and Quality of Life. *Journal of Planning Education and Research,* 20(1), pp. 6-22.

Fredriksson, C., 2017. Big data creating new knowledge as support in decision-making: practical examples of big data use and consequences of using big data as decision support. *Journal of Decision Systems,* 27(1), pp. 1-18.

Fusch, P. & Ness, L., 2015. Are we there yet? data saturation in qualitative research. *Qualitative Report,* 20(9), p. 1408–1416.

Gahm, C., 2020. A conceptual framework for cloud-based advanced planning systems. *Journal of Decision Systems,* 0(0), pp. 1-30.

Galvin, R., 2015. How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge?. *Journal of Building Engineering,* Volume 1, pp. 2-12.

Gopaldas, A., 2016. A front-to-back guide to writing a qualitative research article. *Qualitative Market Research,* 19(1), pp. 115-121.

Harris, R., 2012. *Introduction to decision making.* [Online]
Available at: https://www.virtualsalt.com/crebook5.htm

Imants, P., Theeuwes, J., Bronkhorst, A. & Martens, M., 2021. Effect of multiple traffic information sources on route choice: A driving simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour,* Volume 81, pp. 1-13.

Jetzek, T., Avital, M. & Bjorn-Andersen, N., 2014. Data-Driven Innovation through Open Government Data. *Journal of Theoretical and Applied Electronic Commerce Research,* 9(2), pp. 100-120.

Koberg, D. & Bagnall, J., 1981. *The universal traveler: A soft-systems approach to creativity, problem-solving and the process of reaching goals..* Los Altos: Kaufmann, W.

Kusiak, A., 2009. Innovation: A data-driven approach.. *International Journal of Production Economics,* 122(1), p. 440–448.

Ledbury, J., 2018. 8 - Design and product development in high-performance apparel. In: *High-Performance Apparel (eds. McLoughlin, J., Sabir, T.).* Sawston, Cambridge: Woodhead Publishing, pp. 175-189.

Malmqvist, J. et al., 2019. Conducting the Pilot Study: A Neglected Part of the Research Process? Methodological Findings Supporting the Importance of Piloting in Qualitative Research Studies. *International Journal of Qualitative Methods,* Volume 18.

Meinel, C. & Leifer, L., 2011. Design Thinking Research. In: *Design Thinking, Understand – Improve – Apply, Plattner, H., Meinel, C., Leifer, L. (Eds.).* Berlin, Germany: Springer, pp. 13-20.

Neuendorf, K., 2016. *The content analysis guidebook.* 2nd. ed. s.l.:Sage.

Oltmann, S., 2016. Qualitative interviews: a methodological discussion of the interviewer and respondent contexts. *Forum: Qualitative Social Research,* 17(2), pp. 1-16.

Pollard, J., Spencer, T. & Jude, S., 2018. Big Data Approaches for coastal flood risk assessment and emergency response. *WIREs Climate Change,* 9(5).

Power, D., 2014. Using ‘Big Data’ for analytics and decision support. *Journal of Decision Systems,* 23(2), pp. 222-228.

Power, D., Cyphert, D. & Roth, R., 2019. Analytics, bias, and evidence: the quest for rational decision making. *Journal of Decision Systems,* 28(2), pp. 120-137.

Reddy, G., Reddy, M.P.K., Lakshmanna, K., Kaluri, R., Rajput, D.S., Srivastava, G. & Baker, T., 2020. Analysis of Dimensionality Reduction Techniques on Big Data. *IEEE Access,* Volume 8, pp. 54776-54788.

RIBA, 2020. *RIBA Plan of work,* London: RIBA.

Ritchie, J., Lewis, J. & Elam, G., 2003. Designing and selecting samples.. In: *Qualitative Research Practice(Ritchie, J. & Lewis, J., eds).* Trowbridge, Wiltshire: SAGE Publications Inc, p. 77–108.

Rosenberg, W. & Donald, A., 1995. Evidence based medicine—an approach to clinical problem solving. *BMJ,* 310(6987), pp. 1122-1126.

Schön, D., 1991. *The Reflective Practitioner.* London, Great Britain: Ashgate Publishing Ltd.

Semeraro, C., Lezoche, M., Panetto, H. & Dassisti, M., 2021. Digital twin paradigm: A systematic literature review. *Computers in Industry,* Volume 130, Article 103469.

Stanitsas, M. & Kirytopoulos, K., 2021. Investigating the significance of sustainability indicators for promoting sustainable construction project management. *International Journal of Construction Management,* 0(0), pp. 1-26.

Stanitsas, M. & Kirytopoulos, K., 2022. Underlying factors for successful project management to construct sustainable built assets. *Built Environment Project and Asset Management,* 12(2), pp. 129-146.

Ward, J. & Barker, A., 2013. Undefined By Data: A Survey of Big Data Definitions. *ArXiv.*

West, R., Toplak, M. & Stanovich, K., 2008. Heuristics and biases as measures of critical thinking: Associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology,* 100(4), pp. 930-941.

8 DISCUSSION AND CONCLUSION

8.1 Overview

A summative response to the original research aim, the limitations observed and the novel contribution for each research objective is provided. Summative conclusions and areas of further work are also offered.

8.2 Reflections

8.2.1 Research Question and Response

The stated research aim sought to *“assess the adoption of, and opportunities deriving from, data-driven innovation techniques in the design of urban spaces, by the analysis of pedestrian movement patterns in urban environments, and to evaluate how the integration of evidence-based strategies can be established in supporting decision-making in relation to future urban designs”*.

The research question investigates two different stakeholder groups: *End-users* and *Decision-makers* in designing buildings and places. With respect to the first group, it is already acknowledged that decision-making processes in urban spaces are complex, and that the individuals concerned are heavily influenced by various parameters, such as their previous experiences or the decision-environment (Hilbert, 2012; Ghattas, et al., 2014; Gal & Pfeffer, 2008). However, there remains a gap concerning the identification of the most contributory influencing factors among decision-maker groups and how these factors influence the design process. Technological advancements and the growing utilisation of data-driven innovation techniques within urban research and design applications further add to this complexity (Yi, et al., 2014; Castelli, et al., 2020), introducing additional challenges. Evaluating the importance of 32 individual variables, four key factors were generated for all the types of stakeholders investigated, namely: *Potential for Dynamic Operation, Recency of tools, Thoroughness* and *Control*. Providing a way to extract insights and identifying these influencing factors within decision-making processes represents a step forward towards a more evidence-based decision-making approach and this can lead to more robust decision-making. The factor *Potential for Dynamic Operation* has been highlighted as one of the most important aspects within decision-making process, indicating the need for a dynamic and continuous decision environment, utilising novel tools and methods within this process.

Such techniques have the potential of extracting novel insights that can be applied directly within urban design approaches, complementing traditional methods (e.g., (Whyte, 1980; Gehl, 2011)). Urban spaces vary significantly, and these strongly influence pedestrian movement and behaviour in return (Choi, 2014; Bozovic, et al., 2020). Provision of a consistent methodological framework for identifying pedestrian categories across a range of different periods or contexts and investigating these in conjunction with spatial attributes can reveal insights in relation to individual preferences of end-users and the way they utilise the space. Application of the framework methodology as implemented in a high pedestrian traffic-dense retail urban area in London revealed clear and consistent relationships amongst spatial visibility, individuals' motivation, and knowledge of the area. In addition, the analysis identified the way spatial configuration either undermines or enhances pedestrian flow for each classified end-user group behaviour. The relationship between end-users and spatial configuration was reviewed against six key categories: walkability and accessibility; form and scale; diversity of activities; occupancy; safety, and security; and seasonality.

It is acknowledged that innovation may bring great advantages and opportunities (Denis, et al., 2002; Verhoef, et al., 2021). Nevertheless, innovation also carries significant risks of disadvantaging specific groups (Papaioannou, 2014). Such groups can include or comprise the decision-makers in the design of urban spaces. The research has revealed that although these groups hold domain knowledge, relied upon to support decision-making, they may lack the specialised skillset requirements for utilising of data-driven techniques to aid this process. To overcome this barrier, organisations have created specialised roles, with the specific responsibility of undertaking data-driven innovations and passing results back to design decision-makers. Such "design-team structures" can be problematic as they introduce further implications in the decision-making processes, such as the introduction of additional bias from roles having limited domain knowledge. In addition, although a positive view from the practitioners exists towards the implementation and adoption of data-driven innovations exist, additional limitations have been identified and these can be divided to four key categories: People, Data, Tools and Project Brief. These categories include barriers such as lack of data management and sources, decision-makers ability to analyse and understand such information, time constraints or even stakeholder business models. Nevertheless, although their full application is neither yet possible nor seamless, integrating these in recognised potential application areas, such as their integration in participatory design, can raise awareness among stakeholders, enabling practitioners to employ evidence-based strategies in their decision-making processes.

Another group potentially disadvantaged by the methods adopted include those persons whose presence was not recorded in the urban space under study. This group can include those suffering with long-term impairments (e.g., blind, or neuro-divergent people) (Deluka-Tibljaš, et al., 2022), those physically unable to access the study area, or those who may have been present, yet remained unrecorded as they do not use smartphones, limiting the ability for their movements to be tracked. The omission of these groups highlights the risk of perpetuating discriminations of access (e.g., that the methods consider only the needs of only select or privileged groups). Therefore, the presence of additional information regarding a person’s impairments, constraints, perceptions, or internal motivations could reveal further insights and complement the validity of the results. Nevertheless, this does raise a potentially major concern relating to the protection of privacy, and the need to clearly define the boundaries of any future study clearly. For example, who has the right to access the unique device MAC addresses used to track people’s movement or for how long are the collected datasets stored and how can the designers of the study limit the risks of editing a dataset to reidentify people to the extent where people’s residences and workplaces become obvious.

8.2.2 Limitations

Case study protocol was followed for the mixed-method data collection to enhance the reliability of the inter-related sub-studies as well as the replicability of the experiments. The case studies explored by this research have primarily focussed on pedestrian movement in a retail high-street in London, while it investigated decision-making contexts were investigated at two scales; the first one was addressed at a global scale to better understand decision-making context, and the second one was focused in one UK-based organisation.

Limitations identified with respect to the research stem from methodological choices, as well as constraints placed on the project through it being sponsored by industry. For the latter, the research may have benefited from adopting additional case studies that delved more deeply into end-users' movement. Nevertheless, data availability limitations and COVID-19 implications on observational data collection did not allow for case-study comparison. For the former, as an example of the methodological limitations, employment of the framework methodology developed may have benefited from the comparison of the results against additional datasets derived using similar collection techniques in similar contexts for longer periods. This limitation also includes the use of one case study for pedestrian movement exploration or the specific case studies used in a UK context mainly, although the author maintains that sufficient justification of these design decisions is provided in the earlier chapters of this thesis.

Other examples of limitations (which are addressed in the individual chapters) include the weaknesses recognised with regard to the generalisability of results from questionnaires as well as the recognition that Machine Learning algorithms would provide improved performance with more data. Conversely, increasing model complexity can result in deterioration in performance. As pedestrian movement is of increased complexity, this study assessed the most impactful parameters, such as spatial visibility, to avoid model overfitting and serve the key research hypotheses, while changes were also observed and incorporated from other factors, such as the weather implications. Nonetheless, despite such limitations, the findings remain valuable for extending knowledge in the fields of urban design and decision-making.

8.2.3 Contribution

In conclusion, the specific novel contributions of this research to the topic area are reiterated:

1. This study utilised systematic literature examination to draw from spatial cognition, decision making, and walkability research areas. The work demonstrates that those elements contributing to the sensorial experience of urban spaces should be considered when analysing spatial decision-making and pedestrian routing choices. Further, the work **advances the research methods that can be applied to the study of pedestrian movement in urban environments**.

2. **A quantitative data-driven, evidence-based methodological framework to evaluate the performance of decision-making processes has been developed.** The current analysis captures stakeholder perceptions as to the influencing factors affecting decision-making processes and a quantification as to the way designers make decisions. This research offers a real impact and change in practice by demonstrating the importance of adopting data-driven innovation techniques in decision-making processes in design. The study highlights the need for new metrics, frameworks, and skillsets to improve the ability of designers to extract insights aiming at a better understanding of users' needs.

3. **A novel methodological framework to assess pedestrian routing in urban environments via pedestrian behaviours classification and spatial configuration interactions** to support decision-making. This study utilises an evidence-based method and processes significant amounts of objective movement data to gain insights into pedestrian navigation. More specifically, investigation of the effect of spatial visibility

on walking patterns and weather conditions was conducted. The methodological path followed involves the development and application of Machine Learning approaches used with location data derived from Wi-Fi tracking techniques. This research directly contributes to the existing knowledge surrounding scientific approaches for pedestrian assessment by reinforcing the importance of data-driven environments in supporting improved decision-making in urban design. By understanding the impact of the different physical attributes and environmental factors on pedestrian movement, data-driven design approaches for urban spaces are enhanced, as these factors play an important role on routing choices.

4. This study **raises awareness of the key challenges and opportunities, priorities, and potential areas for evidence-based strategies in informing building and urban design decisions**. Limitations, potential application areas, and perceived constraints for implementing data-driven innovations have been identified, offering opportunities to improve existing design approaches. This contributes in the understanding the implementation potential of strategies that can be applied in decision-making, helping building and urban design organisations make more informed decisions.

8.3 Conclusions

Conclusions are summarised with regards to the intended objectives of this research.

Objective 1	Undertake a critical analysis of the needs of end-users and decision-makers within the design process for urban systems.

Urban and building design stakeholders can be divided into two key groups; the *End-users* - the people who use the designed spaces- and the *Decision-makers* who drive the design process. The way people move in a space influences its design and vice versa. A variety of methods has been explored to understand the needs of both stakeholder groups concerning urban design requirements, details of which are provided below.

Sub-objective 1.1	Assess and evaluate the needs and objectives of stakeholders through a critical review of the current state-of-the-art scientific and practitioner literature.

A review of the state-of-the-art in the pedestrian movement was undertaken to address the first part of this objective. Individuals are unique, but several common elements influence their movement behaviour. Firstly, this review highlighted the parameters influencing individuals' movement and the qualities that space should offer to the users, such as network and visibility connectivity. Secondly, the review also identified the need to classify behaviours systematically as such behaviours vary significantly based on the highlighted personal and spatial characteristics. Finally, the review highlighted the lack of employment of novel data collection techniques in the study of urban space recognition and the role of sensorial experience on movement patterns, limiting their potential in complex urban spaces.

Sub-objective 1.2	Identify the influencing factors of the design decision-making process via stakeholder engagement using a structured questionnaire approach.

The objective of a decision-maker in the design of urban spaces is to utilise available information to provide acceptable outcomes for a large proportion of the population and exhibit sufficient flexibility so that several types of end-users can use a space. Nevertheless, individuals in the design process are also unique, and their decision-making process is influenced by attributes such as previous experiences, roles, and disciplines. A stakeholder engagement through a structured questionnaire for professionals in the building design industry was conducted to meet the second part of this objective, which is to investigate the factors that influence decision-making processes in urban design. The methodology followed included Exploratory Factor Analysis, Average Relative

Importance Index, and Spearman's Rank Correlation Coefficient. The findings revealed that the most influential ones are *Potential for Dynamic Operation, Recency of tools, Thoroughness and Control*. This research also indicated that decision-makers need to rely more on new types of information to identify how a space should be designed rather than their instincts or current knowledge. The lack of appropriate skillsets to manage data collection and analysis drives motivation for undermining novel technologies and information. Stakeholders have already acknowledged the potential of such tools, although uptake is slow due to fundamental cultural and technical barriers. Barriers to implementation are numerous, but mainly they can arise due to difficulties or inabilities to access and easily manipulate data and tools that can improve the final outcomes of design.

Objective 2	*Evaluate the potential of data-driven approaches for revealing new insights in geo-temporal human behaviours and their application within the urban planning process in the building design sector.*

Pedestrian movement patterns influence how urban planning and design of contemporary cities are executed. Decision-makers require a better understanding of pedestrian movement within the urban design and planning industry to improve the design of urban spaces. The introduction of new technologies and data types present opportunities in extracting insights for end-users and built environment interactions. Various methods have been investigated to evaluate the potential of such tools and information, which are further described in the following sub-objectives.

Sub-objective 2.1	Explore the state-of-the-art data collection techniques in the study of pedestrian movement via a systematic literature review, identifying how novel informatics and data-driven technologies inform design decisions and reflecting how humans respond to the built and planned environment and perceive information.

A systematic literature review of the state-of-the-art of novel data collection techniques in the study of pedestrian movement was undertaken to meet the second objective of the research project. The employment of Big Data approaches in the context of pedestrian movement presents several challenges. Such approaches have been previously applied in contexts of safety and mobility. However, other qualities of space have not been extensively researched with such methods. Four key research gaps have been identified in the study of pedestrian movement, while three opportunities of applying new types of data collection techniques are proposed to bridge the research gaps identified. These are: (i) Employing large-scale pedestrian movement monitoring, (ii) Internet of Things (IoT) systems employment, and (iii) Leveraging cross-discipline incorporation.

Complementary to the review, thematic analysis and an online survey questionnaire are employed to identify and report the different types of Big Data challenges and investigate stakeholders' perceptions concerning significant challenges and the potential these approaches to be implemented in research and practice. Due to the high complexity and inter-related parameters that require an understanding of the qualitative attributes of a space, traditional data collection and analysis are usually employed. Novel types of information and model development require domain knowledge. Decision-makers have access to such knowledge; nevertheless, they may lack technical skillsets.

Sub-objective 2.2	Evaluate the scope and effectiveness of Wi-Fi tracking and Machine Learning techniques for extracting enhanced large-scale information, generating new findings and insights in the study of pedestrian movement via a case study approach.

Should such barriers be overcome, novel methods such as emergent Big Data Approaches can be used to reveal new insights into geo-temporal human behaviours and the urban planning process. This study employed diverse information and collection techniques to evaluate their scope and effectiveness. More specifically, a case study in a retail high-street in London was chosen, and Wi-Fi tracking techniques were employed. This research revealed that such techniques could be utilised for observations in large-scale contexts, where traditional approaches currently fail. By investigating walking patterns in conjunction with spatial attributes, such as spatial visibility, insights are revealed concerning individual preferences and behaviours of end-users and utilisation of the urban space.

Nevertheless, one implication is that GDPR considerations can lead to restrictions for the collection of such information due to ethical considerations. In addition, the explored techniques explored include several technical limitations, such as the requirement that individuals switch on their personal mobile device Wi-Fi to allow tracking, or that recorded accuracy is not adequate for detailed analysis on individual movements. Data collection approaches are constantly improving, and significant progress has been observed in the last ten years. However, current approaches still require cross-validation techniques and the employment of more than one data collection technique. Furthermore, this research highlighted the need for more than one types of information. At the same time, it enhanced the value of collecting qualitative information of people's opinions, experiences, and potential physical walking constraints to conclude in other aspects that influence pedestrian movement, such as aesthetics.

Sub-objective 2.3	Appraise data collection, analysis, and visualisation techniques, assessing how they facilitate decision-making and data gathering, and complement traditional urban design approaches.

Urban design guidelines and studies mainly depend on expert knowledge, existing literature, and experience rather than on data-driven innovations. This research has presented a novel means to classify pedestrian behaviours, employing different analysis and visualisation techniques to extract insights into complex relationships between urban design qualities and pedestrian movement, focusing on aspects that may hinder or support walking intentions. This study evidence how spatial visibility influences walking via the use of generalised additive models. It illustrated how the spatio-temporal analysis could help practitioners and researchers visualize results to better understand walking behaviours in diverse urban environments via the employment of heatmaps and partial dependence plots. This analysis has indicated that urban areas of similar type, e.g., retail high streets, have a wide range of end-user behaviour variations. These behaviours are supported via different space configurations and activities. Understanding the effects of the various physical attributes and environmental factors on pedestrian movement can lead to more accurate and robust models applied by decision-makers in urban design processes.

Objective 3	*Identify the role of evidence-based strategies in supporting robust decision-making in informing urban design approaches in practice.*

The final objective of this research was to identify the potential of frameworks to be implemented in urban design decision-making. Framework approaches enable decision impacts to be assessed, facilitating the evaluation of alternative strategies concerning these decisions. Nevertheless, such approaches come with challenges and opportunities. These are explored, and the resultant findings are explained in sub-objective 3.1.

Sub-objective 3.1	Identify the key challenges, opportunities, and priorities for evidence-based strategies in informing building and urban design decisions using semi-structured interviews with practitioners and decision-makers.

A series of semi-structured interviews were undertaken with building and urban design professionals representing a UK-based organisation. Participant responses indicated that evidence-based strategies potentially have a wide range of uses in the building and urban design industry. However, at present, there remains a need to further develop engagement strategies for their practical implementation, clearly communicating the benefits and efficiencies to all stakeholders. These also include considerable effort and time to be invested in training user communities to establish the most appropriate tools and methods to inform the decision-making processes of buildings and places. Together with the limitations, future development areas have been identified, divided into four key categories: People, Data, Tools, and Project Brief.

Should the identified limitations be overcome, frameworks could be introduced at the strategic level, re-enforcing existing design processes in the building and urban construction industry. The results have also indicated that the variety of design methodologies undertaken, the challenges present in each role and sector, and the diversity in experiences and knowledge in big organisations, heavily affect the decision-making in design processes. Thus, highlighting even further the importance of framework implementation to streamline and enhance innovation in the design process of buildings and places.

8.4 Future work

The case studies explored by this research have focussed explicitly on pedestrian movement in urban spaces and decision-making processes in the design of these spaces. However, the findings of this research can provide a robust means to assess behaviours in large-scale contexts and prepare appropriate adaptation strategies for urban spaces or additional assets, such as retail uses. Several other potential avenues for further research have also been identified.

Firstly, this research evaluated the impact of spatial configuration, and more specifically spatial visibility, on pedestrian movement in only a limited capacity. This limitation was due to the lack of accessibility to additional datasets to enable more extended periods to be studied or conduct a comparative case study assessment. In the future, these might become available, and it would be interesting to understand the impact of other spatial attributes, such as aesthetics or green spaces interaction, or additional parameters, such as the rain impact. In addition, due to the lack of observational data, the model choice was limited to unsupervised machine learning. Future analysis could explore supervised machine learning models with collected, labelled data. This exercise could be coupled with the collection of qualitative information to validate some of the assumptions. Finally, due to COVID-19, pedestrian behaviour in urban spaces has been altered; therefore, it would be interesting to compare before and after pedestrian movement patterns.

Secondly, this research collected data from decision-makers globally via an online questionnaire, while it was restricted to only a UK-based company to collect qualitative information via semi-structured interviews. As decision-making processes vary from one part of the world to the other, this research could be expanded in collecting qualitative data from diverse UK organisations to validate further the role of frameworks within the decision-making processes in the design of buildings and places. Additionally, it would be interesting to investigate decision-making processes in more detail by a comparative assessment among organisations located in different parts of the world.

Finally, although evidence-based strategies potentially have a wide range of uses in the building and urban design industry, at present, there remains a need to further develop engagement strategies for their practical implementation, clearly communicating the benefits and efficiencies to all stakeholders. Therefore, further research should explore visualisation methods or other practice-based mechanisms for collaboration and engagement among diverse stakeholders. There are also opportunities to examine further the impact of these engagement initiatives by collating evidence in decision-making process optimisations or failures with in-depth qualitative explorations of practice-based experiences and perceptions in a stepwise manner to document what works and why. Collating such evidence can enable the practical implementation of evidence-based strategies by translating theory into practice.

8.5 Key recommendations

A series of recommendations were also produced, following the findings of this research:

1. **Encourage and support collaboration between scientists and decision-makers**

This study explored how Big Data Approaches could be integrated into decision-making processes. This research identified that decision-makers in the design of buildings and places mostly rely on their personal beliefs or traditional data sources. Such information does not clearly articulate the breadth of information existing in planning and design research, while they do not introduce much innovation, impacting the practice. Further to this, lack of communication between stakeholders of design and scientists presents limitations in utilising these “theoretical foundations” in the practical implementation of the design.

Therefore, it is suggested that the two identified groups, scientists, and decision-makers, require collaboration to resolve fundamental issues of the practical implementation of evidence-based strategies, including educational and practical barriers. Then, they can address future problems that global challenges introduce in contemporary urban contexts. To achieve this, an in-depth understanding of each group's needs, capabilities, and limitations is required. This research identified the uniqueness of individuals; hence if their different values and beliefs as actors within the decision-making process of design of buildings and places are not accommodated in such collaborations, then these are likely to fail. Such collaborations would enhance knowledge sharing among disciplines and stakeholder groups, expressed in the form of partnerships or joint research projects.

2. **Change academic curricula for scientists and decision-makers, providing training and tools**

Building upon the previous recommendation, traditionally trained architects and designers rely on project-based learning, while new training methods have not been introduced in traditional education. Scientists have the specific skillset requirements to introduce innovation and new approaches, such as Big Data Approaches, within the design process. However, scientists may lack an understanding of how decision-makers actually make decisions. This lack of understanding can include the types of information needed or identification of the benefits or impacts of such process transformation. A new generation of scientists is required, enabled to perform a “hybrid” role, applying big data techniques in novel contexts and with domain-specific knowledge, to carefully define the research questions to be explored.

Further, new tools and frameworks need to be developed, bridging such gaps, providing a significant level of robustness and detail while becoming easily transferable across different disciplines, sectors, and stakeholder groups. These tools and frameworks offer great opportunities; however, to be successfully embedded into the design process, they need to convert the information provided by the scientists in a flexible format, easy to manipulate and adjust by the decision-makers, and vice versa.

3. **Support innovation**

Over the last decade, open data initiatives have focused on creating and maintaining open-source data portals, creating ecosystems of applications and services. Nevertheless, understanding human behaviour is a highly complex challenge requiring the inclusion and broader participation of all parts of society. Ethical considerations or lack of accessibility to informational innovations for the citizens can limit their capacity to participate fully and act as agents to their change. Supporting innovation is essential to address global challenges that contemporary cities face, and design places better attuned to end-user needs. This can be achieved by embracing the collection, use, and analysis of new types of information in complex set-ups. If innovation continues to occur in isolation, it will potentially lead to maladaptation or slow implementation, explicitly from individual actors. Citizen empowerment and policymaker leadership will be fundamental in mainstreaming such digital transformations in the design process.

REFERENCES

Bozovic, T., Hinckson, E. & Smith, M., 2020. Why do people walk? role of the built environment and state of development of a social model of walkability. *Travel Behaviour and Society,* Volume 20, pp. 181-191.

Castelli, N., de Carvalho, A.F.P., Vitt, N., Taugerbeck, S., Randall, D., Tolmie, P., Stevens, G. & Wulf, V., 2020. On technology-assisted energy saving: challenges of digital plumbing in industrial settings. *Human–Computer Interaction,* 0(0), pp. 1-29.

Choi, E., 2014. Walkability and the complexity of walking behavior. *A/Z ITU Journal of the Faculty of Architecture,* 11(2), pp. 87-99.

Deluka-Tibljaš, A., Šurdonja, S., Ištoka Otković, I. & Campisi, T., 2022. Child-Pedestrian Traffic Safety at Crosswalks—Literature Review.. *Sustainability,* 14(3), p. 1142.

Denis, J., Hebert, Y., Langley, A., Lozeau, D. & Trottier, L., 2002. Explaining diffusion patterns for complex health care innovations. *Health Care Management Review,* 27(3), pp. 60-73.

Gal, Y. & Pfeffer, A., 2008. Networks of Influence Diagrams: A Formalism for Representing Agents' Beliefs and Decision-Making Processes. *Journal of Artificial Intelligence Research,* Volume 33, pp. 109-147.

Gehl, J., 2011. *Life Between Buildings: Using Public Space.* Copenhagen, Denmark: The Danish Architectural Press.

Ghattas, J., Soffer, P. & Peleg, M., 2014. Improving business process decision making based on past experience. *Decision Support Systems,* Volume 59, pp. 93-107.

Hilbert, M., 2012. Toward a synthesis of cognitive biases: How noisy information processing can bias human decision making.. *Psychological Bulletin,* 138(2), p. 211–237.

Papaioannou, T., 2014. How inclusive can innovation and development be in the twenty-first century?. *Innovation and Development,* 4(2), pp. 187-202.

Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Qi Dong, J., Fabian, N. & Haenlein, M., 2021. Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research,* Volume 122, pp. 889-901.

Whyte, W. H., 1980. *The Social Life of Small Urban Spaces, Project for public spaces.* Washington, D.C.: Conservation Foundation.

Yi, X., Liu, F., Liu, J. & Jin, H., 2014. Building a network highway for big data: architecture and challenges. *IEEE Network,* 28(4), pp. 5-13.

APPENDICES

Appendix A Systematic literature review documents

Table A-1: Overview of the 32 final selected studies for the systematic literature review assessment

	Author	Date	Country (case study area)	Type of urban space researched	Approach/ focus	Methods used
1	Shakibamanesh, A., Ghorbanian., M.	2017	--	Urban spaces	Individual movement and perception of time	Cause-and-effect analysis (field experiment)- virtual reality & exploratory field research
2	Natapov, A., Fisher-Gewirtzman, D.	2016	--	Virtual urban spaces	Individual movement	Immersive Virtual reality
3	ÖZBİL et al.	2015	--	20 2kmx2km urban areas	Pedestrian movement	Observational data, GIS models
4	Wang et al.	2014	Hong Kong, China	Circulation region of a shopping mall	Attractors	Pedestrian simulation model—CityFlow-U
5	Choi, E.	2014	Stockholm, Sweden	Urban spaces	Qualities of built environment	Empirical study
6	Bratina Jurkovič, N.	2014	Ljubljana, Slovenia	Open spaces	Factors that trigger a sense of satisfaction	Empirical research, focus group and socio-spatial schema method
7	Kürkçüoğlu, E., Akin, O.	2013	Üsküdar, Turkey	Water elements	Water elements, urban design and spatial perception processes	Survey (questionnaire)
8	Furman, A.	2012	North America		Walkability	Review
9	Teixeira, C.F.B.	2021	Brazil, South America	Trees and green spaces	Vegetation presence and configuration on human behavior	Survey (questionnaire)

10	Resch et al.	2020	Salzburg, Austria and Cologne, Germany	Urban spaces	Emotions against the background of environmental information	Wearable physiological sensors combined with an eDiary app
11	Ba et al.	2020	Harbin, China	Open spaces	Crowd behaviour/ impact of smell and sound	Experimental analysis
12	Askarizad, R., Safari, H.	2020	Iran	Urban spaces	Social Interactions	Space Syntax technique and empirical observation
13	Liang et al.	2020	Harbin, China	Urban spaces	Severely cold areas	Computer vision technology/ video-based observational study
14	Kim et al.	2020	--	Urban spaces	Physiological response	Physiological saliency cue (PSC), wearable devices
15	Zapata, O., Honey-Rosés, J.	2020	Vancouver, Canada	Pedestrian Street	Social interactions as attractors	Field study
16	Boumezoued et al.	2020	Bejaia, Algeria	Streets	Multi-sensory experience - visual parameter is the main determinant	Qualitative and quantitative methods.
17	Botes, C.M., Zanni, A.m.	2020	Taipei, Taiwan	Urban spaces	Preferences	Discrete choice experiments (DCE)
18	Zhao et al.	2019	--	Urban spaces	Psychological behaviors: social repulsive force	Experimental analysis
18	Capitanio, M.	2019	Kunitachi, Tokyo	Urban spaces	Streetscape features relating to comfort and pleasurability influence pedestrian behavior	Space syntax/ counting of pedestrian frequency on site
20	Engelniederhammer et al.	2019	Hong Kong, China	Crowded spaces	Invasion of a personal space, emotional response	Emotional responses were measured psycho-physiologically via a wearable device, a movement detection sensor

21	Fisher-Gewirtzman, D.	2018	--	Urban spaces	Urban intensification, Predicted Simulation	Experimental analysis, visualization laboratory
22	Qian et al.	2018	Nanjing Hexi, China	Urban spaces	pedestrian friendliness: walking accessibility, pedestrian route directness (PRD), street walking popularity and the ratio of green interface	Survey on Residents' Travel Preferences and Travel Concern Factors, travel destination data through the POI retrieval, Kernel Density, Travel Time, Integral Pedestrian Route Directness, Correlation Analysis
23	Benachio et al.	2018	Hamburg, Germany	Neighbourhoods	Relationship between the concepts of a liveable neighbourhood and human perception	Interviews and surveys
24	Liao et al.	2017	--	Urban spaces	Differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers (self-localization, spatial knowledge acquisition, and decision-making)	Eye tracking, Empirical study, digital interface
25	Mansouri, M., Ujang, N.	2017	Kuala Lumpur, Malaysia	Urban spaces	Tourists' movement patterns	Space syntax, observational data
26	Filingeri et al.	2017	--	Crowd situations (transport hubs, sport events, retail situations)	Experience in crowds	Focus groups and observations
27	Nikolopoulou et al.	2016	Kent, UK, Portsmouth, UK, Glasgow, Scotland	Campus, railway surrounding	Shaping of pedestrian movement via playful interventions: Triangulation, performance and flow	Pilot studies, Questionnaires, and surveys

28	Karndacharuk et al.	2016	Auckland, New Zealand	Shared streets	Movement, access and place functions	Qualitative analysis: interview survey, statistical analysis
29	Ozer, O., Kubat, A.S.	2015	Galata, Istanbul	Urban spaces	Walkability: safety, accessibility (space syntax integration values) and land use pattern.	Multiple regression analysis
30	Phillips et al.	2013	Swansea, Wales	Urban spaces	Familiar and unfamiliar spaces	Experimental analysis, interviews, images display, content analysis
31	Ghahramanpouri et al.	2012	Johor Bahru, Malaysia	Pedestrianised streets	Stationary and sustained activities in pedestrian streets	Observational data
32	Orellana, D., Wachowicz, M.	2011	Amsterdam, Netherlands	Urban spaces	Movement Suspension	Exploratory statistical approach, GPS recordings

Table A-2: Records with full access assessed as part of the systematic literature review process

Author	Date	Country (case study area)	Type of urban space researched	Approach/ focus	Methods used
Liao et al.	2017	--	Crowded spaces with entry/ exit points	Moving pedestrian crowds	Experimental & simulation/ lab based
Shakibamanesh, A., Ghorbanian., M.	2017	--	Urban spaces	Individual movement and perception of time	Cause-and-effect analysis (field experiment)- virtual reality & exploratory field research
Seitz et al.	2016	--		Cognitive heuristics/ Crowd phenomena	Simulation model
Natapov, A., Fisher-Gewirtzman, D.	2016	--	Virtual urban spaces	Individual movement	Immersive Virtual reality
Yin et al.	2015	U.S.A.	Large urban areas	Pedestrian movement/ data collection	Google Street View data
Cheng et al.	2014	--	--	Group dynamics	Review
Wang et al.	2014	Hong Kong, China	Circulation region of a shopping mall	Attractors	Pedestrian simulation model—CityFlow-U
McArdle et al.	2014	Delft, Netherlands	Urban spaces	Pedestrian movement/ data collection and visualisation	Geovisual analytics
Canca et al.	2013	Zaragoza, Spain	Big exhibition events	Information of facilities, such as occupation, queue sizes and links density	Simulation model
Feng et al.	2021	--	--	Pedestrian movement/ data collection	Systematic review
Chavat et al.	2018	--	--	Pedestrian movement/ data collection	Experimental analysis/ Computational Intelligence system

Tang et al.	2020	--	Pedestrian crossing design	Symmetric intersection (SI)	Pedestrian simulation model
Duives et al.	2020	Assen, The Netherlands	Music event	Crowd monitoring system	Data fusion algorithm
Szczepanek, R.	2020	Cracow, Polland	Urban spaces	Pedestrian movement/ data collection	HD webcams
Resch et al.	2020	Salzburg, Austria and Cologne, Germany	Urban spaces	Emotions against the background of environmental information	Wearable physiological sensors combined with an eDiary app
Fiset et al.	2020	--	Pedestrian crossing design	Limb movements	Experimental analysis, lab (virtual environment)
Tan et al.	2020	--	Urban spaces	Visually impaired	MY VISION system/ image processing system - Experimental analysis
Johansson et al.	2020	Sweden	Urban spaces	Nightime lighting	video-based method/ experimental analysis with 62 pedestrians
Hagos et al.	2020	Addis Ababa, Ethiopia	Sidewalks	Sidewalk vendors	Social-force based pedestrian microsimulation model (PTV-Viswalk-11)
Lyons, G.	2020	UK	Urban spaces	Mobility as a Service/ WaaS assessment	Experimental analysis
Rossetti et al.	2020	Brescia, Italy	Transport hubs	--	GIS-based approach
Liu, Y., Kaneda, T.	2020	Shanghai, China	Waterfront	Crowd dynamics, emergency response	Agent-based simulation
Mills et al.	2020	Melbourne, Australia and Aarhus, Denmark	Urban spaces	Pedestrian movement/ data collection	IoT data streams/ Experimental analysis

Greenberg et al.	2020	Haifa and Jerusalem, Israel	Urban spaces	Cognitive model: mutual visibility and physical effort as cognitive bases	Agent-based simulation
Almahmood, M., Skov-Petersen, H.	2020	Copenhagen, Denmark	Urban spaces	Interactive pedestrian simulation	Agent-based simulation, observational data
Zhang et al.	2020	--	Urban spaces	Pedestrian flow	cellular automata model, simulation
Leveque et al.	2020	--	Street crossing	Pedestrians' road crossing gaze behaviour	Review
Kremer et al.	2020	--	--	Distracted behaviours	Agent-based simulations
Aljuaid et al.	2020	Saudi Arabia	Urban spaces	smart system to sense and control the crowd flow	Experimental analysis
Hanna, S.	2020	London, UK	--	Predictor of pedestrian	Turner and Penn's (2002) exosomatic visual architecture (EVA) agents (simulation)
Kolivand et al.	2020	--	--	Social force model	Agent-based modelling
Fernandez-Ares et al.	2020	--	Urban spaces	Detect the movement of people	WiFi data collection,
Wang, L., Shen, S.	2019	--	Urban spaces	heuristic force-based model	Agent-based modelling
Zhao et al.	2019	--	Urban spaces	Psychological behaviors: social repulsive force	Experimental analysis
Li et al.	2019	Nanjing, China	Urban spaces	Pedestrian–level wind (PLW) environments	Simulation
Dias et al.	2019	--	Corners	Corner simulation	Agent-based modelling
Jiang et al.	2019	--	Road-crossing	Virtual agents and human interaction	Virtual environment
Karbovskii et al.	2019	India	Intersections and corners	Crowd safety at mass gatherings	PULSE simulation environment - agent based modelling
Omer, I., Kaplan, N.	2019	--	Urban spaces	Network effects	Agent-based modelling, Space syntax
Akpulat, M., Ekinci, M.	2019	--	Urban spaces	Crowd analysis	Image processing

Zheng, J., Peng, J.	2019	--	--	Pedestrian detection algorithm with multisource face images	Experimental analysis, machine learning
Cheung et al.	2019	--	--	Crowd analysis	Video processing, machine learning
Liao et al.	2019	China	Urban spaces	Eye tracking	Eye movement data mining, machine learning (cross-validated a random forest classifier)
Wang, W., Adamczyk, P.G.	2019	--	Urban spaces	Gait in the Real World	Wearable Movement Sensors, experimental analysis
Crosby et al.	2019	Coventry, UK	Urban environments	Euclidean metrics	Statistics (covariance matrix for Kriging)
Kim et al.	2019	--	Urban spaces	Openness Index (OI)	Simulations
Wu et al.	2019	--	--	Pedestrian dead-reckoning (PDR) system	Algorithm development through experiments
Tamaki et al.	2019	--	--	Method for semantic segmentation of pedestrian trajectories	Agent-based modelling
Jiang et al.	2019	--	--	Pedestrian Tracking	Prototype creation, PTrack, on LG smartwatch
Göçer et al.	2019	Turkey	Outdoor spaces	Pedestrian Tracking	Spatial statistical analyses that in-volve ANN, MC and SD, spatio-temporal mapping method in GIS, Logistic GW
Duives et al.	2019	--	--	Crowd management: a data-driven procedure to forecast crowd movements	GPS traces, Recursive Neural Network (RNN) with a Gated Recurrent Unit (GRU)
Hussein, M., Sayed, T.	2019	Vancouver, BC	Urban spaces	Validation methods	Agent-based modelling, computer vision (video footage)
Li et al.	2019	--	--	Pedestrian simulation modelling	Agent-based modelling

Wang, S.M., Huang, C.J.	2019	Fengjia, China	Urban commercial area	Key factors that influence pedestrian spatial behaviors and pedestrian accessibility	Space Syntax and Information Visualization (deep-learning simulation using the generative adversarial network (GAN))
Pelé et al.	2019	Strasbourg, France and Nagoya, Japan	Road-crossings	Social coordination problems, decision-making	Agent-based modelling
Engelniederhammer et al.	2019	Hong Kong, China	Crowded spaces	invasion of a personal space, emotional response	Emotional responses were measured psychophysiologically via a wearable device, a movement detection sensor
Lin et al.	2019	--	--	Person re-identification	Visual mechanism, recurrent model
Brunyé et al.	2018	--	Intersections	Wayfinding	Experimental analysis: virtual environment
Fisher-Gewirtzman, D.	2018	--	Urban spaces	Urban intensification, Predicted Simulation	Experimental analysis, visualization laboratory
Antigny et al.	2018	--	Urban spaces	Augmented reality applications	AR applications
Crociani et al.	2018	--	Urban spaces	Computer simulation for pedestrian dynamic	Agent-based modelling, experimental analysis
Tordeux et al.	2018	--	Urban spaces	Large-scale simulation of pedestrian dynamics	Agent-based modelling
Lambert et al.	2018	Ibaraki, Japan	Urban spaces	Data collection platform	Lidar, camera, IMU and odometry
Croft, J.L., Panchuk, D.	2018	--	Urban spaces	Collisions with other pedestrians	Experimental analysis, eye tracking device
Biassoni et al.	2018	--	Crossings	Pedestrian crossings by adults and children	Eye tracking in images
Zaki, M.H., Sayed, T.	2018	Vancouver, British Columbia	Urban spaces	Counting of pedestrians in groups	video data collection, classification method

Jiang et al.	2018	China	Crosswalk area	Pedestrians ‘crossing behaviour and visual attention	Experimental study, HD videos and an eye tracker
Yu et al.	2018	--	Lane formation	Field based model for pedestrian dynamics	Agent-based modelling, Simulation
Kalantarov et al.	2018	Be'er Sheva, Israel	Road crossing	Crossing decision, body parts’ movement and fullbody movement	Experimental analysis, 3D motion capturing system
VINAYAGA-SURESHKANTH et al.	2018	--	Crossings	Safety	Experimental analysis, classification
Rashdan et al.	2018	--	Urban spaces	Agent-based social simulation	Agent-based modelling
Hausmann et al.	2017	Germany	Urban spaces	Tools for Participation on Foot	a mobile application and a web app, field study
Graells-Garrido et al.	2017	Santiago, Chile	Urban spaces	Pokémon Go on people’s mobility patterns in a city	Travel Survey, Statistics (Negative Binomial Regression (NB))
Liao et al.	2017	--	Urban spaces	Differences of visual attention in pedestrian navigation when using 2D maps and 3D geo-browsers (self-localization, spatial knowledge acquisition, and decision-making)	Eye tracking, Empirical study, digital interface
Almodfer et al.	2017	--	Crossings	mi-croscopic pedestrian simulation models, crossing speeds	Quantitative and qualitative approaches
Yang et al.	2017	--	Urban spaces	pedestrian dynamics	Agent-based modelling
Omer, I., Kaplan, N.	2017	--	Urban spaces	Network effects	Statistics (Multiple Regression Analysis (MRA)), Space syntax
Negri, P., Garayalde, D.	2017	--	Urban spaces	pedestrian dynamics	Movement Feature Space (MFS), video sequence
Ma et al.	2016	--	road-crossing situation	Human-like microscopic pedestrian flow	Artificial intelligence, microscopic pedestrian movement, simulation experiments
Luo et al.	2016	--	Urban spaces	Fatigue coefficient	Agent-based modelling (multivelocity field floor cellular

					automata (FFCA) model), pedestrian experiments
Xie et al.	2016	Hong Kong, China	Signalized crosswalks	Route choice at signalized crosswalks	Agent-based modelling, simulation, observation data
Mesmer, B.L., Bloebaum, C.L.	2016	--	--	Velocity vector decisions foundin human movement	Decision and game theories, simulation (Vacate-GT emergency egress simulator)
Crociani, L., Lammel, G.	2016	--	Urban spaces	Travel behaviour	Simulation (cellular automaton (CA)model), small-scale experiment
Pouke et al.	2016	Oulu, Finland	Urban spaces	Travel behaviour	Wi-Fi network, 'Time-varying Origin–Destination matrix, Blender modelling software
Jung et al.	2016	Korea	Roads	Tree classification, safety	Parametric modeling approach, classification model
Morrongiello et al.	2016	--	Crossings	risk of injury for child pedestrians	immersive virtual reality system
Fu et al.	2016	--	Long channel	low pedestrian dynamicswith emotion propagation: panic	Agent-based simulation (Multi-grid model)
Takayanagi et al.	2015	--	Urban spaces	Pedestrian flow	Intermittent recording images, A method to visualize pedestrian trajectories using RGB values in an analyzing platform
Crooks et al.	2015	--	Urban spaces	Pedestrian prediction	Agent-based modelling
Lee, S.J.	2015	--	Urban spaces	Vision-driven agents: visual perception and movement	Agent-based modelling
Morrongiello et al.	2015	--	Crossings	Time Pressure, Crossing behaviours	Immersive virtual reality system
Rio et al.	2014	--	Pathways	Walking speed and direction: behavioral dynamics of following	Experimental analysis, virtual environment

Chen, L.	2014	--	High-density urban living	Predict pedestrian traffic	Agent-based modelling, empirical data
Ma et al.	2014	--	Signalised intersection	Pedestrian phase patterns	Agent based modelling, Case study and sensitivity analysis
Kim et al.	2014	--	Crowded spaces	Directional pedestrian counting	Hybrid map-based model, principal component analysis
Dietrich et al.	2014	--	Urban spaces	Agent-based modelling	Optimal Steps Model and the Gradient Navigation Model (comparison analysis)
Li et al.	2014	--	Urban spaces	Real-time pedestrian detection and tracking	Gaussian mixture model
Lee, S.-J., Lee, K.-H.	2014	--	Urban spaces	Natural movement model	Agent-based modelling (visual dynamics analysis (VDA) model) - programmed with NetLog
Mora, R., Astudillo, H., Bravo, S.	2014	Santiago, Chile	Urban spaces	exomatic visual architecture - predict pedestrian movement	Agent-based modelling
Duives, D.C., Daamen, W., Hoogendoorn, S.P.	2013	--	--	Pedestrian simulation models	Review
Kneidl, A., Hartmann, D., Borrmann, A.	2013	--	--	Pedestrian flows	Agent-based modelling
Vizzari, G., Bandini, S.	2013	--	Urban spaces	Pedestrian and Crowd Dynamics	Review- agent-based modelling
Lee et al.	2013	--	Urban spaces	Natural movement model	Pedestrian simulation model (NetLogo)
Qiu, F., Hu, X.	2013	--	--	Pedestrian crowd behavior	Agent-based modelling
Wirz et al.	2013	London, UK	Urban spaces	Mass gatherings monitoring	Location-aware smartphones, CoenoSense platform, statistical analysis
Bunevska Talevska et al.	2012	--	Low speed urban street	Level of service (LOS)	Simulation model (SFStreetSIModel)

Bartsch et al.	2012	--	Urban spaces	Pedestrian recognition	Automotive radar sensors, feature extraction, classification
Raffat, R.	2012	Muzdalifah, Saudi Arabia	Urban spaces	Crowd management	Ral-time virtual environment, spatial analysis, multi-agents system
Ren, M., Karimi, H.A.	2012	Pittsburgh, USA	Urban spaces	pedestrian/wheelchair navigation	Pattern recognition algorithm (feature extraction, classification), GPS positions, orientation data from compass, and movement states recognized from accelerometer data
Yuen, J.K.K., Lee, E.W.M.	2012	--	Urban spaces	Moving crowd	Agent-based modelling
Jodoin et al.	2012	--	Urban spaces	Changes in scene dynamics, behavior subtraction	Agent-based modelling, video analytics
Vogt et al.	2012	Fira, Greece	Urban spaces	Pathfinding	Space syntax (Destination-Based Space Syntax Simulator (DBS3))
Khider et al.	2012	Germany	Localised environments	Pedestrian movement	Bayesian filtering techniques
Torrens et al.	2012	--	--	Urban pedestrian flows	Agent-based modelling
White, D.A., Barber, S.B.	2012	Oaxaca, Mexico	Urban spaces	Modeling past movement	GIS simulation (FETE model)
Wang et al.	2011	--	Urban spaces	Pedestrian detection algorithm with images	Classification and recognition - manifold learning
Terada, Y., Saitoh, F.	2011	Japan	Passageways	Pedestrian detection algorithm with images	Tracking methods- moving objects
Orellana, D., Wachowicz, M.	2011	Amsterdam, Netherlands	Urban spaces	Movement Suspension	Exploratory statistical approach, GPS recordings

Calabrese et al.	2011	Rome, Italy	Urban spaces	Real-time pedestrian detection with cell phones	Localizing and Handling Network Event Systems (LocHNESs) platform
Fridman, N., Kaminka, G.A.	2010	--	Urban spaces	Cognitive model	Agent-based modelling
Zainuddin et al.	2010	--	Urban spaces	decision-making, familiarity	Review (agent-based modelling)
Gidel et al.	2010	--	Urban spaces	Real-time pedestrian detection	Multilayer Laser Scanner

Appendix B Ethical approvals & Data management statement

An ethical approval was obtained for all the data collected and used as part of the Thesis via the Cranfield University Research Ethics System (CURES). The letters of approval and the individual reference numbers for each data collection activity are shown below in Figure B-1, Figure B-2, Figure B-3.

Cranfield University

10 November 2019

Dear Ms Stanitsa ,

Reference: CURES/9406/2019
Title: Digital METRics OPtimising Outcomes for infrastructure Life-cycle Integration Strategies

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/9406/2019 **has been reviewed. You may now proceed with the research activities you have sought approval for.**

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

Figure B-1: Letter of approval for the collection and use of the questionnaire data from the CURES system with reference number CURES/9406/2019

4 May 2020

Dear Ms Stanitsa ,

Reference: CURES/10888/2020
Title: Digital METROPOLiS Wi-Fi Data

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/10888/2020 **has been reviewed. You may now proceed with the research activities you have sought approval for.**

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

Figure B-2: Letter of approval for the collection and use of the Wi-Fi data from the CURES system with reference number CURES/10888/2020

17 August 2021

Dear Ms Stanitsa ,

Reference: CURES/14449/2021
Title: PhD Digital Metropolis

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/14449/2021 **has been reviewed. You may now proceed with the research activities you have sought approval for.**

If you have any queries, please contact CURES Support.

We wish you every success with your project.

Regards,

CURES Team

Figure B-3: Letter of approval for the collection and use of the semi-structured interview data from the CURES system with reference number CURES/14449/2021

A data management plan was also developed and shared with The Crown Estate. This plan formed part of the location data acquisition described in Chapters 5 and 6 and the approval process for its use in this study (Figure B-4). The same data management plan was then followed for all the types of data collected as part of this research.

DIGITAL METROPOLiS

Digital METRics OPtimising Outcomes for infrastructure Life-cycle Integration Strategies

Data management plan

A data management plan will be shared to the diverse stakeholders to be agreed and signed off before the commence of the use of data.

1.0 Security of data and back- up

Data will be stored in a dedicated project space on Cranfield University's network drive, which is regularly backed-up to protect against accidental and malicious data loss. Data management will be aligned to the university wide Management of Research Data Policy with a Research Data Manager on board. Backed-up files will be stored according to the University's policy on data handling, access and data security. Data files will be password protected and controlled access will be assured to these files.

2.0 Data availability

All data collected will be made anonymous prior to analysis and reporting, ensuring that no information will be traced back to an individual. Confidential information (including personal data) will be destroyed and disposed of securely once it is no longer required, after agreed periods of retention have expired, or in cases where destruction is required for legal or ethical reasons.

3.0 Data copyright/ IPR Ownership

Ownership of the intellectual property rights of data produced will be with Cranfield University and SNC Lavalin Atkins. The project will be supported by Cranfield University where professional expertise in intellectual property management will be on hand to advise on any copyright and IPR related issues that arise.

Copyrights and full acknowledgements of all the involved stakeholders funding to conduct the work will be conveyed in all publications, ensuring that approval is sought and received prior to use of official logos where necessary.

Figure B-4: Data management plan developed as part of the PhD data handling requirements

Appendix C Dataset Samples: Raw Data

Table C-1: Sample dataset as collected from the online questionnaire referenced in Chapter 3 and 4.

Response ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18
R_U9kjBIr1L0ufVV7	4	3	5	4	4	4	4	3	2	3	4	2	4	4	5	3	4	2
R_ToV0TnwV0D8wc8x	5	3	4	4	3	5	4	4	2	2	3	1	4	3	4	3	5	4
R_9oj2YaRahYLU1mp	3	2	4	3	1	2	3	1	1	3	2	4	2	2	4	4	3	1
R_21cXzwuRX6vgdGc	3	2	4	2	2	3	3	3	1	2	3	3	4	2	5	3	1	2
R_1rfahkyUdM3osnd	4	4	5	3	2	4	5	3	2	4	5	3	5	4	4	3	2	2
R_XjgwG26Mr2l8B57	3	2	4	4	3	4	4	2	2	3	2	2	4	4	3	2	1	1
R_WCmzCUTFA15QHvz	4	3	4	3	4	4	3	3	2	3	3	2	5	4	3	3	3	2
R_1iaLhhprrTMMZG5	4	4	4	3	3	4	4	1	1	3	2	1	4	4	4	3	3	1
R_s5Ny8qfnNSYfPhL	4	3	4	3	4	3	4	5	1	3	3	2	4	2	4	3	2	1
R_33EjCfKKatOUqTb	5	5	3	5	5	4	4	3	1	2	5	3	3	3	4	3	3	1
R_5vYr2A2HDFaxmlX	4	4	4	3	3	4	4	4	4	3	3	2	3	3	4	3	1	1
R_3lWAKmO5xxfnUiq	4	4	3	3	3	3	4	4	2	3	3	2	4	3	4	3	4	3

Response ID	Q19	Q20	Q21	Q22	Q23	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31	Q32	Q33	Q34	Q35	Q36
R_U9kjBIr1L0ufVV7	4	4	5	1,2,3,6,7,8,10,11	5	5	3	5	5	5	3,4,6	4	5	5	5	5	2,7,8,20	3,9,13,17
R_ToV0TnwV0D8wc8x	4	3	4	1,2,3,6,7,8,10	2	5	4	1	3	3	3,4,6,9	3	4	4	3	4	4,5,6,7,8,9,11,16,18	2,6,7,8,10
R_9oj2YaRahYLU1mp	4	5	4	1,6,8,9	4	4	4	1	4	2	1,2,4	2	5	5	5	5	5,7,8,18	4,6,7,9,15
R_21cXzwuRX6vgdGc	5	3	5	6	5	5	3	5	4	5	8	2	5	5	4	5	8,9,10	2,7,13,16
R_1rfahkyUdM3osnd	4	4	5	1,2,6,8	5	5	4	2	5	2	1,4,5,6,9	4	5	5	5	5	2,4,7,10,15	2,7,9,13
R_XjgwG26Mr2l8B57	4	4	4	2,3,10,11	4	3	4	4	4	2	2,4,5,6,9	4	5	5	5	5	2,8,10	2,5,13
R_WCmzCUTFA15QHvz	4	4	4	1,2,5,7,8,10	4	4	5	4	5	2	2,5,6	3	5	5	4	4	2,4,8,10,17	2,9,14
R_1iaLhhprrTMMZG5	5	3	4	1,2,5,8,10	5	4	4	1	4	1	8	3	4	4	4	3	2,4,5,7,13	2,5,7
R_s5Ny8qfnNSYfPhL	4	4	5	1,2,8,10	5	4	2	2	1	2	1,2,5,6	2	4	4	4	4	4,5,7,8,9,10,12,13,14,15,18	4,5,7,13,15
R_33EjCfKKatOUqTb	5	4	4	2,4,10	5	4	5	4	2	1	6,9,7	4	5	5	4	5	2,10,15	2,5,9,13,14,16,17
R_5vYr2A2HDFaxmlX	4	3	4	2,6,10	4	5	2	5	5	4	1,2	3	3	4	5	5	2,10,20	9,17
R_3lWAKmO5xxfnUiq	3	3	3	2,8	4	4	4	1	3	4	6,9	4	4	4	4	4	1,8,11,18	1

Table C-2: Sample data set of the Wi-Fi data used and referenced in Chapter 5 and 6.

Timestamp	ID	Z	X	Y	Node Id	Signal Strength	Node Id	Signal Strength	Node Id	Signal Strength
1501630610	B072BF5C9CCA	0	51.515289	-0.141437	AC8674932D58	-94				
1501587120	74BADB36014F	0	51.515458	-0.141847	AC8674932C50	-80				
1501587135	74BADB36014F	0	51.515883	-0.141032	AC8674932C50	-76				
1501587150	74BADB36014F	0	51.515991	-0.141943	AC8674932C50	-75				
1501587165	74BADB36014F	0	51.515799	-0.140626	AC8674932C50	-79	AC8674932808	-87		
1501587195	74BADB36014F	0	51.515178	-0.143623	AC8674932C50	-81	AC8674932808	-87		
1501587395	74BADB36014F	0	51.515617	-0.142311	AC8674932808	-80				
1501613580	EC1F7280BA50	0	51.515372	-0.142359	AC8674932D28	-76				
1501593755	4C74BFB70EBC	0	51.421442	-0.3764	AC8674932808	-86				
1501593760	4C74BFB70EBC	0	51.421442	-0.3764	AC8674932808	-86				
1501593770	4C74BFB70EBC	0	51.515617	-0.142311	AC8674932808	-85				
1501593810	4C74BFB70EBC	0	51.462208	-0.264582	AC8674932D28	-82				
1501593815	4C74BFB70EBC	0	51.462208	-0.264582	AC8674932D28	-82				
1501593820	4C74BFB70EBC	0	51.450926	-0.29983	AC8674932D28	-82				
1501618675	8866A5B8993D	0	51.515512	-0.142141	AC8674932808	-85	AC8674932D28	-89	AC8674932C50	-91
1501618685	8866A5B8993D	0	51.515512	-0.142141	AC8674932808	-85	AC8674932D28	-89	AC8674932C50	-91
1501618700	8866A5B8993D	0	51.515526	-0.142145	AC8674932808	-83	AC8674932D28	-89	AC8674932C50	-91
1501618705	8866A5B8993D	0	51.515526	-0.142145	AC8674932808	-83	AC8674932D28	-89	AC8674932C50	-91
1501621175	8866A5B8993D	0	51.515128	-0.141764	AC8674932D58	-87				
1501586655	8CF5A3066245	0	51.515842	-0.142112	AC86745E06D0	-81				
1501586700	8CF5A3066245	0	51.516375	-0.142286	AC8674932808	-81				
1501586740	8CF5A3066245	0	51.516476	-0.142143	AC8674932808	-79	AC8674932D28	-90		

1501586760	8CF5A3066245	0	51.516476	-0.142143	AC8674932808	-79	AC8674932D28	-90		
1501621755	AC5F3E301F24	0	51.515263	-0.141834	AC86745E06D0	-95				
1501621790	AC5F3E301F24	0	51.51454	-0.141957	AC86745E06A0	-83	AC8674932C68	-91		
1501621810	AC5F3E301F24	0	51.514559	-0.141956	AC86745E06A0	-85	AC8674932C68	-91		
1501590355	38F23EA838D0	0	51.515227	-0.141908	AC86745E06D0	-92				
1501590470	38F23EA838D0	0	51.515314	-0.142005	AC8674932D58	-85	AC8674932C50	-86	AC8674932808	-89
1501590475	38F23EA838D0	0	51.515314	-0.142005	AC8674932D58	-85	AC8674932C50	-86	AC8674932808	-89
1501590535	38F23EA838D0	0	51.513889	-0.142367	AC8674932D40	-88				
1501590545	38F23EA838D0	0	51.513889	-0.142367	AC8674932D40	-87				
1501585400	D0C5F3E0CD65	0	51.515554	-0.142567	AC8674932808	-86	AC8674932D28	-91		
1501618520	54EF92527E92	0	51.515251	-0.141795	AC8674932D58	-76	AC8674932C50	-84		
1501618550	54EF92527E92	0	51.515232	-0.141853	AC8674932D58	-76	AC8674932C50	-81	AC8674932D40	-85
1501618580	54EF92527E92	0	51.514926	-0.142023	AC8674932C50	-81	AC8674932D40	-85		
1501618600	54EF92527E92	0	51.513889	-0.142367	AC8674932D40	-85				
1501618620	54EF92527E92	0	51.513981	-0.142191	AC8674932D40	-81				
1501620175	54EF92527E92	0	51.515371	-0.142047	AC8674932C50	-76	AC8674932D28	-82	AC86745E06D0	-88
1501620200	54EF92527E92	0	51.515371	-0.142047	AC8674932C50	-76	AC8674932D28	-82	AC86745E06D0	-88
1501625005	54EF92527E92	0	51.515172	-0.142082	AC86745E06D0	-70	AC8674932D58	-71	AC8674932808	-72